Cardinality Queries over DL-Lite Ontologies

Meghyn Bienvenu[1] , **Quentin Manière**[1] and **Michaël Thomazo**[2]
[1]CNRS, University of Bordeaux, Bordeaux INP, LaBRI, Talence, France
[2]Inria, DI ENS, ENS, CNRS, University PSL, Paris, France
{meghyn.bienvenu, quentin.maniere}@u-bordeaux.fr, michael.thomazo@inria.fr

Abstract

Ontology-mediated query answering (OMQA) employs structured knowledge and automated reasoning in order to facilitate access to incomplete and possibly heterogeneous data. While most research on OMQA adopts (unions of) conjunctive queries as the query language, there has been recent interest in handling queries that involve counting. In this paper, we advance this line of research by investigating cardinality queries (which correspond to Boolean atomic counting queries) coupled with DL-Lite ontologies. Despite its apparent simplicity, we show that such an OMQA setting gives rise to rich and complex behaviour. While we prove that cardinality query answering is tractable (TC^0) in data complexity when the ontology is formulated in DL-Lite$_{core}$, the problem becomes coNP-hard as soon as role inclusions are allowed. For DL-Lite$^{\mathcal{H}}_{pos}$ (which allows only positive axioms), we establish a P-coNP dichotomy and pinpoint the TC^0 cases; for DL-Lite$^{\mathcal{H}}_{core}$ (allowing also negative axioms), we identify new sources of coNP complexity and also exhibit L-complete cases. Interestingly, and in contrast to related tractability results, we observe that the canonical model may not give the optimal count value in the tractable cases, which led us to develop an entirely new approach based upon exploring a space of strategies to determine the minimum possible number of query matches.

1 Introduction

In ontology-mediated query answering (OMQA) [Poggi *et al.*, 2008; Bienvenu and Ortiz, 2015; Xiao *et al.*, 2018], data is enriched with an ontology, which serves both to provide a user-friendly vocabulary for query formulation and to capture domain knowledge that is exploited at query time to obtain a more complete set of answers. While the OMQA approach offers many advantages, it also makes the query answering task more challenging than 'plain' query evaluation. Indeed, instead of having to evaluate the query over the single explicitly given data instance, one must identify the *certain answers*, i.e. those holding in all possible situations (models) compatible with the data and the ontology.

A major topic in OMQA research has thus been to understand the complexity of OMQA and identify tractable settings. Nowadays, for the most commonly considered query language, namely, conjunctive queries (CQs), we have an almost complete picture of the complexity landscape for ontologies formulated in a wide range of different description logics (DLs) [Baader *et al.*, 2017] and rule-based languages [Baget *et al.*, 2011; Calì *et al.*, 2012]. In particular, it has been shown that CQ answering is tractable in data complexity for ontologies expressed in the most commonly considered dialects of the DL-Lite family [Calvanese *et al.*, 2007; Artale *et al.*, 2009], which are often employed in OMQA. A well-known and frequently used property of such DL-Lite dialects and other Horn DLs is that they admit a *canonical model*, which is a single (possibly infinite) model that, by virtue of being homomorphically embeddable into every model, is guaranteed to give the correct answers to all CQs.

While CQs are a natural and well-studied class of queries, there are many other relevant forms of database queries that could be potentially be employed in OMQA. In the present paper, our focus will be on counting queries, which together with other forms of aggregate queries, are widely used for data analysis, yet still not well understood in the context of OMQA. A natural way to equip CQs with counting is to count the number of distinct query matches for each answer. As the count value may differ between models, Kostylev and Reutter (2015) advocated a form of certain answer semantics that considers lower and upper bounds on the count value across different models. Their work provided the first investigation of the complexity of answering counting CQs in the presence of ontologies, revealing such queries to be much more challenging to handle than plain CQs: coNP-complete in data complexity for the well-known DL-Lite$_{core}$ and DL-Lite$^{\mathcal{H}}_{core}$ dialects. A recent work by Bienvenu *et al.* (2020) refined and generalized the complexity results from Kostylev and Reutter to a wider class of counting queries and identified a restricted scenario with very low (TC^0-complete) data complexity: rooted CQs coupled with DL-Lite$_{core}$ ontologies. A similar tractability result for connected rooted CQs was proven independently by Calvanese *et al.* (2020a), who also initiated a study of the impact of other restrictions on query shape and developed the first query rewriting procedure for counting CQs. Notably, both the aforementioned TC^0 result and the rewriting procedure crucially relied upon showing that the canonical model gives the right answers under the considered restrictions. We briefly mention two alternative approaches to counting queries: an epistemic semantics for aggregate queries (which only counts query matches over the data constants, ignoring unnamed elements) was explored by Calvanese *et al.* (2008), while another very recent study by Feier *et al.* (2021) classifies the complexity of counting the number of certain answers (rather than the number of ways a certain answer is obtained) for guarded existential rules.

