Privacy-Preserving Reasoning for Hypergraphical Knowledge Bases

George Voutsadakis[1,2], Jie Bao[3], Giora Slutzki[1], and Vasant Honavar[1]

[1] Department of Computer Science, Iowa State University, Ames, IA 50011
[2] School of Mathematics and Computer Science,
Lake Superior State University, Sault Ste. Marie, MI 49783
[3] Department of Computer Science,
Rensselaer Polytechnic Institute, Troy, NY 12180

Abstract. Many semantic web applications require selective sharing of ontologies between autonomous entities due to copyright, privacy or security concerns. In our previous work it was shown that, on such occasions, an agent who wishes to hide part of its ontology while sharing the rest may still be able to answer safely queries against its knowledge base using inferences based on both hidden and visible knowledge without revealing the hidden knowledge. Moreover, it was shown how this framework may be applied to the case of hierarchical ontologies. We extend the theory to cover privacy-preserving reasoning with information modeled using hypergraphs. We apply this extension to obtain privacy-preserving reachability reasoning for RDF Graphs.

1 Introduction

The widespread adoption and use of networked information systems in virtually every area of human endeavor call for sharing of information among autonomous individuals and across organizations to faciliate productive interaction and collaboration. However, the need to share information among business partners, different governmental agencies (e.g., intelligence, law enforcement, public policy), or independent nations acting on matters of global concern (e.g., counterterrorism) often has to be balanced against the need to protect sensitive or confidential information from unintended disclosure, e.g., due to copyright, privacy, security, or commercial considerations. In such settings, there is often a compelling need for selective sharing of the results of inference using both public and private knowledge, without compromising private knowledge.

The majority of current proposals for *policy languages* [14] forbid access to the private/hidden parts of an ontology when answering queries against the ontology. However, in [1], the authors have argued that such approaches are overly restrictive because there are scenarios where it is possible and may be desirable for a knowledge base to use both hidden and visible knowledge to answer queries without risking disclosure of the hidden knowledge. Such reasoning was termed *privacy-preserving reasoning*. In [1] a precise formulation of the problem of privacy-preserving reasoning was provided and a framework was developed to tackle the problem based on the Open World Assumption (OWA). The notions of a reasoner, of a privacy-preserving reasoner and of a reasoning strategy were formalized and conservative extensions were used to provide a set of privacy-preserving reasoning strategies for description logics. Privacy-preserving reasoning strategies for the special case of hierarchical ontologies, i.e., ontologies that can be represented as directed acyclic graphs, were analyzed based on a reduction of reasoning to graph reachability.

However, many practical applications require more expressive knowledge bases. For example, modeling some aspects of RDF knowledge bases, which are widely used in semantic web applications, requires the use of hypergraphs [8]. It is therefore of interest to study a framework for performing privacy-preserving reasoning using "hypergraphical" knowledge bases. Reasoning on a hypergraph may take various forms. In this paper we formalize this by expressing deductions using rules of inference in much the same way as is done in ordinary logical systems. In this way, we are providing the user with a wide choice of reasoners to pick from depending on the application at hand. As an illustration, we deal in some detail with the case of reachability in hypergraphs. Briefly, a hypergraph representing the information contained in a knowledge base is given. Some of its hyperedges are visible and some are hidden. Using both visible and hidden hyperedges, the reasoner answers queries concerning the reachability of a given vertex from another given vertex without compromising the hidden knowledge. This framework includes the one for hierarchical ontologies as a special case.

As an application of this "hypergraphical" reasoning we look at the problem of *resource connectivity* in RDF Graphs [8]. This is specifically the problem of determining whether two given resources, as described by an RDF knowledge base, are reachable one from the other in the hypergraphical representation of the RDF Graph. Hayes and Gutierrez [8] considered this problem without the privacy-preservation aspect. We revisit the problem but assume that some of the triples in the RDF Graph are hidden. We use both visible and hidden triples to reveal connectivity of resources when it is possible to do so without compromising the hidden information. It is emphasized that we *do not deal* in this work with the RDF semantics of the triples, nor do we use RDF syntax in our queries. We rather represent the RDF triples as hyperedges of a given hypergraph and use our hypergraphical privacy-preserving reasoning framework to study only the problem of connectivity of RDF resources. We plan to extend this work to the more general problem of privacy-preserving reasoning for RDF Graphs (that takes into account the RDF semantics of the triples) by examining the special case of ρDF entailment, first studied by Muñoz, Pérez and Gutierrez in [13].

2 Preliminaries: Privacy-Preserving Reasoning

A *knowledge base* (KB) K over a language L consists of a set of *axioms* $K = \{\alpha_1, ..., \alpha_n\}$. We assume that K is consistent and does not contain tautologies. We use $\mathrm{SIG}(\alpha_i)$ to denote the set of names occurring in an axiom α_i and $\mathrm{SIG}(K)$ to denote the *signature* of a KB K, i.e., $\mathrm{SIG}(K) = \cup_{i=1}^{n} \mathrm{SIG}(\alpha_i)$. The set of axioms that make up a KB K is divided into two mutually exclusive parts: a *visible part* K_v and a *hidden part* K_h, with the corresponding signatures $\mathrm{SIG}(K_v)$ and $\mathrm{SIG}(K_h)$. We call $\mathrm{SIG}(K_v)$ the *visible signature*, $\mathrm{SIG}(K_h) - \mathrm{SIG}(K_v)$ the *hidden signature* and we write $K = (K_v, K_h)$.

