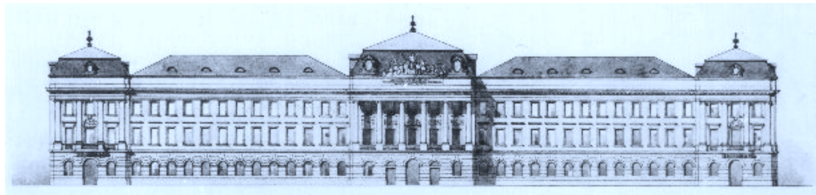
INFSYS
RESEARCH
REPORT



INSTITUT FÜR INFORMATIONSSYSTEME

ARBEITSBEREICH WISSENSBASIERTE SYSTEME

THE COMPLEXITY OF EXPLAINING NEGATIVE QUERY ANSWERS IN *DL-Lite*

Diego Calvanese   Magdalena Ortiz

Mantas Šimkus   Giorgio Stefanoni

INFSYS RESEARCH REPORT 1843-12-03

MARCH 2012

Institut für Informationssysteme
AB Wissensbasierte Systeme
Technische Universität Wien
Favoritenstraße 9-11
A-1040 Wien, Austria
Tel: +43-1-58801-18405
Fax: +43-1-58801-18493
sek@kr.tuwien.ac.at
www.kr.tuwien.ac.at

INFSYS RESEARCH REPORT
INFSYS RESEARCH REPORT 1843-12-03, MARCH 2012

THE COMPLEXITY OF EXPLAINING NEGATIVE QUERY ANSWERS IN *DL-Lite*

Diego Calvanese[1] Magdalena Ortiz[2] Mantas Šimkus[3]
Giorgio Stefanoni[4]

Abstract. In order to meet usability requirements, most logic-based applications provide explanation facilities for reasoning services. This holds also for DLs, where research has focused on the explanation of both TBox reasoning and, more recently, query answering. Besides explaining the presence of a tuple in a query answer, it is important to explain also why a given tuple is missing. We address the latter problem for (conjunctive) query answering over DL-Lite ontologies, by adopting abductive reasoning, that is, we look for additions to the ABox that force a given tuple to be in the result. As reasoning tasks we consider existence and recognition of an explanation, and relevance and necessity of a certain assertion for an explanation. We characterize the computational complexity of these problems for subset minimal and cardinality minimal explanations.

[1]Free Univ of Bozen-Bolzano, Italy. Email: calvanese@inf.unibz.it
[2]Vienna University of Technology, Austria. Email: ortiz@kr.tuwien.ac.at
[3]Vienna University of Technology, Austria. Email: ortiz@kr.tuwien.ac.at
[4]University of Oxford, United Kingdom. Email: Giorgio.Stefanoni@cs.ox.ac.uk

Acknowledgements: This work was partially supported by the Austrian Science Fund (FWF) grants P20840 and T515, the EU FP7 project ACSI (FP7-257593), and by Alcatel-Lucent and EPSRC.

This report accompanies the short paper with the same title that appears in the Proceedings of the 13th International Conference on Principles of Knowledge Representation and Reasoning (KR 2012). A preliminary version appears in the Proceedings of the 24th International Workshop on Description Logics (DL 2011).

1 Introduction

Query answering over ontologies formulated in Description Logics (DLs) has received considerable attention in both research and industry. Given an ontology, users pose queries over the conceptual schema and get answers that take into account the constraints specified at the conceptual level. Many efforts have concentrated on lightweight DLs. For instance *DL-Lite*$_{\mathcal{A}}$, the language at the basis of the OWL 2 QL profile [Motik et al., 2009], has been tailored for query answering over large data sets [Calvanese et al., 2009]. In this setting, expressive power is traded in favour of a better computational behaviour in terms of data-complexity. In fact, conjunctive query answering in *DL-Lite*$_{\mathcal{A}}$ enjoys FO-rewritability, i.e., it can be reduced to the problem of evaluating a suitable FO query over a database instance.

In order to meet usability requirements set by domain users, most logic-based applications provide explanation algorithms for reasoning services. This holds also for DLs, where research has focused on the explanation of TBox reasoning (cf. [McGuinness and Borgida, 1995; Borgida, Franconi, and Horrocks, 2000; Penaloza and Sertkaya, 2010; Horridge, Parsia, and Sattler, 2008]). Additionally, the problem of explaining positive answers to conjunctive queries over *DL-Lite*$_{\mathcal{A}}$ ontologies has been studied in [Borgida, Calvanese, and Rodríguez-Muro, 2008], where a procedure for computing the reasons for a tuple to be in the answers to a query is outlined. The same paper advocates the importance of computing explanations also for the absence of query answers. To the best of our knowledge, in the literature, the problem of explaining negative answers has been considered only for relational databases extended with provenance information. In particular, [Chapman and Jagadish, 2009] studied the problem of determining the database operations that prevented a given tuple to be in the answers to the query. Also, [Huang et al., 2008] focused on computing database updates fixing missing answers to given SQL queries. Unfortunately, typical ontologies do not provide provenance information and, thus, negative query answers can not be explained by adapting one of the available solutions.

For this reason, we formalize the problem of explaining the absence of a tuple in the context of query answering over DL ontologies. We adopt *abductive reasoning* [Eiter and Gottlob, 1995; Klarman, Endriss, and Schlobach, 2011], that is, we consider which additions need to be made to the ABox to force the given tuple to be in the result. More precisely, given a TBox $\mathcal{T}$, an ABox $\mathcal{A}$, and a query q, an *explanation* for a given tuple $\vec{c}$ is a new ABox $\mathcal{E}$ such that the answer to q over $\langle \mathcal{T}, \mathcal{A} \cup \mathcal{E} \rangle$ contains $\vec{c}$. An important aspect in explanations is to provide users with explanations that are simple to understand and free of redundancy, hence as small as possible. To address this requirement, we study various restrictions on explanations, in particular, we focus on subset minimal and cardinality minimal ones. We consider standard decision problems associated to logic-based abduction: *(i) existence* of an explanation; *(ii) recognition* of a given ABox as being an explanation; *(iii) relevance* and *(iv) necessity* of an ABox assertion, i.e., whether it occurs in some or all explanations. Additionally, it is important to allow one to restrict the signature of explanations. This can be used to consider only solutions that do not extend the ABox vocabulary: an important property in the context of accessing relational databases through ontologies, where database instances are defined over a small, fixed, vocabulary, and the terminological component is used to enrich that vocabulary. The idea of restricting the explanation signature is an adaptation of a concept introduced in [Baader et al., 2010], which studies among others the $\mathcal{CQ}$-emptiness problem. That is, given a query q over a TBox $\mathcal{T}$ decide whether for all ABoxes $\mathcal{A}$ over a given signature Σ, we have that evaluating q over $\langle\mathcal{T},\mathcal{A}\rangle$ leads to an empty result. In our framework, deciding the existence of an explanation generalizes the $\mathcal{CQ}$ non-emptiness problem. In fact, deciding whether there exists an explanation for a negative answer amounts to check whether a query admits a solution w.r.t. a TBox $\mathcal{T}$ *and* an ABox $\mathcal{A}$. In the following we sketch algorithms to solve the relevant reasoning tasks and give a precise characterization of their computational complexity for *DL-Lite*$_\mathcal{A}$. The complexity results for the various reasoning tasks are summarized in Table 1.

2 Preliminaries

***DL-Lite*$_\mathcal{A}$.** *DL-Lite*$_\mathcal{A}$ is a member of the *DL-Lite* family of DLs [Calvanese et al., 2009], which have been designed for dealing efficiently with large amounts of extensional information. Let N_C, N_R, N_I be, respectively, countably infinite sets of concept, role, and constant names. In the following, when the distinction between concept and role names is inessential, we refer to them as *predicate* names. Then in *DL-Lite*$_\mathcal{A}$, concept expressions (or, *concepts*) C, denoting sets of objects, and role expressions (or, *roles*) R, denoting binary relations between objects, are formed as follows:

$$C \longrightarrow A \mid \exists R, \qquad\qquad R \longrightarrow P \mid P^-,$$

where A is a concept name from N_C and P a role name from N_R[1]. In a *DL-Lite*$_\mathcal{A}$ ontology $\mathcal{O} = \langle\mathcal{T},\mathcal{A}\rangle$, the TBox $\mathcal{T}$ consists of axioms of the form:

$$\begin{array}{lll} C_1 \sqsubseteq C_2, & R_1 \sqsubseteq R_2, & \\ C_1 \sqsubseteq \neg C_2, & R_1 \sqsubseteq \neg R_2, & (\mathsf{funct}\ R), \end{array}$$

where the first row consists of *positive inclusions* among concepts or roles, while the second row contains *disjointness axioms* among concepts or roles and *functionality assertions* on roles. The ABox $\mathcal{A}$ consists of assertions of the form $A(c)$ and $P(c, c')$, where c, c' are constants in N_I.

The semantics of *DL-Lite*$_\mathcal{A}$, as usual in DLs, is based on first-order interpretations $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$. We adopt the *unique name assumption* (UNA), i.e., for every interpretation $\mathcal{I}$ and constant pair $c_1 \neq c_2$, we have $c_1^\mathcal{I} \neq c_2^\mathcal{I}$. An interpretation $\mathcal{I}$ is a *model* of $\mathcal{O} = \langle\mathcal{T},\mathcal{A}\rangle$, if it satisfies all the axioms in the TBox $\mathcal{T}$ and all the assertions in the ABox $\mathcal{A}$. We call $\mathcal{O}$ *consistent* if it admits at least one model.

Queries. Let N_V be a countably infinite set of variables. Expressions $A(t)$ and $P(t, t')$ are called *atoms*, where $t, t' \in N_V \cup N_I$. A *conjunctive query (CQ)* q of arity $n \geq 0$ is an expression $q(x_1, \ldots, x_n) \leftarrow a_1, \ldots, a_m$, where each a_i, $i \in \{1, \ldots, m\}$, is an atom. We consider *safe* CQs, i.e., each x_i, for $i \in \{1 \ldots, n\}$, occurs in at least one of the atoms of the query. Let $N_V(q)$ denote the set of variables occurring in q, $N_I(q)$ the set of constants in q, and let $at(q) = \bigcup_{1\leq i\leq m}\{a_i\}$. A *match* for q in an interpretation $\mathcal{I}$ is a mapping $\pi : N_V(q) \cup N_I(q) \rightarrow \Delta^\mathcal{I}$ such that

[1] We ignore here the distinction between data values and objects present in *DL-Lite*$_\mathcal{A}$ and OWL 2 QL, since it is immaterial for our results. That is, we do not consider value domains and attributes.