While recent studies have improved our understanding of the complexity of counting CQs, there nevertheless remain many unanswered questions. In this paper, we focus on Boolean atomic counting queries of the form $\exists z.\mathrm{A}(z)$ and $\exists z_1, z_2.\mathrm{R}(z_1, z_2)$, which we term *cardinality queries* as they correspond to the natural task of determining (bounds on) the cardinality of a given concept or role name. The data complexity of answering such basic counting queries remains completely open for DL-Lite$_{core}$ ontologies, whilst for DL-Lite$^{\mathcal{H}}_{core}$, the problem is known to be P-hard and in coNP [Calvanese *et al.*, 2020a]. The main results of our investigation are displayed in Table 1. We show that when ontologies are expressed in DL-Lite$_{core}$, cardinality query answering is tractable in data complexity and enjoys the lower possible complexity (TC^0-complete). For cardinality queries based upon a concept atom, TC^0 membership holds even for the fragment of DL-Lite$^{\mathcal{H}}_{core}$ obtained by disallowing negative role inclusions. By contrast, for role cardinality queries, we show that coNP-hard situations arise in DL-Lite$^{\mathcal{H}}_{pos}$, which allows only positive concept and role inclusions. In fact, we obtain a complete data complexity classification for DL-Lite$^{\mathcal{H}}_{pos}$, showing that every ontology-mediated query is either TC^0-complete, coNP-complete, or is in P and logspace-equivalent to the complement of PERFECT MATCHING (whose precise complexity is a longstanding open problem). The preceding classification does not extend to DL-Lite$^{\mathcal{H}}_{core}$: we identify new sources of coNP-hardness and further exhibit L-complete cases. We find it intriguing that such complex behaviour arises in what appears at first glance to be a simple OMQA setting. Moreover, in all of the tractable cases we identify, the canonical model may not yield the minimum cardinality, and query answering involves solving non-trivial optimization problems. This led us to devise an entirely new approach based upon exploring a space of strategies to find the optimal way of merging witnesses for existential axioms.

The paper is organized as follows. Section 2 recalls relevant background material and presents the considered OMQA setting. Section 3 introduces strategies and uses them to establish TC^0 membership. Our complexity classification for DL-Lite$^{\mathcal{H}}_{pos}$ is the topic of Section 4, while Section 5 presents our results for DL-Lite$^{\mathcal{H}}_{core}$. Section 6 concludes with a brief discussion of related and future work.

An appendix with full proofs can be found in the long version of this paper, available on arXiv.

2 Preliminaries

We recall standard definitions and notation for OMQA in DL-Lite and introduce the particular setting studied in this paper.

	Concept	Role
DL-Lite$_{core}$	TC^0-c	TC^0-c
DL-Lite$^{\mathcal{H}}_{pos}$	TC^0-c†	TC^0-c | co-PM-c | coNP-c
DL-Lite$^{\mathcal{H}}_{core}$	TC^0-c | L-c | coNP-c | ?	TC^0-c | L-c | co-PM-c | coNP-c | ?

Table 1: Data complexity of cardinality queries based upon concept and role atoms for various DL-Lite dialects. †: upper bound holds for all DL-Lite$^{\mathcal{H}}_{core}$ ontologies without negative role inclusions.