Example 1: Consider a company, say *U-Travel* that provides travel information to online customers. Suppose *U-Travel* offers a query service that provides limited information to the public but more detailed information to paying subscribers. The *U-Travel*'s ontology contains the following knowledge: (a) `Sun Lodge` is a `2-star hotel` (b) a `2-star hotel` is an `inn` (c) `Sun Lodge` is `AAA-discountable` and (d) an `inn` is a `hotel`.

Suppose *U-Travel* is willing to reveal that "`Sun Lodge` is a `hotel`" to the public, yet it wants to hide the fact that "`Sun Lodge` is a `2-star hotel`" from all but its paying subscribers. If the *U-Travel* query service could not use hidden information, i.e., that `Sun Lodge` is a `2-star hotel`, it would not be able to inform a non-paying subscriber that `Sun Lodge` is a `hotel`, although it is possible to do so, without compromising hidden knowledge. This is formalized by defining an ontology $K = (K_v, K_h)$ of the *U-Travel* company. We use the *partial-order* relation $\leq$ to indicate concept inclusion. The hidden part K_h contains

$$\texttt{SunLodge} \leq \texttt{2StarHotel} \qquad \texttt{2StarHotel} \leq \texttt{Inn}$$

and the visible part K_v contains

$$\texttt{SunLodge} \leq \texttt{AAADiscountable} \qquad \texttt{Inn} \leq \texttt{Hotel}$$

Thus, the visible signature is $\mathrm{SIG}(K_v) = \{\texttt{SunLodge}, \texttt{Inn}, \texttt{AAADiscountable}, \texttt{Hotel}\}$, $\mathrm{SIG}(K_h) = \{\texttt{SunLodge}, \texttt{2StarHotel}, \texttt{Inn}\}$. Hence, the hidden signature is $\mathrm{SIG}(K_h) - \mathrm{SIG}(K_v) = \{\texttt{2StarHotel}\}$. □

Let K be a KB over a language L, Q the query space over L, i.e., the set of possible assertions to be tested against K, and A an answer space. A **reasoner** R for K is an algorithm that defines a function $R : Q \to A$. Some natural requirements that need to be met by a reasoner operating in this privacy-preserving setting are:

1. **Honesty**. The reasoner should not "lie". That is, answers produced by the reasoner should always be *consistent* with its KB.
2. **History Independence**. The reasoner should always respond to a given query q against a fixed KB K with the same answer regardless of the *history* of queries that have been posed against K.
3. **Safety**. The reasoner must ensure that the answers it produces are *safe*, in the sense that it is not possible for a querying agent to infer any piece of hidden knowledge based on the answers to past queries and the visible part of the KB.

An immediate consequence of this definition is that a reasoner R is "history independent" in the sense suggested by Requirement 2 above.

Our basic approach to designing privacy-preserving reasoners for KBs that contain hidden knowledge is to ensure that the answers to queries do not reveal hidden knowledge. The central idea is to design a reasoner that exploits the *Open World Assumption* (OWA) of ontology languages to make it impossible for the querying agent to distinguish between information that is unknown to the reasoner (because of the incompleteness of the KB) and the knowledge that is being protected by the reasoner. A query that cannot be safely answered without running the risk of disclosing hidden knowledge will be answered *as if* the reasoner *lacks the complete knowledge* to answer the query.

We use $K \vdash \gamma$ to mean that γ is *classically provable* from K. Thus $\vdash \alpha$ means that α is a tautology. If every axiom in a KB K_2 is classically provable from another KB K_1, we say that K_1 *entails* K_2 and denote it as $K_1 \vdash K_2$. A reasoner R might employ an *inference engine* which can be viewed as a classical reasoner C with answer space $A = \{Y, N\}$, such that $\forall q \in Q, C(q) = Y$ iff $K \vdash q$. While an inference engine always responds in a truthful manner, the reasoner, in order to protect some parts of K, may use an answering strategy which does not respond with the "whole truth". For example, a reasoner may answer "U" (Unknown) even if the correct answer (from the inference engine) is "Y" or "N". The answer to a query q may be "U" either because the reasoner has incomplete knowledge (i.e., $K \nvdash q$ and $K \nvdash \neg q$) under the Open World Assumption (OWA), or because the "truthful" answer to q might risk disclosure of hidden knowledge.