(i) $\pi(c) = c^{\mathcal{I}}$ for all $c \in N_I(q)$,
(ii) $\pi(t) \in A^{\mathcal{I}}$ for each $A(t) \in at(q)$, and
(iii) $\langle \pi(t), \pi(t') \rangle \in P^{\mathcal{I}}$ for each $P(t, t') \in at(q)$.

The tuple $\langle x_1, \ldots, x_n \rangle$ is the tuple of *answer variables* of q. The *answer* to q over $\mathcal{I}$, denoted $\mathsf{ans}(q, \mathcal{I})$, is the set of all n-tuples $\langle d_1, \ldots, d_n \rangle \in (N_I)^n$ such that $\langle d_1^{\mathcal{I}}, \ldots, d_n^{\mathcal{I}} \rangle = \langle \pi(x_1), \ldots, \pi(x_n) \rangle$, for some match π for q in $\mathcal{I}$. A CQ of arity 0 is called Boolean, and returns as answer either $\emptyset$ or the empty tuple $\langle\rangle$. We will write a Boolean CQ simply as a set of atoms. A *union of conjunctive queries* (UCQ) is a set of CQs of the same arity. For a UCQ q, we let $\mathsf{ans}(q, \mathcal{I}) = \bigcup_{q' \in q} \mathsf{ans}(q', \mathcal{I})$. The *certain answer* to a CQ or a UCQ q of arity n over $\mathcal{O}$ is defined as $\mathsf{cert}(q, \mathcal{O}) = \{\vec{c} \in (N_I)^n \mid \vec{c} \in \mathsf{ans}(q, \mathcal{I}), \text{ for each model } \mathcal{I} \text{ of } \mathcal{O}\}$.

Complexity Theory. We briefly outline the definition of some non-canonical complexity classes used in the paper, and refer to [Papadimitriou, 1994] for more details. The class Σ_2^{P} is a member of the Polynomial Hierarchy; it is the class of all decision problems solvable in non-deterministic polynomial time using an NP oracle. The class $\mathsf{P}_{\parallel}^{\mathsf{NP}}$ contains all the decision problems that can be solved in polynomial time with an NP oracle, where all oracle calls must be first prepared and then issued in parallel. The class DP contains all problems that, considered as languages, can be characterized as the intersection of a language in NP and a language in coNP. It is believed that PTIME $\subseteq$ NP $\subseteq$ DP $\subseteq \mathsf{P}_{\parallel}^{\mathsf{NP}} \subseteq \Sigma_2^{\mathsf{P}}$ is a strict hierarchy of inclusions. Here we make such an assumption.

In the following, we will use $[i..j]$ to denote the integer interval $\{i, \ldots, j\}$.

3 Explaining Negative Query Answers

In this paper we deal with the following problem:

Definition 1. Let $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a *DL-Lite*$_{\mathcal{A}}$ ontology, $q(\vec{x})$ a UCQ, and $\vec{c}$ a tuple of constants of arity $|\vec{x}|$. Further, assume a set $\Sigma \subseteq N_C \cup N_R$. We call $\mathcal{P} = \langle \mathcal{T}, \mathcal{A}, q, \vec{c}, \Sigma \rangle$ a *Query Abduction Problem* (QAP). An *explanation* for (or, a solution to) $\mathcal{P}$ is an ABox $\mathcal{E}$ such that:
(i) the concept and role names of $\mathcal{E}$ are contained in Σ,
(ii) the ontology $\mathcal{O}' = \langle \mathcal{T}, \mathcal{A} \cup \mathcal{E} \rangle$ is consistent, and
(iii) $\vec{c} \in \mathsf{cert}(q, \mathcal{O}')$.

The set of all explanations for $\mathcal{P}$ is denoted by $\mathsf{expl}(\mathcal{P})$. If $\Sigma = N_C \cup N_R$, we say that $\mathcal{P}$ has *an unrestricted explanation signature*. ∎

The predicates in Σ are the ones allowed in explanations, hence we call them *abducible predicates*. If $\vec{c} \notin \mathsf{cert}(q, \mathcal{O})$, we call $\vec{c}$ a *negative answer* to q over $\mathcal{O}$. Note that a query over the ontology can have a negative answer only if the ontology is consistent. Also, by condition *(ii)*, if the ontology is inconsistent then $\mathcal{P}$ does not have any explanation. Note also that $\mathcal{E}$ may contain constant names not present in $\mathcal{A}$. Next, we provide an example of a QAP.

Example 1. Let $\mathcal{A}$ be the following set of assertions about a particular university:

$$\begin{array}{ll} \mathsf{DPhil}(\mathit{Anna}), & \mathsf{DPhil}(\mathit{Beppe}), \\ \mathsf{enroll}(\mathit{Anna}, \mathit{KR}), & \mathsf{teach}(\mathit{Marco}, \mathit{KR}), \\ \mathsf{enroll}(\mathit{Luca}, \mathit{IDB}), & \mathsf{teach}(\mathit{Carlo}, \mathit{IDB}). \end{array}$$

That is, Anna and Beppe are doctoral students. Anna is enrolled in the KR course, which is taught by Marco, and Luca is enrolled in the introductory DB course (IDB), which is taught by Carlo. Now, consider the following TBox $\mathcal{T}$, in standard *DL-Lite*$_{\mathcal{A}}$ syntax, formalizing the university domain, of which $\mathcal{A}$ is a (partial) instance:

$$\begin{array}{rclrcl} \exists\mathsf{enroll} & \sqsubseteq & \mathsf{Student}, & \exists\mathsf{teach} & \sqsubseteq & \mathsf{Lecturer}, \\ \exists\mathsf{enroll}^- & \sqsubseteq & \mathsf{Course}, & \exists\mathsf{teach}^- & \sqsubseteq & \mathsf{Course}, \\ \mathsf{DPhil} & \sqsubseteq & \mathsf{Student}, & \mathsf{Course} & \sqsubseteq & \exists\mathsf{teach}^-. \end{array}$$

$\mathcal{T}$ models that objects in the domain of enroll are Students, and objects in the domain of teach are Lecturers, while objects in the range of enroll or of teach are Courses. Among the students we have DPhil students. Finally, every Course must be taught by someone.

Now, assume that the university administration is interested in finding all those who are teaching a course in which at least one of the enrolled students is a doctoral student, which is captured by the following query:

$$q(x) \leftarrow \mathsf{teach}(x, y), \mathsf{enroll}(z, y), \mathsf{DPhil}(z).$$

Assume that Carlo is expected to be part of the result, i.e., $\mathit{Carlo} \in \mathsf{cert}(q, \langle\mathcal{T}, \mathcal{A}\rangle)$. This is not the case, as Luca is the only student of Carlo and he is not known to be a doctoral student. Suppose that we have complete information on all the predicates but enroll and teach, i.e., only the latter predicates are abducible. It is easy to see that:

$$\{\mathsf{teach}(\mathit{Carlo}, \mathit{AI}), \mathsf{enroll}(\mathit{Beppe}, \mathit{AI}), \mathsf{enroll}(\mathit{Luca}, \mathit{AI})\}$$

is an explanation for the given negative answer, which suggests the existence of a course, AI, not present in $\mathcal{A}$. ■

This example shows that certain explanations may be too assumptive in that they include assertions that are not required to solve the problem. Indeed, in the example's explanation there is no reason to assume that Luca is enrolled in the freshly introduced course on Artificial Intelligence. In the following, we will examine various restrictions to $\mathsf{expl}(\mathcal{P})$ to reduce redundancy in explanations, achieved by introducing a preference relation among explanations. This relation is reflexive and transitive, i.e., we have a pre-order among explanations. For such a pre-order $\preceq$ on $\mathsf{expl}(\mathcal{P})$, we write $\mathcal{E} \prec \mathcal{E}'$ if $\mathcal{E} \preceq \mathcal{E}'$ and $\mathcal{E}' \npreceq \mathcal{E}$.

Definition 2. The preferred explanations $\mathsf{expl}_{\preceq}(\mathcal{P})$ of a QAP $\mathcal{P}$ under the pre-order $\preceq$, called $\preceq$-*explanations*, are defined as follows: $\mathsf{expl}_{\preceq}(\mathcal{P}) = \{\, \mathcal{E} \in \mathsf{expl}(\mathcal{P}) \mid \text{there is no } \mathcal{E}' \in \mathsf{expl}(\mathcal{P}) \text{ s.t. } \mathcal{E}' \prec \mathcal{E} \,\}$, i.e., $\mathsf{expl}_{\preceq}(\mathcal{P})$ contains all the explanations that are *minimal* under $\preceq$. ■

$\preceq$	$\preceq$-EXIST	$\preceq$-NEC	$\preceq$-REL	$\preceq$-REC
none	NP	coNP	NP	NP
$\leq$	NP	$\mathsf{P}_{\parallel}^{\mathsf{NP}}$	$\mathsf{P}_{\parallel}^{\mathsf{NP}}$	DP
$\subseteq$	NP	coNP	Σ_2^{P}	DP

Table 1: Summary of main complexity results for *DL-Lite*$_{\mathcal{A}}$ explanation (all are completeness results)

We consider two preference orders that are commonly adopted when comparing abductive solutions: the *subset-minimality order*, denoted by $\subseteq$, and the *minimum explanation size order*, denoted by $\leq$. The latter order is defined by $\mathcal{E} \leq \mathcal{E}'$ iff $|\mathcal{E}| \leq |\mathcal{E}'|$. Observe that $\mathsf{expl}_{\leq}(\mathcal{P}) \subseteq \mathsf{expl}_{\subseteq}(\mathcal{P})$.

Example 2. $\{\mathsf{teach}(Carlo, AI), \mathsf{enroll}(Beppe, AI)\}$ is a $\subseteq$-explanation in our example, whereas $\{\mathsf{enroll}(Beppe, IDB)\}$ is a $\leq$-explanation (and hence also a $\subseteq$-explanation). ■

We study here the four basic decision problems related to (minimal) explanations [Eiter and Gottlob, 1995], which are parametric w.r.t. the chosen preference order $\preceq$.