Knowledge Bases. We assume mutually disjoint sets $\mathsf{N_C}$ of concept names (unary predicates), $\mathsf{N_R}$ of role names (binary predicates), and $\mathsf{N_I}$ of individual names (constants). We denote by $\mathsf{N_R^{\pm}}$ the set $\mathsf{N_R} \cup \{\mathrm{R}^- \mid \mathrm{R} \in \mathsf{N_R}\}$ of role names and their inverses. A *knowledge base (KB)* $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ consists of an ABox (dataset) $\mathcal{A}$ and a TBox (ontology) $\mathcal{T}$. An *ABox* is a finite set of *concept assertions* $\mathrm{A}(\mathsf{b})$ (with $\mathrm{A} \in \mathsf{N_C}$, $b \in \mathsf{N_I}$) and *role assertions* $\mathrm{P}(\mathsf{a}, \mathsf{b})$ (with $\mathrm{P} \in \mathsf{N_R}$, $\mathsf{a}, \mathsf{b} \in \mathsf{N_I}$), while the TBox consists of a finite set of axioms, whose forms are dictated by the considered description logic.

In this paper, our focus will be on DL-Lite$^{\mathcal{H}}_{core}$ (alternatively referred to as DL-Lite$_{\mathcal{R}}$), which is the logic underlying the OWL 2 QL profile. DL-Lite$^{\mathcal{H}}_{core}$ TBoxes contain four types of axioms: *positive concept inclusions* $\mathrm{B}_1 \sqsubseteq \mathrm{B}_2$, *negative concept inclusions* $\mathrm{B}_1 \sqsubseteq \neg\mathrm{B}_2$, *positive role inclusions* $\mathrm{R}_1 \sqsubseteq \mathrm{R}_2$, and *negative role inclusions* $\mathrm{R}_1 \sqsubseteq \neg\mathrm{R}_2$, where the $\mathrm{B_i}$ and $\mathrm{R_i}$ are *positive concepts and roles* given by:

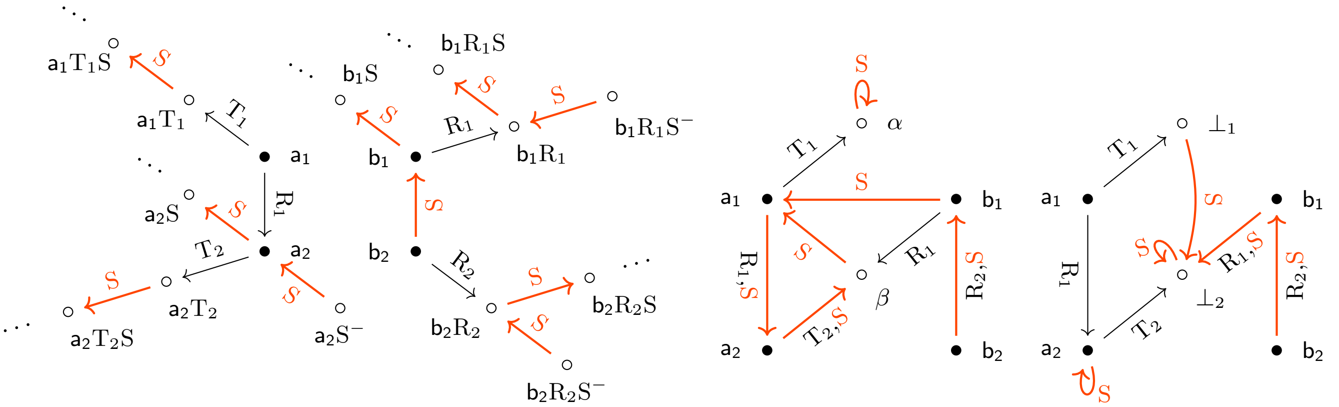
$$\mathrm{B_i} := \mathrm{A} \mid \exists \mathrm{R_i} \qquad \mathrm{R_i} := \mathrm{P} \mid \mathrm{P}^- \qquad (\mathrm{A} \in \mathsf{N_C},\ \mathrm{P} \in \mathsf{N_R})$$

The sublogic DL-Lite$_{core}$ allows only concept inclusions (which may be either positive or negative), while DL-Lite$^{\mathcal{H}}_{pos}$ is restricted to positive (concept and role) inclusions.