Definition 1 (Privacy-Preserving Reasoner). *Let $K = (K_v, K_h)$ be a* KB *over a language L, Q the query space in L, $A = \{\mathrm{U}, \mathrm{Y}, \mathrm{N}\}$ the answer space, and R a reasoner for K. We define: $Q_Y = R^{-1}(Y), Q_N = R^{-1}(N), Q_U = R^{-1}(U)$ and further assume that $q \in Q_Y$ iff $\neg q \in Q_N$.*

*(a) R is **strongly privacy-preserving** w.r.t. K if it satisfies the following two axioms:*
- *Honesty Axiom: $q \in Q_Y \Rightarrow K \vdash q$.*
- *Strong Safety Axiom: $\forall \alpha$ such that $\nvdash \alpha$ and $\mathrm{SIG}(\alpha) \subseteq \mathrm{SIG}(K_h)$, $K_h \vdash \alpha \Rightarrow (K_v \cup Q_Y \nvdash \alpha)$.*

*(b) R is **weakly privacy-preserving** w.r.t. K if it satisfies the Honesty Axiom and the following axiom:*
- *Weak Safety Axiom: $\forall \alpha$, $\alpha \in K_h \Rightarrow (K_v \cup Q_Y \nvdash \alpha)$*

The honesty axiom requires that reasoners provide answers that do not contradict the given KB (i.e., $K \cup Q_Y$ is consistent). The strong safety axiom requires that the answers provided by reasoners do not disclose any consequence that can be drawn from the hidden knowledge alone. The weak safety axiom requires the reasoner to protect only axioms (and their semantically equivalent syntactic variants) that are explicitly mentioned in the hidden part of the KB (but not necessarily their consequences).

Definition 2 (Strategy). *Let L be a language, $\boldsymbol{K}_L$ the class of all knowledge bases over L, and $\boldsymbol{R}_L$ the class of all reasoners over $\boldsymbol{K}_L$. A* **strategy for** L *is a function $\mathfrak{R} : \boldsymbol{K}_L \to \boldsymbol{R}_L$, such that, for every $K \in \boldsymbol{K}_L$, $R = \mathfrak{R}(K)$ is* *a reasoner for K. The strong/weak safety scope of a strategy $\mathfrak{R}$, $Scope(\mathfrak{R}) = \{K \in \boldsymbol{K}_L | \ \mathfrak{R}(K)$ is a strongly/weakly privacy-preserving reasoner for $K\}$.*

A strategy needs to compromise between the two apparently conflicting goals of *generality*, i.e., of having the largest possible scope, and of *informativeness*, i.e., of yielding reasoners that provide as much information as possible.

3 Privacy-Preserving Reasoning with Hypergraphs

In this section, we provide a new application of the framework for privacy-preserving reasoning that was developed in [1] and was briefly reviewed in Section 2. More precisely, we show how the framework can be adapted to perform various types of privacy-preserving reasoning with information that can be represented in the form of a hypergraph. Even though "hypergraphical ontologies" form a rather special class of arbitrary ontologies, they still merit special consideration since, on the one hand, they extend hierarchical ontologies, which have many applications in practice, and, on the other, provide additional tools for dealing with information that is not directly expressible in hierarchical form. The RDF connectivity problem studied in the next section provides such an example.

Definition 3 (Hypergraph). *A* **hypergraph** *is a pair $G = \langle V, \mathcal{E} \rangle$, where V is the set of* **nodes** *and $\mathcal{E} = \{E_i\}_{i \in I}$ is a family of subsets of V, called* **edges**. *G is* **simple** *if all edges are distinct. G is r-***uniform** *if all edges have cardinality r. An r-uniform hypergraph is* **ordered** *if the occurrence of nodes in every edge is ordered. The class of all hypergraphs is denoted by* **HG**.

Let $G = \langle V, \mathcal{E} \rangle$, with $\mathcal{E} = \{E_i\}_{i \in I}$, be a simple hypergraph. Assume $\mathcal{E} = \mathcal{E}_v \cup \mathcal{E}_h$, with $\mathcal{E}_v = \{E_j\}_{j \in J}$ and $\mathcal{E}_h = \{E_k\}_{k \in K}$, such that $J \cup K = I$ and $J \cap K = \emptyset$. $\mathcal{E}_v$ is the set of *visible edges* and $\mathcal{E}_h$ is the set of *hidden edges*. G is called r-*ordered* if it is both r-uniform and ordered.

Example 2: Consider the hypergraph described pictorially in Figure 1. It has three visible edges E_1, E_2, E_3 and one hidden edge E_4. ■

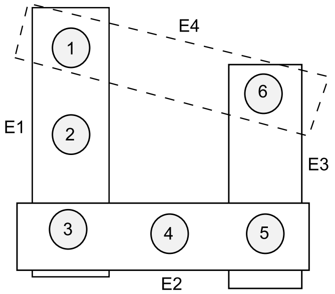

Fig. 1. A Pictorial Representation of a Hypergraph.

To extend our framework for privacy-preserving reasoning with hierarchical ontologies to privacy-preserving reasoning with arbitrary hypergraphs we introduce *closure properties* on the potential edges of a hypergraph. More specifically, given a hypergraph $G = \langle V, \mathcal{E} \rangle$, a closure property P on the set of subsets of V is a property that may be expressed by "rules of inference". The following examples illustrate this idea:

Examples: (a) For 2-ordered hypergraphs, i.e., for directed graphs, consider **reachability**. Formally, the closure property can be expressed as $P =$"for all vertices x, y, z, (x, z) follows from (x, y) and (y, z)". The corresponding rule of inference is

$$\frac{(x, y), (y, z)}{(x, z)}.$$

(b) For arbitrary hypergraphs, the t-**exclusive closure** is $P =$"for all $E, F, D \subseteq V$, such that $D \subseteq E \cap F$ and $|D| = t$, $(E \cup F) \backslash D$ follows from E, F".