Definition 3. Given a QAP $\mathcal{P}$, define the following decision problems:

- $\prec$-EXIST(ENCE): Does there exist a $\preceq$-explanation for $\mathcal{P}$?
- $\prec$-NEC(ESSITY): Does a given assertion α occur in all $\preceq$-explanations for $\mathcal{P}$?
- $\prec$-REL(EVANCE): Does a given assertion α occur in some $\preceq$-explanation for $\mathcal{P}$?
- $\prec$-REC(OGNITION): Is a given ABox $\mathcal{E}$ a $\preceq$-explanation for $\mathcal{P}$?

Whenever no preference is applied (i.e., when $\preceq$ is the identity) we omit to write $\preceq$ in front of the problems' names. ■

In the next section, we study the complexity of these four problems in the light of the different preference relations.

4 Complexity of Explanations

In Table 1 we give an overview of our complexity results for explanation in *DL-Lite*$_{\mathcal{A}}$. We measure the complexity of a QAP $\mathcal{P} = \langle \mathcal{T}, \mathcal{A}, q, \vec{c}, \Sigma \rangle$ in terms of the combined size of $\mathcal{T}$, $\mathcal{A}$, and q, i.e., we consider *combined complexity*. Notice that we do not explicitly count the explanation signature Σ towards the complexity, since when it is restricted, its size is bounded by the size of the other parameters (see Proposition 2), and when it is unrestricted it is actually countably infinite and defined outside of the actual problem instance.

We briefly comment the results and provide then the details in the rest of the section. An algorithm for EXIST can non-deterministically guess an ABox $\mathcal{E}$ and check in polynomial time whether $\mathcal{E}$ is an explanation for the given QAP. The NP-hardness is proved by reducing the well-known problem of finding a homomorphism between two graphs. The latter lower-bound is subsequently used to characterize the complexity of REL, NEC, and $\subseteq$-NEC. Differently, $\leq$-REL and $\leq$-NEC are harder. The reason being that in order to solve these problems one has to compute first the minimal size of an explanation and, then, inspect all the explanations of that size. Additionally, there is another increase in complexity when dealing with $\subseteq$-REL. The intuition is that there is an exponential number of candidate explanations to examine and for each of them one has to check that none of its subsets is itself an explanation, which requires a coNP computation. Finally, the intuition for the NP upper bound of REC is that in order to solve the problem one needs to check consistency of the explanation with the ontology, and check whether the tuple is in the certain answer to the query. In case a preference order is in place, one has to check minimality as well. This check is coNP-hard for $\subseteq$- and $\leq$-minimality, leading to completeness for DP..

Next, we provide formal proofs of our results. One can prove that all the reductions we provide can be computed using logarithmic space.

4.1 Existence of Explanations

Before discussing our complexity results, we show that whenever a QAP $\mathcal{P} = \langle \mathcal{T}, \mathcal{A}, q, \vec{c}, \Sigma \rangle$ has an explanation, then $\mathcal{P}$ has an explanation $\mathcal{E}$ that is small in two senses.

First, all concepts and roles occurring in $\mathcal{E}$ occur either in $\mathcal{T}$ or in q. Indeed, we can remove from an arbitrary explanation all assertions that make use of predicates not in $\mathcal{T}$ or in q and all conditions in Definition 1 continue to be satisfied (the removed assertions are irrelevant for certain answers and for ontology consistency). Second, $\mathcal{E}$ is built from a small number of fresh constants. This can be shown using the FOL-rewritability of queries in *DL-Lite*$_\mathcal{A}$ [Calvanese et al., 2009].

Definition 4. Given an ABox $\mathcal{A}$, let $DB_\mathcal{A}$ be the interpretation whose domain $\Delta^{DB_\mathcal{A}}$ is the set of constants occurring in $\mathcal{A}$, and
(i) $A^{DB_\mathcal{A}} = \{o \mid A(o) \in \mathcal{A}\}$, for all $A \in N_C$;
(ii) $P^{DB_\mathcal{A}} = \{\langle o, o' \rangle \mid P(o, o') \in \mathcal{A}\}$, for all $P \in N_R$.

Proposition 1. [Calvanese et al., 2009] *Let $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a DL-Lite$_\mathcal{A}$ ontology, q a UCQ over $\mathcal{O}$, and PerfectRef$(q, \mathcal{T})$ the* perfect reformulation *of q w.r.t. $\mathcal{T}$, which is a UCQ obtained by applying the rewrite rules given in [Calvanese et al., 2009]. Then,* $\mathsf{cert}(q, \mathcal{O}) = \bigcup_{r \in PerfectRef(q,\mathcal{T})} \mathsf{ans}(r, DB_\mathcal{A})$.

In other words, the certain answers to a UCQ over an ontology $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ can be computed by rewriting each CQ q_i in it into a UCQ Q'_i to be evaluated over $\mathcal{A}$ alone, seen as a standard relational database. From this and Definition 1 it follows that if there is a solution to $\mathcal{P}$, then there exists an explanation $\mathcal{E}$ such that some CQ q'_i in Q'_i has a match in $\mathcal{E} \cup \mathcal{A}$. Furthermore, $|\mathcal{E}|$ is bounded by $|q'_i|$, while the number of fresh constants in $\mathcal{E}$ is bounded by the number of variables in q'_i. Indeed, a match for q'_i needs to map only the terms and the atoms occurring in the query. Since it follows from [Calvanese et al., 2009] that each q'_i in Q'_i has at most $|q_i|$ atoms and $2 \cdot |q_i|$ terms, we obtain:

Proposition 2. *Let $\langle \mathcal{T}, \mathcal{A} \rangle$ be a DL-Lite$_\mathcal{A}$ ontology. If $\mathcal{P} = \langle \mathcal{T}, \mathcal{A}, q, \vec{c}, \Sigma \rangle$ has an explanation, then $\mathcal{P}$ has an explanation $\mathcal{E}$ with concepts and roles only from $\mathcal{T}$ and q, at most* $\max(q)$ *atoms, and at most* $2 \cdot \max(q)$ *fresh ABox constants, where* $\max(q) = \max_{q_i \in q} |q_i|$.

We now turn to the complexity of finding explanations.

Algorithm 1 someExplanation

INPUT: QAP $\mathcal{P} = \langle \mathcal{T}, \mathcal{A}, q, \vec{c}, \Sigma \rangle$.
OUTPUT: yes iff $\mathcal{P}$ has an explanation.
Guess an ABox $\mathcal{E}$ constructed from Σ using original constants in $\mathcal{A}$ and fresh constants $o_1, \ldots, o_{\max(q)}$.
Guess a rewriting $r(\vec{c})$ of $q(\vec{c})$ w.r.t. $\mathcal{T}$.
Guess a variable mapping π for $r(\vec{c})$ in $DB_{\mathcal{A}\cup\mathcal{E}}$.
Check that $\langle \mathcal{T}, \mathcal{A} \cup \mathcal{E} \rangle$ is consistent.
Check that π is a match for $r(\vec{c})$ over $DB_{\mathcal{A}\cup\mathcal{E}}$.

Theorem 3. *For DL-Lite$_\mathcal{A}$,* EXIST, $\subseteq$-EXIST, *and* $\leq$-EXIST *are NP-complete. NP-hardness holds already for QAPs with an empty TBox and a CQ.*

Proof. Observe that existence of an explanation for a QAP $\mathcal{P}$ implies existence of a $\subseteq$-minimal and a $\leq$-minimal explanation for $\mathcal{P}$. Thus, we further consider EXIST only.

(MEMBERSHIP) The upper bound follows from guess-and-check Algorithm 1, which is immediate by Propositions 1 and 2. It guesses non-deterministically a candidate explanation $\mathcal{E}$, a rewriting $r(\vec{c})$ of $q(\vec{c})$ w.r.t. $\mathcal{T}$[2] and a candidate match for $r(\vec{c})$ in $DB_{\mathcal{A}\cup\mathcal{E}}$. The algorithm then checks in polynomial time that the ontology $\langle \mathcal{T}, \mathcal{A} \cup \mathcal{E} \rangle$ is consistent [Calvanese et al., 2009]. Finally, it checks that π is indeed a match for $r(\vec{c})$ witnessing $\vec{c} \in \mathsf{ans}(r, DB_{\mathcal{A}\cup\mathcal{E}})$ and thus $\vec{c} \in \mathsf{cert}(q, \langle \mathcal{T}, \mathcal{A} \cup \mathcal{E} \rangle)$.

(HARDNESS) In [Baader et al., 2010] the authors show CONP-completeness of $\mathcal{CQ}$-query emptiness, which is equivalent to checking non-existence of an explanation in QAPs with empty ABoxes. Hence, the lower bound carries over to EXIST. Their reduction uses a TBox and an acyclic conjunctive query. We show next that, quite unsurprisingly, NP-hardness of EXIST can be shown without a TBox, but using a non-empty ABox and (cyclic) conjunctive queries.

We reduce the NP-complete problem of deciding the existence of a homomorphism from a directed graph $G = (V, E)$ to a directed graph $G' = (V', E')$. Assume such a pair $\langle G, G' \rangle$ and build a QAP $\mathcal{P}_{G,G'} = \langle \emptyset, \mathcal{A}, q, \langle\rangle, \Sigma \rangle$ as follows. Let $\mathcal{A} = \{e(c_a, c_b) \mid (a, b) \in E'\}$ and $q = \{e(x_a, x_b) \mid (a, b) \in E\} \cup \{A(o)\}$ with o a constant not occurring in $\mathcal{A}$. Finally, let $\Sigma = \{A\}$.

We claim that there is a homomorphism from G to G' iff there is a solution to $\mathcal{P}_{G,G'}$. Indeed, if there is a homomorphism from G to G', then $\{A(o)\}$ is a solution to $\mathcal{P}$. For the other direction, assume there is an explanation $\mathcal{E}$ for $\mathcal{P}$. Since binary atoms are prohibited from occurring in $\mathcal{E}$ by the selection of Σ, there must exist a match π from q to $DB_\mathcal{A}$. Such a mapping π also witnesses the existence a homomorphism from G to G'. □

The NP-hardness of EXIST is caused by the restriction of the alphabet over which solutions can be found. In fact, this forbids us to explicitly encode the body of the query into the ABox, and forces us to search for a match. However, if the signature is not constrained, i.e., $\Sigma = N_C \cup N_R$, the problem can be solved in polynomial time. To see this, keep in mind that CQs, seen as FO formulae, are always satisfiable. Then an explanation does not exist only if the structure of the

[2]The procedure for guessing the derivation of a rewriting in non-deterministic polynomial time is presented in [Calvanese et al., 2009].

query is not compliant with the constraints expressed in the ontology. A naïve method to check whether a UCQ q is compliant with the ontological constraints is to iteratively go through all the CQs in q and instantiate them in the ABox, introducing fresh constants for the variables. If for none of the CQs we obtain a consistent ontology, then the query violates some of the constraints imposed at the conceptual level. However, we need to take into account that the introduced constants might not correspond to distinct individuals. Indeed, it can be proved that EXIST is equivalent to the PTIME-complete consistency problem for *DL-Lite*$_{\mathcal{A}}$ without the unique name assumption [Artale et al., 2009].