We denote by $\mathsf{Ind}(\mathcal{A})$ the set of individuals occurring in an ABox $\mathcal{A}$. A *signature* is a finite set of concept and role names. Given a signature Σ, we denote by $\Sigma^{\pm}_{\mathsf{C}}$ (resp. $\Sigma^{\pm}_{\mathsf{R}}$) the set of positive concepts (resp. roles) built from Σ. The signature of a TBox $\mathcal{T}$ (resp. ABox $\mathcal{A}$) is the set of concept and role names it contains, denoted $\mathsf{sig}(\mathcal{T})$ (resp. $\mathsf{sig}(\mathcal{A})$). To simplify the presentation, we will assume w.l.o.g. that $\mathsf{sig}(\mathcal{A}) \subseteq \mathsf{sig}(\mathcal{T})$.

Semantics of KBs. An interpretation takes the form $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is a non-empty set (called the domain) and $\cdot^{\mathcal{I}}$ is the interpretation function that maps each $\mathrm{A} \in \mathsf{N_C}$ to $\mathrm{A}^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, each $\mathrm{P} \in \mathsf{N_R}$ to $\mathrm{P}^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, and each $\mathsf{a} \in \mathsf{N_I}$ to $\mathsf{a}^{\mathcal{I}}$. In this paper, we will make the *Standard Names Assumption* by setting $\mathsf{a}^{\mathcal{I}} = \mathsf{a}$. Note however that our results only rely upon the weaker *Unique Names Assumption* (UNA), which stipulates that $\mathsf{a}^{\mathcal{I}} \neq \mathsf{b}^{\mathcal{I}}$ whenever $\mathsf{a} \neq \mathsf{b}$. The UNA is commonly adopted for DL-Lite KBs and enables more interesting reasoning in the context of counting queries.

The function $\cdot^{\mathcal{I}}$ is extended to general concepts and roles as follows: $(\mathrm{P}^-)^{\mathcal{I}} = \{(e, d) \mid (d, e) \in \mathrm{P}^{\mathcal{I}}\}$, $(\exists\mathrm{R})^{\mathcal{I}} = \{d \mid (d, e) \in \mathrm{R}^{\mathcal{I}}\}$, and $(\neg\mathrm{G})^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus \mathrm{G}^{\mathcal{I}}$. An inclusion $\mathrm{G} \sqsubseteq \mathrm{H}$ is satisfied in $\mathcal{I}$ if $\mathrm{G}^{\mathcal{I}} \subseteq \mathrm{H}^{\mathcal{I}}$; an assertion $\mathrm{A}(\mathsf{b})$ (resp. $\mathrm{P}(\mathsf{a}, \mathsf{b})$) is satisfied in $\mathcal{I}$ if $\mathsf{b} \in \mathrm{A}^{\mathcal{I}}$ (resp. $(\mathsf{a}, \mathsf{b}) \in \mathrm{P}^{\mathcal{I}}$). An interpretation is a model of a TBox $\mathcal{T}$ (resp. ABox $\mathcal{A}$)) if it satisfies all axioms in $\mathcal{T}$ (resp. assertions in $\mathcal{A}$), and it is a *model of a KB* $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ if it is a model of both $\mathcal{T}$ and $\mathcal{A}$. A KB



(a) Initial portion of the canonical model of $\mathcal{K}_e$.

(b) Another model of $\mathcal{K}_e$.

(c) Interpretation of strategy σ_e.