The corresponding rule of inference is

$$\frac{\{x_1, \ldots, x_n, z_1, \ldots, z_t\}, \{z_1, \ldots, z_t, y_1, \ldots, y_m\}}{\{x_1, \ldots, x_n, y_1, \ldots, y_m\}}.$$

This rule will be used when studying privacy-preserving resource connectivity in RDF Graphs. ■

For instance, looking back at the example of Figure 1, it is easily seen that 1-exclusive closure allows the following two derivations.

$$\frac{\{1, 2, 3\}, \{3, 4, 5\}}{\{1, 2, 4, 5\}}, \qquad \frac{\{3, 4, 5\}, \{5, 6\}}{\{3, 4, 6\}}.$$

Given a hypergraph $G = \langle V, \mathcal{E} \rangle$, together with a collection IR of rules of inference, $E \subseteq V$ is said to be *directly derivable by*, or *inferred from*, a set $E_1, \ldots, E_n \in \mathcal{P}(V)$ if there exists a rule R in IR, such that $\frac{E_1, \ldots, E_n}{E}$ is an instance of R. Given $\mathcal{A} \cup \{E\} \subseteq \mathcal{P}(V)$, we say that $\mathcal{A}$ *entails* E, written $\mathcal{A} \vdash_G^{\mathrm{IR}} E$, if there exists a sequence $A_0, \ldots, A_n \in \mathcal{P}(V)$, such that $A_n = E$ and, for all $i \leq n$, $A_i \in \mathcal{A}$ or A_i is inferred from $A_{j_1}, \ldots, A_{j_k}$, for some $j_1, \ldots, j_k < i$. Such a sequence is called a *proof of E from $\mathcal{A}$*. Define $C_G^{\mathrm{IR}} : 2^{\mathcal{P}(V)} \to 2^{\mathcal{P}(V)}$ by

$$C_G^{\mathrm{IR}}(\mathcal{A}) = \{E \subseteq V : \mathcal{A} \vdash_G^{\mathrm{IR}} E\}.$$

Note that the entailment relation $\vdash$, that was used in the general definition of reasoner (see Section 2), is replaced in this context by the specific entailment relation $\vdash_G^{\mathrm{IR}}$. Evidently, different rules of inference give rise to different entailment relations.

Given a hypergraph $G = \langle V, \mathcal{E} \rangle$, together with a collection IR of rules of inference for G, we set the query space to be $Q = \mathcal{P}(V)$ and the answer space $A = \{Y, N, U\}$. A *reasoner* for G is then a function $R : \mathcal{P}(V) \to \{Y, N, U\}$.

Definition 4 (Privacy-Preserving Reasoner). *Let the set of edges $\mathcal{E}$ of the hypergraph $G = \langle V, \mathcal{E} \rangle$ be partitioned into a visible part $\mathcal{E}_v$ and a hidden part $\mathcal{E}_h$ and let* IR *be a collection of rules of inference. Then, a* weakly-privacy preserving reasoner *for G (w.r.t.* IR*) is a reasoner $R : Q \rightarrow A$, that satisfies the axioms*

1. **Honesty:** $Q_Y \subseteq C_G^{\mathrm{IR}}(\mathcal{E})$;
2. **Weak Safety:** $C_G^{\mathrm{IR}}(\mathcal{E}_v \cup Q_Y) \cap \mathcal{E}_h = \emptyset$.

A strongly-privacy preserving reasoner *for G, on the other hand, is a reasoner $R : Q \rightarrow A$, that satisfies Honesty and*

2'. **Strong Safety:** $C_G^{\mathrm{IR}}(\mathcal{E}_v \cup Q_Y) \cap C_G^{\mathrm{IR}}(\mathcal{E}_h) = \emptyset$.

As an illustration of the concepts presented in Definition 4, consider again Example 2. Let us focus on reachability reasoning, i.e., on inferring conclusions about whether a given vertex is reachable from another vertex by following a sequence of partially overlapping edges. Then, it is clear that a privacy-preserving reasoner for G cannot answer "Y" to all three queries concerning the visible edges of G, since, in that case, it would compromise the hidden information that $\{1, 6\}$ is an edge in the hypergraph.

These definitions of privacy-preserving reasoners on hypergraphs generalize the corresponding definitions for weakly and strongly privacy-preserving reasoners, respectively, for hierarchical ontologies that were presented in [1]. In fact, very similarly to the case of hierarchical ontologies, one obtains

Lemma 1. *R is a strongly privacy-preserving reasoner for $G = \langle V, \mathcal{E}_v \cup \mathcal{E}_h \rangle$ iff R is a weakly privacy-preserving reasoner for $G^+ = \langle V, \mathcal{E}_v \cup C_G^{\mathrm{IR}}(\mathcal{E}_h) \rangle$.*

Proof: For honesty, notice that $C_G^{\mathrm{IR}}(\mathcal{E}_v \cup \mathcal{E}_h) = C_G^{\mathrm{IR}}(\mathcal{E}_v \cup C_G^{\mathrm{IR}}(\mathcal{E}_h))$. For the Safety condition, R is a strongly privacy-preserving reasoner for $G = \langle V, \mathcal{E}_v \cup \mathcal{E}_h \rangle$ iff $C_G^{\mathrm{IR}}(\mathcal{E}_v \cup Q_Y) \cap C_G^{\mathrm{IR}}(\mathcal{E}_h) = \emptyset$ iff R is a weakly privacy-preserving reasoner for $G = \langle V, \mathcal{E}_v \cup C_G^{\mathrm{IR}}(\mathcal{E}_h) \rangle$. ■

Moreover, as in the case of hierarchical ontologies (see [1]), a hierarchy of privacy-preserving reasoning strategies for hypergraphical ontologies may be obtained based on the generality and the informativeness of the reasoners that they produce.