Theorem 4. *For DL-Lite*$_{\mathcal{A}}$ *and QAPs with unrestricted signatures,* EXIST*,* $\subseteq$-EXIST*, and* $\leq$-EXIST *are* PTIME*-complete.*

Proof. As noted in the proof of Theorem 3, it suffices to show the result for EXIST.

(MEMBERSHIP) Let $\Sigma = N_C \cup N_R$. Note that $\mathcal{P} = \langle \mathcal{T}, \mathcal{A}, q, \vec{c}, \Sigma \rangle$, with q a UCQ, has an explanation iff $\mathcal{P}_{q'} = \langle \mathcal{T}, \mathcal{A}, q', \vec{c}, \Sigma \rangle$ has an explanation for some $q' \in q$. Hence, it suffices to show the upper bound for CQs. To this end, we provide a logarithmic space reduction to consistency in *DL-Lite*$_{\mathcal{A}}$ without UNA. Assume a QAP $\mathcal{P} = \langle \mathcal{T}, \mathcal{A}, q, \vec{c}, \Sigma \rangle$ with q a CQ. Take an ontology $\mathcal{O} = \langle \mathcal{T} \cup \mathcal{T}', \mathcal{A} \cup \mathcal{E}_q \cup \mathcal{A}' \rangle$ as follows. The ABox $\mathcal{E}_q$ is obtained from $at(q(\vec{c}))$ by replacing each variable x with a fresh constant name a_x. The ABox $\mathcal{A}'$ consists of assertions $A_o(o)$ for all constants o occurring in $\mathcal{A} \cup \mathcal{E}_q$, where each A_o is a fresh concept name. The TBox $\mathcal{T}'$ consists of axioms $A_o \sqsubseteq \neg A_{o'}$ for all pairs $o \neq o'$ of constants occurring in $\mathcal{A}$ and $q(\vec{c})$. We now show that $\mathcal{P}$ has a solution iff $\mathcal{O}$ is consistent.

Assume that $\mathcal{P}$ has an explanation $\mathcal{E}$. Then, due to consistency of $\mathcal{O}' = \langle \mathcal{T}, \mathcal{A} \cup \mathcal{E} \rangle$, there is a model $\mathcal{I}$ of $\mathcal{O}'$ under the UNA. Additionally, $\mathcal{I}$ admits a match π for $q(\vec{c})$. Let $\mathcal{I}'$ be the extension of $\mathcal{I}$ that additionally interprets *(i)* constants in $\mathcal{E}_q$ as $a_x^{\mathcal{I}'} = \pi(x)$ for each variable x in q, and *(ii)* $A_o^{\mathcal{I}'} = \{o^{\mathcal{I}}\}$ for each freshly introduced A_o. It remains to show that $\mathcal{I}'$ is a model of $\mathcal{O}$. Observe that since $\mathcal{I}$ is under the UNA, we have that $A_o^{\mathcal{I}'} \cap A_{o'}^{\mathcal{I}'} = \emptyset$, for all constant pairs $o \neq o'$. Thus $\mathcal{I}'$ satisfies all the disjointness axioms $A_o \sqsubseteq \neg A_{o'}$ in $\mathcal{T}'$. The assertions in $\mathcal{A}'$ are satisfied due to *(ii)*, and the assertions in $\mathcal{E}_q$ due to *(i)* above.

For the other direction, observe that given any model $\mathcal{I}$ of $\mathcal{O}$, an explanation for $\mathcal{P}$ can be easily extraced from $\mathcal{I}$.

(HARDNESS) We sketch a reduction from consistency in *DL-Lite*$_{\mathcal{A}}$ without UNA. Given $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$, simply create a QAP $\mathcal{P} = \langle \mathcal{T}, \emptyset, q, \langle\rangle, N_C \cup N_R \rangle$, where q is obtained from $\mathcal{A}$ by replacing each constant a in $\mathcal{A}$ by a variable name x_a. It is easy to see that $\mathcal{O}$ is consistent under UNA iff $\mathcal{P}$ has an explanation. □

4.2 Deciding Necessity

The existence of an explanation is most of the times not very informative to the user. In fact, given a negative answer to a query, it is important to delineate the fundamental reasons leading to the absence of the expected tuple. That is, users would like to know which assertions necessarily occur in all the explanations for a QAP $\mathcal{P}$.

Theorem 5. *For DL-Lite$_\mathcal{A}$,* NEC *and* $\subseteq$-NEC *are* CONP*-complete. Moreover, for QAPs with unrestricted explanation signature,* NEC *and* $\subseteq$-NEC *are* PTIME*-complete.*

Proof. We consider NEC first.

(MEMBERSHIP) We provide a reduction to non-existence of an explanation, which is CONP-complete in general, but PTIME-complete for QAPs with unrestricted explanation signature (see Theorems 3 and 4). Assume $\mathcal{P} = \langle \mathcal{T}, \mathcal{A}, q, \vec{c}, \Sigma \rangle$ and an assertion $\varphi(\vec{t})$. Take a new QAP $\mathcal{P} = \langle \mathcal{T}', \mathcal{A}', q, \vec{c}, \Sigma \rangle$, where $\mathcal{A}' = \mathcal{A} \cup \{\bar{\varphi}(\vec{t})\}$ and $\mathcal{T}' = \mathcal{T} \cup \{\bar{\varphi} \sqsubseteq \neg\varphi\}$, where $\bar{\varphi}$ is a fresh predicate name. Intuitively, the models of $\langle \mathcal{T}', \mathcal{A}' \rangle$ are exactly the models of $\langle \mathcal{T}, \mathcal{A} \rangle$ in which $\varphi(\vec{t})$ is false. The following holds: $\varphi(\vec{t})$ occurs in all explanations for $\mathcal{P}$ iff $\mathcal{P}'$ has no explanation. For the "only if" direction, assume $\varphi(\vec{t})$ occurs in all explanations for $\mathcal{P}$, and $\mathcal{P}'$ has an explanation $\mathcal{E}$. Clearly, $\mathcal{E}$ is an explanation for $\mathcal{P}$ and $\varphi(\vec{t}) \notin \mathcal{E}$, which contradicts the assumption. For the other direction, observe that an explanation $\mathcal{E}$ with $\varphi(\vec{t}) \notin \mathcal{E}$ is also an explanation for $\mathcal{P}'$.

(HARDNESS) We provide a reduction from checking non-existence of an explanation. Assume a QAP $\mathcal{P} = \langle \mathcal{T}, \mathcal{A}, q, \vec{c}, \Sigma \rangle$ and take $A(o)$, where A and o do not occur in $\mathcal{P}$. The following holds: $\mathcal{P}$ has no explanation iff $A(o)$ occurs in all explanations for $\mathcal{P}$. The "only if" direction is immediate. For the "if" direction, note that by Proposition 2, if $\mathcal{P}$ admits an explanation $\mathcal{E}$, then $\mathcal{P}$ also admits an explanation $\mathcal{E}'$ restricted to the signature of $\mathcal{P}$, i.e., with $A(o) \notin \mathcal{E}'$.

To obtain the completeness result for $\subseteq$-NEC, observe that an assertion α occurs in all $\subseteq$-minimal explanations for a QAP $\mathcal{P}$ iff α occurs in all explanations for $\mathcal{P}$. Thus, NEC and $\subseteq$-NEC are equivalent. □

Now, we consider necessity under the minimum explanation size order and we show that under common assumptions the problem is harder than NEC. Intuitively, this is because one has to compute first the minimal size of an explanation and, then, inspect all the explanations of that size.

Theorem 6. *For DL-Lite$_\mathcal{A}$,* $\leq$-NEC *is* $\mathsf{P}^{\mathsf{NP}}_{\parallel}$*-complete.* $\mathsf{P}^{\mathsf{NP}}_{\parallel}$*-hardness holds already for QAPs with an empty TBox, a CQ, and an unrestricted explanation signature.*

Proof. (MEMBERSHIP) Let's assume a QAP $\mathcal{P} = \langle \mathcal{T}, \mathcal{A}, q, \vec{c}, \Sigma \rangle$ and an assertion α. From Proposition 2, we know that if $\mathcal{P}$ has an explanation, then there exists an explanation, whose size m is polynomial in the number of terms contained in q. Observe that $\langle \mathcal{P}, \alpha \rangle$ is a negative instance of $\leq$-NEC iff there is an $i \in [0..m]$ such that *(a)* $\mathcal{P}$ has an explanation $\mathcal{E}$ with $|\mathcal{E}| = i$ and $\alpha \notin \mathcal{E}$, and *(b)* $\mathcal{E}$ is $\leq$-minimal. Thus, we use an auxiliary problem SIZE-OUT, which is to decide given a tuple $\langle \mathcal{P}', \alpha', n' \rangle$, where $\mathcal{P}'$ is a QAP, α' is an assertion, and n' is an integer, whether there exists an explanation $\mathcal{E}'$ for $\mathcal{P}'$ such that $|\mathcal{E}'| = n'$ and $\alpha' \notin \mathcal{E}'$. Furthermore, the problem NO-SMALLER is to decide, given a tuple $\langle \mathcal{P}', n' \rangle$ of a QAP and an integer, whether there is no explanation $\mathcal{E}'$ for $\mathcal{P}'$ such that $|\mathcal{E}'| < n'$. Observe that SIZE-OUT is in NP, while NO-SMALLER is in CONP. Take the tuple $S = \langle A_0, B_0, \ldots, A_m, B_m \rangle$, where $A_i = \langle \mathcal{P}, \alpha, i \rangle$ and $B_i = \langle \mathcal{P}, i \rangle$, for all $i \in [0..m]$. Due to the above observation, α occurs in all $\leq$-minimal explanations $\mathcal{E}$ for $\mathcal{P}$ iff for all $i \in [0..m]$, one of the following holds: *(i)* A_i is a negative instance of SIZE-OUT, or *(ii)* B_i is a negative instance of NO-SMALLER. Note that S can be built in polynomial time in the size of the input, while the positivity of the instances in S can be decided by making $2m$ parallel calls to an NP oracle. Thus we obtain membership in $\mathsf{P}^{\mathsf{NP}}_{\parallel}$.