Figure 1: Models of the example KB $\mathcal{K}_e$. For readability, we have omitted concepts and highlighted the role S from the cardinality query.

is *satisfiable* if it has at least one model. An inclusion (resp. assertion) Φ is *entailed* from $\mathcal{T}$ (resp. $\mathcal{K}$), written $\mathcal{T} \models \Phi$ (resp. $\mathcal{K} \models \Phi$), if Φ is satisfied in every model of $\mathcal{T}$ (resp. $\mathcal{K}$). We use $\mathcal{K} \models \exists \mathrm{R}(\mathsf{a})$ (resp. $\mathcal{K} \models \mathrm{R}(\mathsf{a}, \mathsf{b})$ with $\mathrm{R} \in \mathsf{N}_\mathsf{R}^{\pm}$) to indicate $\mathsf{a} \in \exists \mathrm{R}^{\mathcal{I}}$ (resp. $(\mathsf{a}, \mathsf{b}) \in \mathrm{R}^{\mathcal{I}}$) for every model $\mathcal{I}$ of $\mathcal{K}$.

Example 1. *As a running example, we will consider the KB* $\mathcal{K}_e = (\mathcal{T}_e, \mathcal{A}_e)$ *whose TBox contains the following inclusions*

$$\begin{array}{llll} \mathrm{A}_1 \sqsubseteq \exists \mathrm{T}_1 & \mathrm{A}_2 \sqsubseteq \exists \mathrm{T}_2 & \exists \mathrm{T}_1^- \sqsubseteq \exists \mathrm{S} & \exists \mathrm{R}_1^- \sqsubseteq \neg \exists \mathrm{R}_2^- \\ \mathrm{B}_1 \sqsubseteq \exists \mathrm{R}_1 & \mathrm{B}_2 \sqsubseteq \exists \mathrm{R}_2 & \exists \mathrm{R}_1^- \sqsubseteq \exists \mathrm{S}^- & \exists \mathrm{R}_1^- \sqsubseteq \neg \exists \mathrm{T}_1^- \\ \exists \mathrm{T}_2^- \sqsubseteq \exists \mathrm{S} & \exists \mathrm{S}^- \sqsubseteq \exists \mathrm{S} & \exists \mathrm{R}_2^- \sqsubseteq \exists \mathrm{S}^- & \end{array}$$

and whose ABox contains the assertions

$$\{\mathrm{A}_1(\mathsf{a}_1), \mathrm{A}_2(\mathsf{a}_2), \mathrm{B}_1(\mathsf{b}_1), \mathrm{B}_2(\mathsf{b}_2), \mathrm{R}_1(\mathsf{a}_1, \mathsf{a}_2), \mathrm{S}(\mathsf{b}_2, \mathsf{b}_1)\}$$

Two finite models of $\mathcal{K}_e$ *are displayed in Figures 1b and 1c.*

Canonical Model. Every satisfiable DL-Lite$^{\mathcal{H}}_{core}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ has a *canonical model* $\mathcal{C}_\mathcal{K}$, defined as follows. The domain of $\mathcal{C}_\mathcal{K}$ contains $\mathsf{Ind}(\mathcal{A})$ and all words $\mathsf{a}\mathrm{R}_1 \ldots \mathrm{R}_n$, with $\mathsf{a} \in \mathsf{Ind}(\mathcal{A})$, $\mathrm{R}_i \in \mathsf{N}_\mathsf{R}^{\pm}$, and $n \geqslant 1$, such that:

- $\mathcal{K} \models \exists \mathrm{R}_1(\mathsf{a})$ and there is no $\mathrm{R}_1(\mathsf{a}, \mathsf{b}) \in \mathcal{A}$;
- for $1 \leq i < n$, $\mathcal{T} \models \exists \mathrm{R}_i^- \sqsubseteq \exists \mathrm{R}_{i+1}$ and $\mathrm{R}_i^- \neq \mathrm{R}_{i+1}$.