Let $G = \langle V, \mathcal{E}_v \cup \mathcal{E}_h \rangle$ be a hypergraph and IR a collection of rules of inference for reasoning with G. As before, $\vdash_G^{\mathrm{IR}}$ denotes the hypergraphical inference according to IR and C_G^{IR} the corresponding closure operator. We now proceed to define several classes of hypergraphs that have safe strategies with different degrees of informativeness. Define, first, given any integer $n \geq 0$,

$$C_G^n(\mathcal{A}) := C_G^{\mathrm{IR}^n}(\mathcal{A}) = \{E \subseteq V : \mathcal{A} \vdash_G^{\mathrm{IR}} E \text{ via a proof of length at most } n\}.$$

Since the set IR of rules of inference will be assumed to be fixed in a specific context, we choose to omit it in order to simplify notation.

Now define the following classes of hypergraphs, for all $m, n \geq 0$:

$$\mathbf{S}_{m,n} = \{G \in \mathbf{HG} : C_G^n(C_G^m(\mathcal{E}) - \mathcal{E}_h) \cap \mathcal{E}_h = \emptyset\}.$$

If m or n are substituted by $+$, then the full closure operator $C_G^+ := C_G^{\mathrm{IR}}$ will be assumed. Hypergraphs in the class $\mathbf{S}_{m,n}$ are termed (m, n)-*safe*. Intuitively, m represents a restriction on the ability of the reasoner to detect possible safety hazards and n a similar restriction on the ability of the querying agent to discover knowledge from previous answers and the visible part of the hypergraph. Note that $\mathbf{S}_{+,+} := \bigcap_{m=1}^{\infty} \mathbf{S}_{m,+}$. The following reasoners are members of the hierarchy $\mathbf{S}_{m,+}, m \geq 1$.

The dummy reasoner: A dummy reasoner responds to every query with the answer "U". It preserves the safety of precisely those hypergraphs that satisfy $C_G(\mathcal{E}_v) \cap \mathcal{E}_h = \emptyset$, i.e., its safety scope is $\mathbf{S}_{1,+}$. This strategy has the widest safety scope but is the least informative.

The obvious reasoner: An obvious reasoner responds with an answer "Y" to only those queries that follow from $\mathcal{E}_v$. Its weak safety scope is also $\mathbf{S}_{1,+}$.

The safe reasoner: A safety reasoner has $Q_Y = C_G(\mathcal{E}) - \mathcal{E}_h$. This reasoner satisfies by definition Honesty and it satisfies Weak Safety iff $C_G(C_G(\mathcal{E}) - \mathcal{E}_h) \cap \mathcal{E}_h = \emptyset$. Thus, its weak safety scope is $\mathbf{S}_{+,+}$. If one sets $Q_Y = C_G^m(\mathcal{E}) - \mathcal{E}_h$, then the safety scope becomes $\mathbf{S}_{m,+}$.

The naive reasoner: A naive reasoner always gives away all the information that it has, i.e., $Q_Y = C_G(\mathcal{E})$. It is trivially honest but its weak safety scope consists only of those hypergraphs with no hidden edges.

4 Privacy-Preserving Reachability Reasoning for RDFs

In [8], the authors introduce a representation of RDF Graphs, i.e., sets of RDF triples, in the form of *RDF bipartite graphs*, which are ordinary labeled bipartite graphs. Using this representation they provide a satisfactory algorithmic answer to the problem of *resource connectivity* in RDF Graphs. We sketch in this section how the framework of Section 3 may be applied in order to obtain privacy-preserving resource connectivity reasoning with RDF Graphs. Contrasted to the work presented in [8], our work introduces the privacy-preservation aspect in the reasoning and, moreover, resource connectivity is tackled more directly, using the hypegraphical representation of an RDF Graph rather than, first, transforming the RDF Graph to an ordinary bipartite graph.

We now provide some more details regarding this framework.