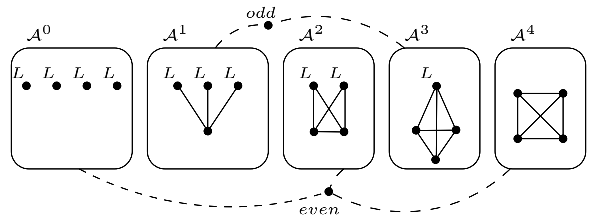


Figure 1: The structure of $\mathcal{A}_G$ for a graph G with m vertices. Solid arcs in $\mathcal{A}^\ell$ represent assertions $Egde(a, b)$ in $\mathcal{A}^\ell$ introduced in (b). A dashed arc from an ABox $\mathcal{A}^\ell$ to the constant $par(\ell)$ represents the collection of assertions that relate each constant in $\mathcal{A}^\ell$ to $par(\ell)$ via the role P.

(HARDNESS) We give a reduction from ODDMINVERTEXCOVER, which is $\mathsf{P}_{\|}^{\mathsf{NP}}$-complete [Wagner, 1987]. An instance of this problem is given by a graph $G = (V, E)$, and we are asked whether the least cardinality over all vertex covers in G is odd. That is, is there an odd integer $k \in [1..|V|]$ such that G has a vertex cover C with $|C| = k$, and there is no vertex cover C' in G with $|C'| < k$?

In the reduction we exploit the following property. Given an integer k and an undirected graph $G = (V, E)$ with m vertices, construct a new graph $G' = ([1..m], E')$ such that there is an edge between each $i \in [1..k]$ and $j \in [1..m]$. The following holds: if there is an injective homomorphism h from G to G', then G has a vertex cover of size k. Indeed, take $C = \{v \in V \mid h(v) \leq k\}$. Due to injectivity, $|C| = k$. Assume an arbitrary edge $\{v_1, v_2\} \in E$. Since h is a homomorphism, due to the selection of edges we must have $h(v_1) \leq k$ or $h(v_2) \leq k$. Then $\{v_1, v_2\} \cap C \neq \emptyset$ by the selection of C.

Assume an arbitrary graph $G = (V, E)$ with vertices $V = \{v_1, \ldots, v_m\}$. W.l.o.g., G is connected, directed, and has at least 2 nodes. We construct next a QAP $\mathcal{P}_G = \langle \emptyset, \mathcal{A}_G, q_G, \langle\rangle, \Sigma_G \rangle$ and an atom α_G such that G is a positive instance of ODDMINVERTEXCOVER iff α_G is $\leq$-necessary for $\mathcal{P}_G$. In the reduction we use constants odd, $even$, c^i_j, where $i, j \in [0..m]$, concept names M, L, and roles P, $\neq$, $Edge$.

Let q_G be the Boolean query consisting of atoms

(i) $Edge(x_{i_1}, x_{i_2})$, for each edge $(v_{i_1}, v_{i_2}) \in E$,

(ii) $\neq(x_{i_1}, x_{i_2})$, for each $i_1, i_2 \in [1..m]$, $i_1 \neq i_2$, and

(iii) $L(x_1), \ldots, L(x_m)$ and $P(x_1, y)$, $M(y)$.

Intuitively, in *(i)* we represent the graph G in the query. We will use atoms in *(ii)* to ensure that different variables are mapped to distinct elements. The atoms $L(x_i)$ will be used to measure the size of vertex covers, while the atoms $P(x_1, y)$ and $M(y)$ will be used to determine their parity. We allow explanations only over concept names, and thus set $\Sigma_G = \{M, L\}$.

To define $\mathcal{A}_G$, we first construct a collection $\mathcal{A}^0, \ldots, \mathcal{A}^m$ of ABoxes, where each $\mathcal{A}^j$ consists of the assertions

(a) $L(c^j_i)$, for $i \in [j..m]$,

(b) $Edge(c^j_{i_1}, c^j_{i_2})$, for all $i_1, i_2 \in [1..m]$ with $i_1 \leq j$ or $i_2 \leq j$, and

(c) $\neq(c^j_{i_1}, c^j_{i_2})$, for all $i_1, i_2 \in [1..m]$ with $i_1 \neq i_2$.

For an integer k, let $par(k) = odd$ if k is odd, and $par(k) = even$, otherwise. Let $\mathcal{A}' = \{P(c^j_i, par(j)) \mid i, j \in [0..m]\}$. Then $\mathcal{A}_G = \mathcal{A}^0 \cup \cdots \cup \mathcal{A}^m \cup \mathcal{A}'$. See Figure 1 for an example.

Finally, we let $\alpha_G = M(odd)$. To prove the correctness of the reduction, we define $up(k) = \{L(c^k_1), \dots, L(c^k_k), M(par(k))\}$, and claim the following:

CLAIM 1: *If C is a vertex cover in G of size k, then $up(k)$ is an explanation for $\mathcal{P}_G$.* Let $\mathcal{A}^* = \mathcal{A}_G \cup up(k)$. Since we have no TBox axioms, it suffices to show the existence of a match π for q_G in $DB_{\mathcal{A}^*}$. Take an enumeration $z_1, \dots, z_m$ of variables $x_1, \dots, x_m$ such that $\{z_1, \dots, z_k\} = \{x_i \mid v_i \in C\}$. Take the mapping π such that $\pi(z_i) = c^k_i$ for all $i \in [1..m]$, and $\pi(y) = par(k)$. Assume an atom $Edge(x_{i_1}, x_{i_2})$ in q_G. Due to (b) in the definition of $\mathcal{A}^j$, it suffices to show that $\pi(x_{i_1}) = c^k_\ell$ or $\pi(x_{i_2}) = c^k_\ell$ for some $\ell \leq k$. Indeed, since C is a vertex cover, $v_{i_1} \in C$ or $v_{i_2} \in C$. Then due to the enumeration of variables, $x_{i_1} = z_\ell$ or $x_{i_2} = z_\ell$ for some $\ell \leq k$. Due to the definition of π, $\pi(x_{i_1}) = c^k_\ell$ or $\pi(x_{i_2}) = c^k_\ell$ for $\ell \leq k$. The atoms $\neq(x_{i_1}, x_{i_2})$ in q_G are properly mapped due to (c) in the construction of $\mathcal{A}^j$ and the fact that π is injective by construction. For an atom $L(x_i)$ in q_G we have two options. If $\pi(x_i) = c^k_\ell$ with $\ell \leq k$, then $L(c^k_\ell) \in up(k)$ by the definition of $up(k)$. Otherwise, if $\ell > k$, then $L(c^k_\ell) \in \mathcal{A}^k$ by the definition of $\mathcal{A}^k$. The atom $P(\pi(x_1), \pi(y))$ belongs to $\mathcal{A}^*$ due to the definition of $\mathcal{A}'$, while $M(\pi(y)) \in up(k)$ by construction of $up(k)$.

CLAIM 2: *Assume $up(k)$ is an explanation for $\mathcal{P}_G$. Then G has a vertex cover of size k.* Let $\mathcal{A}^* = \mathcal{A}_G \cup up(k)$ and let π be a match for q_G in $DB_{\mathcal{A}^*}$. Observe that due irreflexivity of the role $\neq$ and the atoms (ii) in q_G, π must be injective. Observe also that for all $\ell \in [1..m]$, where $\ell \neq k$, we have $|\{c^\ell_i \mid L(c^\ell_i) \in \mathcal{A}^\ell\}| < m$. Due to the connectedness of G and atoms $L(x_1), \dots, L(x_m)$ in q_G, π must use only the atoms in $\mathcal{A}^k \cup \mathcal{A}' \cup up(k)$. That is, π is also a match for q_G in $DB_{\mathcal{A}^k \cup \mathcal{A}' \cup up(k)}$. Let $C = \{v_i \in V \mid \pi(x_i) = c^k_n, n \in [1..k]\}$. Then $|C| = k$ due to the injectivity of π. To see that C is a vertex cover, assume an edge $(v_{i_1}, v_{i_2}) \in E$. By construction, q_G has the atom $Edge(x_{i_1}, x_{i_2})$. Since π is a match in $DB_{\mathcal{A}^k \cup \mathcal{A}' \cup up(k)}$, $Edge(\pi(x_{i_1}), \pi(x_{i_2})) \in \mathcal{A}^k$. Then, by construction of $\mathcal{A}^k$, we have $\pi(x_{i_1}) = c^k_n$ or $\pi(x_{i_2}) = c^k_n$ with $n \leq k$. Then by the selection of C, $\{\pi(x_{i_1}), \pi(x_{i_2})\} \cap C \neq \emptyset$.

CLAIM 3: *Assume $\mathcal{E}$ is a $\leq$-minimal explanation for $\mathcal{P}_G$ with size k. Then $\mathcal{E} = up(k-1)$.* Since G is connected and $\mathcal{E}$ is $\leq$-minimal, there exist an index $\ell \in [1..m]$ such that $\mathcal{E} \subseteq \{L(c^\ell_1), \dots, L(c^\ell_m), M(par(\ell))\}$ and there is a match for q_G in $\mathcal{A}^\ell \cup \mathcal{A}' \cup \mathcal{E}$. Since $L(c^\ell_i) \in \mathcal{A}^\ell$ for $i \in [\ell{+}1..m]$ by the definition of $\mathcal{A}^\ell$, we have by cardinality minimality that $\mathcal{E} \subseteq \{L(c^\ell_1), \dots, L(c^\ell_\ell), M(par(\ell))\}$. By the definition of $\mathcal{A}^\ell$, $|\{c^\ell_i \mid L(c^\ell_i) \in \mathcal{A}^\ell\}| = m - \ell$. Thus, due to the injectivity of any match π for q_G, we must have $|\{c^\ell_i \mid L(c^\ell_i) \in \mathcal{E}\}| \geq \ell$. Hence, $\mathcal{E} = \{L(c^\ell_1), \dots, L(c^\ell_\ell), M(par(\ell))\} = up(\ell)$. Since $|\mathcal{E}| = k$, we have $\ell = k - 1$.

We can now finalize the correctness proof:

"$\Rightarrow$" Suppose there exists an odd integer $k \in [1..|V|]$ such that G has a vertex cover C with $|C| = k$, and there is no vertex cover C' in G with $|C'| < k$. By CLAIM 1, $up(k)$ is an explanation for $\mathcal{P}_G$. We make sure that $up(k)$ is $\leq$-minimal. Suppose there exists an explanation $\mathcal{E}'$ with size $|\mathcal{E}'| < |up(k)|$, i.e., $|\mathcal{E}'| = \ell$ for some $\ell \leq k$. We can assume that $\mathcal{E}'$ is $\leq$-minimal. Then by CLAIM 3, $\mathcal{E}' = up(\ell - 1)$. It follows from CLAIM 2 that G has a vertex cover of size $\ell - 1$. Since $\ell - 1 < k$, we arrive at a contradiction to the assumption that G has no vertex cover of size $< k$. Thus $up(k)$ is $\leq$-minimal. Since k is odd, we have $M(odd) \in up(k)$. By CLAIM 3, apart from $up(k)$ there is no other $\leq$-minimal explanation for $\mathcal{P}_G$. That is, $M(odd)$ occurs in all $\leq$-minimal explanations for $\mathcal{P}_G$.

"$\Leftarrow$" Assume $M(odd)$ occurs in all $\leq$-minimal explanations for $\mathcal{P}_G$. By CLAIM 3, we know that $up(k)$ is the unique $\leq$-minimal explanation, for some integer k. Since $M(odd) \in up(k)$, we get that k is odd. Then, by CLAIM 2, there is a vertex cover C with size k. It remains to ensure that there is no vertex cover C' of size $\ell < k$. Assume the opposite. Then by CLAIM 1 we have that $up(\ell)$ is an explanation with size $|up(\ell)| < |up(k)|$, which contradicts the assumption that $up(k)$ is $\leq$-minimal. Thus G is a positive instance of ODDMINVERTEXCOVER.

The definition of Σ_G prohibits binary atoms from occurring in $\leq$-minimal explanations. The same effect can be achieved using $\Sigma_G = N_C \cup N_R$ and modifying $\mathcal{A}_G$ and q_G to make it prohibitively expensive to have binary atoms in $\leq$-minimal explanations. Simply replace each binary assertion $r(c, d)$ in $\mathcal{A}_G$ by fresh assertions $r_1(c, d), \ldots, r_{m+2}(c, d)$, and each binary $r(x, y)$ in q_G by $r_1(x, y), \ldots, r_{m+2}(x, y)$. In this way the lower-bound can be shown for unrestricted explanation signatures. □

4.3 Deciding Relevance

A domain user faced with a negative answer to a query may ask herself whether the absence of a certain ABox assertion α in the ontology is related with the lack of the tuple in the results. That is, she would like to know whether α occurs in some explanation to QAP $\mathcal{P}$. First, we consider REL.

Theorem 7. *For DL-Lite$_\mathcal{A}$, REL is NP-complete. NP-hardness holds already for QAPs with an empty TBox and a CQ.*

Proof. (MEMBERSHIP) This part can be shown by a reduction to EXIST, which we have shown previously to be NP-complete. For a given QAP $\mathcal{P}$ and an assertion α, construct a new QAP $\mathcal{P}'$ from $\mathcal{P}$ by extending the ABox in $\mathcal{P}$ with the assertion α. Intuitively, we restrict the search to those explanations which do not contradict α. Clearly, α is relevant for $\mathcal{P}$ iff an explanation for $\mathcal{P}'$ exists.

(HARDNESS) We can again use a reduction and an argument almost identical to the one in the proof of Theorem 3. For a pair G, G' of directed graphs, let $\mathcal{P}_{G,G'}$ be defined as in the proof of Theorem 3, and let $\alpha = A(o)$. Then there is a homomorphism from G to G' iff α is relevant for $\mathcal{P}_{G,G'}$. □

We now turn our attention to $\subseteq$-REL.

Theorem 8. *For DL-Lite$_\mathcal{A}$, $\subseteq$-REL is Σ_2^P-complete. Σ_2^P-hardness holds already for (i) QAPs with an empty TBox and a CQ, and (ii) QAPs with an empty TBox, a UCQ, and an unrestricted explanation signature.*

Proof. (MEMBERSHIP) Let $\mathcal{P} = \langle \mathcal{T}, \mathcal{A}, q, \vec{c}, \Sigma \rangle$ be a QAP and let α be an ABox assertion. We now provide an extended version of the algorithm solving existence (see Algorithm 1), which in turn decides whether α is $\subseteq$-relevant for $\mathcal{P}$. First of all, an ABox $\mathcal{E}$ containing α is non-deterministically guessed. Let HAS-SUBEXPL solve the problem of deciding whether an explanation $\mathcal{E}$ has a subset which is itself an explanation. The problem can be easily proved to be in NP. Then, the algorithm checks the complement of HAS-SUBEXPL in order to assure that none of the subsets of $\mathcal{E}$ is itself an explanation, from which it follows that α is $\subseteq$-relevant. Checking the complement of HAS-SUBEXPL can be done in coNP. For this reason, the problem is solvable in non-deterministic polynomial time by a TM with an NP oracle.

(HARDNESS) We make use of a reduction of the Σ_2^{P}-complete problem co-CERT3COL [Stewart, 1991] (see also [Bonatti, Lutz, and Wolter, 2009]). An instance of co-CERT3COL is given by a graph $G = (V, E)$ with vertices $V = \{1, \ldots, n\}$ such that every edge is labeled with a disjunction of two literals over the Boolean variables $\{p_{(i,j)} \mid i, j \in [1..n]\}$. G is a positive instance if there is a truth value assignment t to the Boolean variables such that the graph $t(G)$ obtained from G by including only those edges whose label evaluates to true under t is not 3-colorable. Assume an instance G of co-CERT3COL. We show how to build in polynomial time a QAP $\mathcal{P}_G = \langle \mathcal{T}_G, \mathcal{A}_G, q_G, \vec{c}_G, \Sigma_G \rangle$ and an ABox assertion α_G such that: G is a positive instance of co-CERT3COL iff α_G is $\subseteq$-relevant for $\mathcal{P}_G$. We use an empty TBox and a Boolean query, thus $\mathcal{T}_G = \emptyset$ and $\vec{c}_G = \langle\rangle$. Let $\mathcal{B} = \{\mathbf{t}, \mathbf{f}\}$ denote the set of truth values. The query q_G has the following atoms for each edge $e = (i, j)$ in E: *(a)* $B(x_i)$, $R_e(x_i, y_e)$, $R_e(y_e, x_j)$, $B(x_j)$, and *(b)* $P(y_e, z_{p_i})$, $A_{p_i}(z_{p_i})$, $W_{p_i}(z_{p_i}, z'_{p_i})$, where $p_i \in \{p_1, p_2\}$ and p_1, p_2 are the first and the second proposition in the labeling of e, respectively. The query q_G simply incorporates G together with the disjunctions on the edges. Observe that if two edges have the same proposition in their label, then this will be reflected in q_G by some shared variables z_{p_i}.

To build $\mathcal{A}_G$ we use constants c_p and $c_{\neg p}$ for the truth value of proposition p. Intuitively, the truth value of p will be determined by either $A_p(c_p)$ or $A_p(c_{\neg p})$ being in the explanation. Assume a tuple $\vec{t} = \langle e, v_1, v_2, a, b \rangle$, where $e \in E$, $\{v_1, v_2\} \subseteq \mathcal{B}$, and a, b are constants. Let p_1, p_2 be the first and the second propositions of e. For $i \in \{1, 2\}$ and $v_i = \mathbf{t}$, let $l_i = p_i$ if p_i is positive and $l_i = \neg p_i$ otherwise. Similarly, for $i \in \{1, 2\}$ and $v_i = \mathbf{f}$, let $l_i = \neg p_i$ if p_i is positive and $l_i = p_i$ otherwise. Then, the ABox $\mathcal{A}(\vec{t})$ consists of the assertions $R_e(a, d_T)$, $R_e(d_T, b)$, $P(d_T, c_{l_1})$ and $P(d_T, c_{l_2})$ depending on the Boolean values in input.

The ABox $\mathcal{A}_G$ is the union of the following ABoxes:

(A1) $\mathcal{A}(\langle e, v, v', a_i, a_j \rangle)$ for all $e \in E$, $v, v' \in \mathcal{B}$, $i, j \in [0..2]$, and $i \neq j$;
(A2) $\mathcal{A}(\langle e, \mathbf{f}, \mathbf{f}, a_i, a_i \rangle)$ for all $e \in E$, and $i \in [0..2]$;
(A3) $\mathcal{A}(\langle e, v, v', b, b \rangle)$ for all $e \in E$, $v, v' \in \mathcal{B}$;
(A4) The ABox $\{B(a_0), B(a_1), B(a_2)\}$;
(A5) The assertions $W_p(c_p, c_{\neg p})$ and $W_p(c_{\neg p}, c_p)$ for all propositions.

Let $\alpha_G = B(b)$ and $\Sigma_G = \{A_p \mid A_p \in q_G\} \cup \{B\}$. It is not too difficult to see that G is a positive instance of co-CERT3COL iff there exists an $\subseteq$-explanation $\mathcal{E}$ to $\mathcal{P}$ such that $\alpha_G \in \mathcal{E}$. Basically, definitions (A1)–(A3) encode a triangular structure T in which edges in G that evaluate to false according to a given truth assignment can be mapped on any edge of T, reflexive edges included. If an edge of G evaluates to true, then it must be mapped to one of the non-reflexive edges. This ensures that if G can be mapped to T under truth assignment t, then $t(G)$ is 3-colorable. Instead, definitions (A4)–(A5) define a cyclic structure C into which any graph G can be embedded. It has to be noted that the node b is not asserted to be a member of B, hence q_G cannot be mapped there directly with any truth assignment. We see this more formally next:

"$\Rightarrow$" Suppose there is a truth assignment t such that $t(G)$ is not 3-colorable. Let $\mathcal{E} = \{B(b)\} \cup \mathcal{E}_1$, where $\mathcal{E}_1 = \{A_p(c_p) \mid t(p) = \mathbf{t}\} \cup \{A_p(c_{\neg p}) \mid t(p) = \mathbf{f}\}$. It remains to argue that $\mathcal{E}$ is a $\subseteq$-explanation to $\mathcal{P}$. It is not hard to see that $\mathcal{E}$ is an explanation. Indeed q_G matches already in the

ABox obtained by point (A3) (hint: since $B(b) \in \mathcal{E}$, we match q_G by mapping all variables of q_G to (interpretation of) b). Suppose there is a smaller explanation $\mathcal{E}' \subset \mathcal{E}$. Observe that $\mathcal{E}_1 \subseteq \mathcal{E}'$. This is because for all propositions p, the symbol A_p does not occur in $\mathcal{A}_G$ but does occur in q_G. Then, $\mathcal{E} \setminus \{B(b)\}$ must be an explanation. If this is the case, then q_G can be matched in $\mathcal{A}_G$ without the ABox from (A3), i.e., in the triangle part. Then $t(G)$ is 3-colorable, which contradicts the assumption.