Definition 5. *An* RDF statement *is a triple* (a, b, c)*.* a *is called the* subject*,* b *the* predicate *and* c *the* object *of the statement.* a, b *and* c *can be* URI*'s,* literals *or* blank nodes*. They are collectively referred to as* values*. The only restriction that is applied to the syntax is that* b *must be a URI. An* RDF Graph T *is a set of RDF triples. The set of all values occurring in* T *is denoted by* $\mathrm{univ}(T)$*.*

Let T *be an RDF Graph. A* path *or* triple path P *is a sequence of RDF triples* $(t_1, t_2, \ldots, t_n)$*, with* $t_k = (s_k, p_k, o_k)$*, such that* $\{s_i, p_i, o_i\} \cap \{s_{i+1}, p_{i+1}, o_{i+1}\} \neq \emptyset$*,* $i < n$*. A triple path* $(t_1, \ldots, t_n)$ *is said to* connect resources x *and* y *if* $n = 1$ *and* $x, y \in t_1$ *or* $x \in \{s_1, p_1, o_1\}, x \notin \{s_i, p_i, o_i : 1 < i \leq n\}$ *and* $y \in \{s_n, p_n, o_n\}, y \notin \{s_i, p_i, o_i : 1 \leq i < n\}$*.* x *is* reachable from y *if there exists a triple path that connects* x *and* y*.*

Thus, the problem of RDF reachability or, equivalently, connectivity of RDF resources is the problem of determining, given an RDF Graph and two of its resources, whether there exists a triple path connecting these resources.

To solve the RDF resource connectivity problem, Hayes and Gutierrez [8] transform the given RDF Graph T into a labeled bipartite graph $\beta(T)$. They show that two resources x, y are connected in T if and only if there exists a path in $\beta(T)$ between the corresponding nodes v_x and v_y. This enables them to apply ordinary graph reachability algorithms to solve the resource connectivity problem.

Although the framework of [8] does not involve *privacy-preserving reasoning*, the results of Section 3 may be applied in this context to study privacy-preserving resource connectivity in RDFs. Roughly speaking, this is the problem of inferring connectivity between various RDF resources without revealing hidden connections. We formulate this problem and provide a solution in what follows.

Let T be a *partially hidden RDF Graph*, i.e., a set of triples T together with a partition $T = T_v \cup T_h$ of T into a subset T_v of *visible triples* and a subset T_h of *hidden triples*. Since we concentrate on the problem of connectivity of resources, the ordering of the elements in the triple is irrelevant. So T will be represented by the hypergraph $G := G(T) = \langle V, \mathcal{E} \rangle$, with $V = \text{univ}(T)$, $\mathcal{E} = \{\{s,p,o\} : (s,p,o) \in T\}$, such that $\mathcal{E} = \mathcal{E}_v \cup \mathcal{E}_h$ is also partitioned into a visible part $\mathcal{E}_v = \{\{s,p,o\} : (s,p,o) \in T_v\}$ and a hidden part $\mathcal{E}_h = \{\{s,p,o\} : (s,p,o) \in T_h\}$.

Consider the class $\mathbf{HG}_{\leq 3}$ of all hypergraphs $G = \langle V, \mathcal{E} \rangle$, such that $|E| \leq 3$, for all $E \in \mathcal{E}$. Let EW consists of the following two inference rules:

- **1-Exclusion Rule:** $\frac{\{x,y,z\},\{z,w\}}{\{x,y,w\}}$
- **Weakening Rule:** $\frac{\{x,y,z\}}{\{x,y\}}$.

The 1-Exclusion Rule simulates in this context the reasoning taking place when one uses transitivity to reveal new relations from given ones in a partial ordering. To provide an explanation for the Weakening Rule, consider as the intended meaning of a triple the existence of a mutual pairwise relationship between its members. Then, the Weakening Rule expresses the fact that, if three elements are pairwise-related, then any two of them also are.

A *proof of* $E \subseteq V$ from a set $\mathcal{A} \subseteq \mathcal{P}(V)$ is defined as before and, if there exists a proof of E from $\mathcal{A}$, we write $\mathcal{A} \vdash_G^{\text{EW}} E$. $C_G^{\text{EW}}(\mathcal{A}) = \{E \subseteq V : \mathcal{A} \vdash_G^{\text{EW}} E\}$ denotes the corresponding closure operator.

Example 3: Consider an RDF Graph T, whose standard pictorial representation is given in Figure 2. The hypergraph $G(T)$ associated with T is

$$\begin{aligned} G(T) = \{ & \{\text{``}\mathit{Slutzki}\text{''}, \mathit{type}, \mathit{Literal}\}, && (1)\\ & \{\mathit{slutzki}, \mathit{last\text{-}name}, \text{``}\mathit{Slutzki}\text{''}\}, && (2)\\ & \{\mathit{slutzki}, \mathit{type}, \mathit{Faculty}\}, && (3)\\ & \{\mathit{slutzki}, \mathit{teaches}, \mathit{game\text{-}theory}\}, && (4)\\ & \{\mathit{teaches}, \mathit{Domain}, \mathit{Faculty}\}, && (5)\\ & \{\mathit{teaches}, \mathit{Range}, \mathit{Course}\}, && (6)\\ & \{\mathit{game\text{-}theory}, \mathit{type}, \mathit{Course}\}\} && (7) \end{aligned}$$

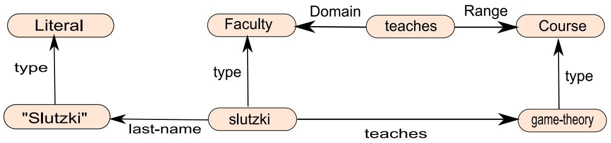

Fig. 2. A Pictorial Representation of an RDF Graph.