"$\Leftarrow$" Let $\mathcal{E}$ be a $\subseteq$-minimal explanation containing $B(b)$. Due to the signature restriction, the predicate W_p cannot occur in $\mathcal{E}$ for any proposition p. Since $\mathcal{E}$ is an explanation, by the definition of q' and (A5) we have that $A_p(c_p) \in \mathcal{E}$ or $A_p(c_{\neg p}) \in \mathcal{E}$ for all propositions p. Since for any proposition p we have that A_p occurs in q_G with one and only variable z_p, we know that exactly one of $A_p(c_p) \in \mathcal{E}$ and $A_p(c_{\neg p}) \in \mathcal{E}$ holds. Due to the atoms $W_p(z_p, z'_p)$ in q_G, we also have that constants of the form c_p and $c_{\neg p}$ are the only ones that can get an A_p label. Consider the assignment t defined as follows: $t(e) = \mathbf{t}$ if $A_p(c_p) \in \mathcal{E}$, and $t(e) = \mathbf{f}$ if $A_p(c_{\neg p}) \in \mathcal{E}$. It is not difficult to argue that $t(G)$ is not 3-colorable and thus G is a positive instance of co-CERT3COL. Indeed, if $t(G)$ was 3-colorable, Q should be mappable into the triangle part obtained in (A1)–(A3). Then $\mathcal{E} \setminus \{B(b)\}$ would be a smaller explanation, which would mean a contradiction.

In [Calvanese et al., 2011] we have proven that this lower bound holds also for QAPs with a UCQ and unrestricted explanation signature. □

Unsurprisingly, $\leq$-REL has the same complexity as $\leq$-NEC. Indeed, the two problems share the same source of complexity, namely the need to inspect all explanations up to a computed size, which allows us to reduce the ODDMINVERTEXCOVER problem. In fact, $\mathrm{P}_{\|}^{\mathrm{NP}}$-hardness can be shown using the same reduction as in the proof of Theorem 6. A matching upper bound can also be obtained by slightly modifying the algorithm for $\leq$-NEC.

Theorem 9. *For DL-Lite$_\mathcal{A}$, $\leq$-REL is $\mathrm{P}_{\|}^{\mathrm{NP}}$-complete. $\mathrm{P}_{\|}^{\mathrm{NP}}$-hardness holds already for QAPs with an empty TBox, a CQ, and an unrestricted explanation signature.*

Proof. (MEMBERSHIP) $\leq$-REL can be tackled in a way similar to $\leq$-NEC. In fact, the algorithm described in Theorem 6 can be modified in order to solve this problem. Let SIZE-IN solve the following problem: given a tuple $\langle \mathcal{P}, \alpha, n\rangle$, where $\mathcal{P}$ is a QAP, α an assertion, and n an integer, decide whether there exists an explanation $\mathcal{E}$, with $|\mathcal{E}| = n$ and $\alpha \in \mathcal{E}$. Then, we change the positivity condition of the $\leq$-NEC algorithm as follows: *α occurs in some $\leq$-minimal explanations $\mathcal{E}$ for $\mathcal{P}$ iff for some $i \in [0..m]$ it holds that: (i) A_i is a positive instance of* SIZE-IN, *and (ii) B_i is a positive instance of* NO-SMALLER. It is easy to see that SIZE-IN is solvable in NP, hence the whole problem is again in $\mathrm{P}_{\|}^{\mathrm{NP}}$.

(HARDNESS) Recall the reduction from ODDMINVERTEXCOVER to $\leq$-NEC in the proof of Theorem 6. We argue that exactly the same reduction also shows $\mathrm{P}_{\|}^{\mathrm{NP}}$-hardness of $\leq$-REL. Assume a directed graph G and let $\mathcal{P}_G$ and α_G be the QAP and the assertion resulting in the reduction. To prove the claim is suffices to show the following equivalence: α_G is $\leq$-necessary for $\mathcal{P}_G$ iff α_G is $\leq$-relevant for $\mathcal{P}_G$. This equivalence follows directly from CLAIM 3, which states that $\mathcal{P}_G$ has a unique $\leq$-minimal explanation. □

4.4 Recognizing Explanations

In order to decide REC, we need to check consistency of the explanation with the ontology, and check whether the tuple is in the certain answer to the query. The former is polynomial and the latter in NP, therefore REC is in NP.

Theorem 10. *For DL-Lite$_\mathcal{A}$,* REC *is NP-complete. NP-hardness holds already for QAPs with an empty TBox, a CQ, and an unrestricted explanation signature.*

Proof. (MEMBERSHIP) Given a QAP $\mathcal{P} = \langle \mathcal{T}, \mathcal{A}, q, \vec{c}, \Sigma \rangle$ and an ABox $\mathcal{E}$, we devise an algorithm deciding REC as follows. Firstly, the procedure checks that $N_C(\mathcal{E}) \cup N_R(\mathcal{E}) \subseteq \Sigma$. Then, it makes sure that extending the ontology with ABox $\mathcal{E}$ does not lead to an inconsistent theory. Then, it non-deterministically guesses a rewriting r_j of q w.r.t. $\mathcal{T}$ and a match π for r_j in $DB_{\mathcal{A}\cup\mathcal{E}}$. Finally, the algorithm checks that π is a match witnessing $\vec{c}$ in $DB_{\mathcal{A}\cup\mathcal{E}}$.

(HARDNESS) This part can be shown by a reduction and an argument that is almost identical to the one in the proof of Theorem 3. For a pair G, G' of directed graphs, let $\mathcal{P}_{G,G'}$ be defined as in the proof of Theorem 3 except that $\Sigma = N_C \cup N_R$, and let $\mathcal{E} = \{A(o)\}$. Then there is a homomorphism from G to G' iff $\mathcal{E}$ is an explanation for $\mathcal{P}_{G,G'}$. □

In case a preference order is in place, to recognize an explanation one has to check minimality as well. This check is coNP-hard for $\subseteq$- and $\leq$-minimality, leading to completeness for DP.

Theorem 11. *For DL-Lite$_\mathcal{A}$, $\leq$-*REC *and $\subseteq$-*REC *are DP-complete. DP-hardness holds already for QAPs with an empty TBox, a CQ, and an unrestricted explanation signature.*

Proof. (MEMBERSHIP) Membership of a problem Π in DP can be shown by providing two languages $L_1 \in$ NP and $L_2 \in$ coNP, such that the set of all yes-instances of Π is $L_1 \cap L_2$. For $\leq$-REC, simply let $L_1 = \{(\mathcal{P}, \mathcal{E}) \mid \mathcal{E}$ is an explanation for $\mathcal{P}\}$ and $L_2 = \{(\mathcal{P}, \mathcal{E}) \mid \mathcal{P}$ has no explanation $\mathcal{E}'$ s.t. $|\mathcal{E}'| \leq |\mathcal{E}|\}$. For $\subseteq$-REC, we take L_1 as above and $L_2 = \{(\mathcal{P}, \mathcal{E}) | \mathcal{P}$ does not have an explanation $\mathcal{E}'$ s.t. $\mathcal{E}' \subset \mathcal{E}\}$.

(HARDNESS) DP-hardness is shown by a reduction from the problem HP-NOHP. An instance of HP-NOHP is given by two directed graphs $G = (V, E)$ and $G' = (V', E')$, where $\langle G, G' \rangle$ is a positive instance iff G has an Hamilton path and G' does not have one. For such a pair $\langle G, G' \rangle$, we define a QAP $\mathcal{P} = \langle \emptyset, \mathcal{A}, q, \langle\rangle, \Sigma \rangle$ and a set $\mathcal{E}$ such that:

(a) $\langle G, G' \rangle$ is a positive instance of HP-NOHP iff $\mathcal{E}$ is a $\leq$-minimal explanation for $\mathcal{P}$, and
(b) $\langle G, G' \rangle$ is a positive instance HP-NOHP iff $\mathcal{E}$ is a $\subseteq$-minimal explanation for $\mathcal{P}$.

W.l.o.g., nodes in G and G' are disjoint and are ordinary constants. Construct an ABox $\mathcal{A}_G = \{e(v_i, v_j) \mid (v_i, v_j) \in E\} \cup \{d(v_i, v_j) \mid v_i, v_j \in V, v_i \neq v_i\}$. Intuitively, an assertion $e(v_i, v_j)$ encodes an edge (v_i, v_j) in the graph G, whereas an assertion $d(v_i, v_j)$ encodes that nodes v_i and v_j are distinct. The ABox $\mathcal{A}_{G'}$ encodes G' in a similar way as before, using roles e' instead of e, and in addition it has an assertion $A(v'_i)$ for each $v'_i \in V'$. Take a set of fresh constants $O = \{o_1, \ldots, o_{|V'|}\}$ and an ABox $\mathcal{A}_C = \{e'(o_i, o_j), d(o_i, o_j) \mid o_i, o_j \in O\}$. Then $\mathcal{A}$ in $\mathcal{P}$ is simply $\mathcal{A} = \mathcal{A}_G \cup \mathcal{A}_{G'} \cup \mathcal{A}_C$.

Let $q = q_1 \wedge q_1' \wedge q_2 \wedge q_2' \wedge q_3$ be a Boolean CQ with

$$
\begin{array}{rcl}
q_1 & = & \{e(x_1, x_2), e(x_2, x_3), \ldots, e(x_{|V|-1}, x_{|V|}))\}, \\
q_1' & = & \{d(x_i, x_j) \mid v_i, v_j \in V, v_i \neq v_j\}, \\
q_2 & = & \{e'(y_1, y_2), e'(y_2, y_3), \ldots, e'(y_{|V'|-1}, y_{|V'|})\}, \\
q_2' & = & \{d(y_i, y_j) \mid v_i', v_j' \in V', v_i' \neq v_j'\}, \\
q_3 & = & \{A(y_1), \ldots, A(y_{|V'|})\}.
\end{array}
$$

Intuitively, $q_1 \wedge q_1'$ asks for a simple path with $|V|$ vertices related via the role e. Analogously, $q_2 \wedge q_2'$ asks for a simple path with $|V'|$ vertices related via the role e'. Additionaly, q_3 asks that each node on the latter path satisfies A.

Finally, let $\mathcal{E} = \{A(o) \mid o \in O\}$.

"$\Rightarrow$" Assume that $\langle G, G' \rangle$ is a positive instance of HP-NOHP, and let $a_1, \ldots a_{|V|}$ be a Hamilton path in G. We show that $\mathcal{E}$ is a $\leq$-minimal and a $\subseteq$-minimal explanation for $\mathcal{P}$. To this end, first take a mapping π for variables in q such that $\pi(x_1) = a_1, \ldots, \pi(x_{|V|}) = a_{|V|}$ and $\pi(y_1) = a_1, \ldots, \pi(y_{|V'|}) = o_{|V|}$. Then clearly π is a match for q in $DB_{\mathcal{A} \cup \mathcal{E}}$, and hence $\mathcal{E}$ is an explanation to $\mathcal{P}$. Indeed, the subquery $q_1 \wedge q_1'$ of q is fulfilled because $a_1, \ldots a_{|V|}$ is a Hamilton path in G, $q_2 \wedge q_2'$ is fulfilled because $\mathcal{A}_C$ has a clique of size $|V'|$, while q_3 is fulfilled by $\mathcal{E}$. To assure minimality, assume towards a contradiction that there is an explanation $\mathcal{E}'$ with $|\mathcal{E}'| < |\mathcal{E}'|$ or $|\mathcal{E}'| \subset |\mathcal{E}'|$. In any case, $|\mathcal{E}'| < |V'|$. Assume π' is a match for q in $DB_{\mathcal{A} \cup \mathcal{E}'}$. Note that $\mathcal{A}_G'$ and $\mathcal{A}_G$ do not share constants. Since $q_3 \wedge q_2'$ asks for $|V'|$ elements satisfying A and $|\mathcal{E}'| < |V'|$, π' must map the variables $y_1, \ldots, y_{|V'|}$ to the $|V'|$ distinct constants of $\mathcal{A}_G'$. Then the presence of q_2 in q implies the existence of a Hamilton path in G'. Contradiction.

"$\Leftarrow$" Assume that $\mathcal{E} \in \mathsf{expl}_{\leq}(\mathcal{P})$ (resp., $\mathcal{E} \in \mathsf{expl}_{\subseteq}(\mathcal{P})$) and π is a match for q in $DB_{\mathcal{A} \cup \mathcal{E}}$. Note that e' does not occur in $\mathcal{A}_G$ and e does not occur in $\mathcal{A}_G' \cup \mathcal{A}_C$. Then by construction of $q_1 \wedge q_1'$ and $\mathcal{A}_G$, π maps the variables $x_1, \ldots, x_{|V|}$ to the $|V|$ distinct constants of $\mathcal{A}_G$ and G must have a Hamilton path. Towards a contradiction suppose G' also has a Hamilton path. Then by construction of $\mathcal{A}_G$, $q_2 \wedge q_2' \wedge q_3$ has a match in $DB_{\mathcal{A}_G}$. This means we can build a match π' for q in $DB_{\mathcal{A}_G}$, which in turn means that $\emptyset$ is an explanation to $\mathcal{P}$. This contradicts the assumption that $\mathcal{E}$ is $\leq$-minimal (resp., $\subseteq$-minimal). $\square$

Computing Explanations We discuss now the problem of actually computing a solution to a QAP $\mathcal{P}$. The complexity of this problem is determined by the established complexity bounds for reasoning tasks over QAPs. Consider first the problem of finding an arbitrary solution $\mathcal{E}$ to $\mathcal{P}$ with unrestricted explanation signature. By Theorem 4, one can do so in polynomial-time by creating a suitable instantiation of the query in the ABox. In general, however, one cannot do better than guessing an ABox $\mathcal{E}$ and deciding whether $\mathcal{E} \in \mathsf{expl}_{\preceq}(\mathcal{P})$. The intuition is that the search space for solutions is intrinsically exponential in the size of the query and minimality criteria require a check over all the solutions.

5 Conclusions

In this paper we characterize the computational complexity of the novel problem of explanation of negative answers to user queries over *DL-Lite*$_{\mathcal{A}}$ ontologies. All the lower bounds proved in the paper do not rely on the notion of FO-rewritability. Our upper bounds rely on FO-rewritability only to argue that solutions are of polynomial size w.r.t. the input query, and on the fact that query answering can be done in NP. For this reason, we expect our results to carry over to other DLs that admit "small" explanations and for which query answering is in NP. For instance, the complexity bounds are applicable to OWL 2-QL, which is obtained from *DL-Lite*$_{\mathcal{A}}$ by forbidding functionality assertions and dropping the unique name assumption (as our results do not rely on functionality axioms, the unique name assumption is irrelevant).

For more expressive DLs, some bounds on the complexity of our reasoning tasks can also be inferred. For QAPs with unrestricted signature, deciding the existence of an explanation has in general the same complexity as checking ontology consistency. If we consider restricted signature, lower bounds follow from results on $\mathcal{CQ}$-emptiness from [Baader et al., 2010], while we expect some upper bounds to be inherited from the complexity of query entailment.

In this work we have focused on combined complexity. With respect to *data complexity* (i.e., when both the query and the TBox are considered fixed), we observe that those inference tasks that we have shown to be NP-complete essentially rely on checking ontology consistency. It follows that they are FOL rewritable, and hence in AC^0 in data complexity. Moreover, given that explanations are bounded by the size of the query (see Proposition 2), it is easy to see that for a fixed query, there are only polynomially many explanations. Hence all our reasoning tasks are polynomial in data complexity and in *ontology complexity* (i.e., when only the query is considered fixed).

Finally, it would be interesting to apply this framework to other lightweight description logics, starting with those of the $\mathcal{EL}$-family. Also, we would like to investigate other minimality criteria. For instance, semantic criteria allow one to reward explanations that are less/more constraining in terms of the models of an ontology.

References

[Artale et al., 2009] Artale, A.; Calvanese, D.; Kontchakov, R.; and Zakharyaschev, M. 2009. The *DL-Lite* family and relations. *J. of Artificial Intelligence Research* 36:1–69.

[Baader et al., 2010] Baader, F.; Bienvenu, M.; Lutz, C.; and Wolter, F. 2010. Query and predicate emptiness in description logics. In *Proc. of KR 2010*.

[Bonatti, Lutz, and Wolter, 2009] Bonatti, P. A.; Lutz, C.; and Wolter, F. 2009. The complexity of circumscription in description logics. *J. of Artificial Intelligence Research* 35:717–773.

[Borgida, Calvanese, and Rodríguez-Muro, 2008] Borgida, A.; Calvanese, D.; and Rodríguez-Muro, M. 2008. Explanation in the *DL-Lite* family of description logics. In *Proc. of ODBASE 2008*, volume 5332 of *LNCS*, 1440–1457. Springer.

[Borgida, Franconi, and Horrocks, 2000] Borgida, A.; Franconi, E.; and Horrocks, I. 2000. Explaining $\mathcal{ALC}$ subsumption. In *Proc. of ECAI 2000*.

[Calvanese et al., 2009] Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; Poggi, A.; Rodríguez-Muro, M.; and Rosati, R. 2009. Ontologies and databases: The *DL-Lite* approach. In Tessaris, S., and Franconi, E., eds., *Semantic Technologies for Informations Systems – 5th Int. Reasoning Web Summer School (RW 2009)*, volume 5689 of *LNCS*. Springer. 255–356.

[Calvanese et al., 2011] Calvanese, D.; Ortiz, M.; Simkus, M.; and Stefanoni, G. 2011. The complexity of conjunctive query abduction in dl-lite. In *Proc. of the 24th Int. Workshop on Description Logics (DL 2011)*, volume 745 of *CEUR Electronic Workshop Proceedings, http://ceur-ws.org/*.

[Chapman and Jagadish, 2009] Chapman, A., and Jagadish, H. V. 2009. Why not? In *Proc. of ACM SIGMOD 2009*, 523–534.

[Eiter and Gottlob, 1995] Eiter, T., and Gottlob, G. 1995. The complexity of logic-based abduction. *J. of the ACM* 42(1):3–42.

[Horridge, Parsia, and Sattler, 2008] Horridge, M.; Parsia, B.; and Sattler, U. 2008. Laconic and precise justifications in OWL. In *Proc. of ISWC 2008*, volume 5318 of *LNCS*, 323–338. Springer.

[Huang et al., 2008] Huang, J.; Chen, T.; Doan, A.; and Naughton, J. 2008. On the provenance of non-answers to queries over extracted data. *PVLDB* 1(1):736–747.

[Klarman, Endriss, and Schlobach, 2011] Klarman, S.; Endriss, U.; and Schlobach, S. 2011. ABox abduction in the description logic $\mathcal{ALC}$. *J. of Automated Reasoning* 46(1):43–80.

[McGuinness and Borgida, 1995] McGuinness, D. L., and Borgida, A. 1995. Explaining subsumption in description logics. In *Proc. of IJCAI'95*, 816–821.

[Motik et al., 2009] Motik, B.; Fokoue, A.; Horrocks, I.; Wu, Z.; Lutz, C.; and Grau, B. C. 2009. OWL 2 Web Ontology Language Profiles. W3C Recommendation, World Wide Web Consortium.

[Papadimitriou, 1994] Papadimitriou, C. H. 1994. *Computational Complexity*. Addison Wesley Publ. Co.

[Penaloza and Sertkaya, 2010] Penaloza, R., and Sertkaya, B. 2010. Complexity of axiom pinpointing in the *DL-Lite* family. In *Proc. of DL 2010*, volume 573 of *CEUR,* `ceur-ws.org`.

[Stewart, 1991] Stewart, I. A. 1991. Complete problems involving boolean labelled structures and projection transactions. *J. of Logic and Computation* 1(6):861–882.

[Wagner, 1987] Wagner, K. W. 1987. More complicated questions about maxima and minima, and some closures of NP. *Theoretical Computer Science* 51(1–2):53–80.