Consider $\mathcal{A} = \{(5), (6), (7)\} \subseteq G(T)$ and $E = \{\{game\text{-}theory, Faculty\}\}$. Then $\mathcal{A} \vdash^{\mathrm{EW}}_{G(T)} E$ is witnessed by the following proof:

(5)	(An Axiom)
$\{teaches, Faculty\}$	(By Weakening)
(6)	(An Axiom)
$\{Faculty, Range, Course\}$	(By 1-Exclusion)
$\{Faculty, Course\}$	(By Weakening)
(7)	(An Axiom)
$\{game\text{-}theory, type, Faculty\}$	(By 1-Exclusion)
$\{game\text{-}theory, Faculty\}$	(By Weakening)

■

A *weakly* EW-*privacy preserving reasoner* for reachability in $G \in \mathbf{HG}_{\leq 3}$ is a privacy-preserving reasoner according to Definition 4, where IR is replaced by EW. Similarly, we may define a *strongly* EW-*privacy-preserving reasoner* for G. As a corollary of Lemma 1, we obtain

Corollary 1. *R is a strongly EW-privacy-preserving reasoner for $G = \langle V, \mathcal{E}_v \cup \mathcal{E}_h \rangle$ iff R is a weakly EW-privacy-preserving reasoner for $G^+ = \langle V, \mathcal{E}_v \cup C^{\mathrm{EW}}_G(\mathcal{E}_h) \rangle$.*

Example 3 (Cont'd): Suppose now that $\mathcal{E}_h = \{(2)\}$ and let $E = \{Faculty, Literal\}$. $E \in C^{\mathrm{EW}}_{G(T)}(\mathcal{E})$ but $C^{\mathrm{EW}}_{G(T)}(\mathcal{E}_v \cup \{E\}) \cap \mathcal{E}_h \neq \emptyset$. Therefore, every weakly privacy-preserving reasoner for $G(T)$ must answer "U" to the query E.

Suppose, next, that $\mathcal{E}_h = \{(2), (4)\}$ and $E = \{\text{"}Slutzki\text{"}, Faculty\}$. In this case, $E \in C^{\mathrm{EW}}_{G(T)}(\mathcal{E})$ and $C^{\mathrm{EW}}_{G(T)}(\mathcal{E}_v \cup \{E\}) \cap C^{\mathrm{EW}}_{G(T)}(\mathcal{E}_h) \neq \emptyset$. Thus, every strongly privacy-preserving reasoner for $G(T)$ must answer "U" to the query E. ■

The following proposition attests to the fact that privacy-preserving reasoning using EW corresponds exactly to reasoning about connectivity of resources in RDF Graphs.

Proposition 1. *In an RDF Graph T, resource b is reachable from resource a iff $G(T) \vdash^{\mathrm{EW}}_{G} \{a, b\}$.*

Proof:

$\Rightarrow$: Suppose b is reachable from a. Then, there exists a sequence of RDF triples $(s_1, p_1, o_1), \ldots, (s_n, p_n, o_n)$, such that $a \in \{s_1, p_1, o_1\}, b \in \{s_n, p_n, o_n\}$ and $\{s_i, p_i, o_i\} \cap \{s_{i+1}, p_{i+1}, o_{i+1}\} \neq \emptyset$, for all $i < n$. Suppose, without loss of generality, that $a = s_1, b = o_n$ and $o_i = s_{i+1}$, for all $i < n$. We show by induction on i that, for all $i \leq n$, $G(T) \vdash_G^{\mathrm{EW}} \{a, o_i\}$.

For $i = 1$, this follows by the Weakening Rule. Assume it is true for $i = k-1$. Then $G(T) \vdash_G^{\mathrm{EW}} \{a, o_{k-1}\}$ and, by the definition of $\vdash_G^{\mathrm{EW}}$ and the fact that $s_k = o_{k-1}$, $G(T) \vdash_G^{\mathrm{EW}} \{o_{k-1}, p_k, o_k\}$. Hence, by the 1-Exclusion Rule, $G(T) \vdash_G^{\mathrm{EW}} \{a, p_k, o_k\}$ and, by the Weakening Rule, $G(T) \vdash_G^{\mathrm{EW}} \{a, o_k\}$, as desired.

$\Leftarrow$: Suppose, conversely, that $G(T) \vdash_G^{\mathrm{EW}} \{a, b\}$. Then, there exists a proof $E_1, \ldots, E_n$ of $\{a, b\}$ from $\mathcal{E}$ in $\vdash_G^{\mathrm{EW}}$. We show, by induction on $i \leq n$ that, for all $x, y \in E_i$, y is reachable from x in the RDF graph T.

For $i = 1$, E_1 is an edge in $\mathcal{E}$, whence, if $x, y \in E_1$, y is reachable from x. Assume that, for all $i < k$, and all $x, y \in E_i$, y is reachable from x in T. If $E_k \in \mathcal{E}$, the conclusion follows as in the base of the induction. If E_k follows from E_j, $j < k$, by the Weakening Rule, then the conclusion follows trivially from the induction hypothesis. If $E_k = \{x, y, w\}$ follows from $E_i = \{x, y, z\}, E_j = \{z, w\}$, $i, j < k$, by the 1-Exclusion Rule, then, by the induction hypothesis, z is reachable from x, y in T and w is reachable from z in T, whence any element in any pair chosen from among x, y, w is reachable from the other element in the pair in T. ■

Proposition 1 shows that the study of privacy-preserving resource connectivity in an RDF Graph $T = T_v \cup T_h$ is equivalent to $\vdash_G^{\mathrm{EW}}$-privacy-preserving reasoning on $G(T) = \langle V, \mathcal{E}_v \cup \mathcal{E}_h \rangle$.

5 Summary and Discussion

Related Work: Problems of trust, privacy and security in information systems in general, and networked information systems (e.g., the web) in particular, are topics of significant current interest. For early work on security and access control policies in computer systems and databases see the survey [2]. Recent work on *policy languages* for the web [4, 14, 3, 15, 11, 10, 7, 12] focuses on specifying syntax-based restrictions on access to specific resources or operations on the web. Research on encryption of sensitive information focuses on preventing unauthorized access to such information using cryptographic protocols [6]. In contrast to our work, access control policies and encryption techniques do not allow the use of hidden knowledge to answer queries even if it might be possible to do so without risking its disclosure. Farkas et al. [5, 9] have proposed a *privacy information flow model* to prevent unwanted inferences in data repositories. Their framework, as opposed to ours, uses *closed world semantics*. Jain and Farkas [9] have proposed an RDF authorization model that assigns a security label to each (stored or inferred) RDF triple using a pre-specified set of *syntactic* rules. On the other hand, our approach to privacy-preserving reasoning, and, as a consequence, also to RDF resource connectivity reasoning uses a *semantics*-based approach.

Summary: In this paper we extended on our previous work on privacy-preserving reasoning. We developed a framework for performing privacy-preserving reasoning with various inference systems on hypergraphical ontologies, i.e., on knowledge bases that may be represented by hypergraphs. We used the hypergraphical inference framework to give an example on how privacy-preserving reachability reasoning, which is tantamount to privacy-preserving resource connectivity, in RDF Graphs may be carried out. In future work we hope to tackle general privacy-preserving reasoning with RDF Graphs starting for simplicity with the ρdf fragment, presented in [13].

References

1. Jie Bao, Giora Slutzki, and Vasant Honavar. Privacy-preserving reasoning on the semantic web. In *Web Intelligence*, pages 791–797, 2007.
2. Elisa Bertino, Latifur R. Khan, Ravi S. Sandhu, and Bhavani M. Thuraisingham. Secure knowledge management: confidentiality, trust, and privacy. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 36(3):429–438, 2006.
3. Piero A. Bonatti, Claudiu Duma, Norbert Fuchs, Wolfgang Nejdl, Daniel Olmedilla, Joachim Peer, and Nahid Shahmehri. Semantic web policies - a discussion of requirements and research issues. In *ESWC*, pages 712–724, 2006.
4. Piero A. Bonatti and Daniel Olmedilla. Rule-based policy representation and reasoning for the semantic web. In *Reasoning Web*, pages 240–268, 2007.
5. Csilla Farkas, Alexander Brodsky, and Sushil Jajodia. Unauthorized inferences in semi-structured databases. *Information Sciences*, 176(22):3269–3299, Nov. 2006.
6. Mark Giereth. On partial encryption of rdf-graphs. In Yolanda Gil, Enrico Motta, V. Richard Benjamins, and Mark A. Musen, editors, *International Semantic Web Conference*, volume 3729 of *Lecture Notes in Computer Science*, pages 308–322. Springer, 2005.
7. Simon Godik and Tim Moses (ed.). Oasis extensible access control markup language (xacml). OASIS Committee Secification cs-xacml-specification-1.0, November 2002, http://www.oasis-open.org/committees/xacml/, 2002.
8. Jonathan Hayes and Claudio Gutiérrez. Bipartite graphs as intermediate model for rdf. In *International Semantic Web Conference*, pages 47–61, 2004.
9. Amit Jain and Csilla Farkas. Secure resource description framework: an access control model. In *SACMAT*, pages 121–129, 2006.
10. Lalana Kagal, Timothy W. Finin, and Anupam Joshi. A policy based approach to security for the semantic web. In *International Semantic Web Conference*, pages 402–418, 2003.
11. Lalana Kagal, Massimo Paolucci, Naveen Srinivasan, Grit Denker, Timothy W. Finin, and Katia P. Sycara. Authorization and privacy for semantic web services. *IEEE Intelligent Systems*, 19(4):50–56, 2004.
12. Vladimir Kolovski, James A. Hendler, and Bijan Parsia. Analyzing web access control policies. In *WWW*, pages 677–686, 2007.
13. Sergio Muñoz, Jorge Pérez, and Claudio Gutiérrez. Minimal deductive systems for rdf. In *ESWC*, pages 53–67, 2007.
14. Gianluca Tonti, Jeffrey M. Bradshaw, Renia Jeffers, Rebecca Montanari, Niranjan Suri, and Andrzej Uszok. Semantic web languages for policy representation and reasoning: A comparison of kaos, rei, and ponder. In *International Semantic Web Conference*, pages 419–437, 2003.
15. Daniel J. Weitzner, James Hendler, Tim Berners-Lee, and Dan Connolly. Creating a policy-aware web: Discretionary, rule-based access for the world wide web. In E. Ferrari and B.Thuraisingham, editors, *Web and Information Security*. Idea Group, In press.