Concept and role names are interpreted as follows:

$$\begin{aligned} \mathrm{A}^{\mathcal{C}_\mathcal{K}} &= \{\mathsf{a} \in \mathsf{Ind}(\mathcal{A}) \mid \mathcal{K} \models \mathrm{A}(a)\} \\ &\quad \cup \{\mathsf{a}\mathrm{R}_1 \ldots \mathrm{R}_n \in \Delta^{\mathcal{C}_\mathcal{K}} \setminus \mathsf{Ind}(\mathcal{A}) \mid \mathcal{T} \models \exists \mathrm{R}_n^- \sqsubseteq \mathrm{A}\} \\ \mathrm{P}^{\mathcal{C}_\mathcal{K}} &= \{(\mathsf{a}, \mathsf{b}) \mid \mathrm{P}(\mathsf{a}, \mathsf{b}) \in \mathcal{A}\} \\ &\quad \cup \{(e_1, e_2) \mid e_2 = e_1\mathrm{R} \text{ and } \mathcal{T} \models \mathrm{R} \sqsubseteq \mathrm{P}\} \\ &\quad \cup \{(e_2, e_1) \mid \ e_2 = e_1\mathrm{R} \text{ and } \mathcal{T} \models \mathrm{R} \sqsubseteq \mathrm{P}^-\} \end{aligned}$$

We use $\mathsf{gen}_\mathcal{K}$ to refer to the set of *generated roles*, i.e. those $\mathrm{R} \in \mathsf{N}_\mathsf{R}^{\pm}$ such that $\Delta^{\mathcal{C}_\mathcal{K}}$ contains an element $w\mathrm{R}$.

Example 2. *An initial portion of (the infinite) canonical model of* $\mathcal{K}_e$ *is displayed in Figure 1a. Observe that* $\mathsf{gen}_\mathcal{K} = \{\mathrm{S}, \mathrm{S}^-, \mathrm{R}_1, \mathrm{R}_2, \mathrm{T}_1, \mathrm{T}_2\}$.

It is well known (see e.g. [Calvanese *et al.*, 2007]) that, for every model $\mathcal{I}$ of $\mathcal{K}$, there is a *homomorphism* from $\mathcal{C}_\mathcal{K}$ to $\mathcal{I}$, i.e. a function $f : \Delta^{\mathcal{C}_\mathcal{K}} \to \Delta^{\mathcal{I}}$ such that (i) $f(\mathsf{a}) = \mathsf{a}$ for all $\mathsf{a} \in \mathsf{Ind}$, (ii) $e \in \mathrm{A}^{\mathcal{C}_\mathcal{K}}$ implies $f(e) \in \mathrm{A}^{\mathcal{I}}$, and (iii) $(d, e) \in \mathrm{P}^{\mathcal{C}_\mathcal{K}}$ implies $(f(d), f(e)) \in \mathrm{P}^{\mathcal{I}}$.

Cardinality Queries. A *cardinality query* is either a *concept cardinality query* $\exists z.\mathrm{C}(z)$ or a *role cardinality query* $\exists z_1, z_2.\mathrm{S}(z_1, z_2)$. Throughout the paper, we use q_C (resp. q_S) as a shorthand for the cardinality query based upon C (resp. S). A *match* for a cardinality query q_C (resp. q_S) in an interpretation $\mathcal{I}$ is an element of $\mathrm{C}^{\mathcal{I}}$ (resp. $\mathrm{S}^{\mathcal{I}}$). We define the *answer to a cardinality query* q *in an interpretation* $\mathcal{I}$, denoted $q^{\mathcal{I}}$, as the number of matches of q in $\mathcal{I}$, or equivalently, as the cardinality of $\mathrm{F}^{\mathcal{I}}$, with F the concept or role name in q. A *certain answer* to q w.r.t. $\mathcal{K}$ is an interval $[m, M] \in \mathbb{N} \times \mathbb{N}$ such that $q^{\mathcal{I}} \in [m, M]$ for every model $\mathcal{I}$ of $\mathcal{K}$.

Example 3. *Consider the role cardinality query* q_S. *The answer to* q_S *is* $+\infty$ *in* $\mathcal{C}_{\mathcal{K}_e}$, 6 *in the model from Figure 1b, and* 5 *in the model from Figure 1c. The latter implies that* $[6, +\infty]$ is not *a certain answer. We leave it is an exercise to find a model with 3 matches and show there is no model with fewer matches, which means that* $[m, +\infty]$ *is a certain answer to* q_S *over* $\mathcal{K}_e$ *if and only if* $m \leq 3$.

Cardinality queries as defined above correspond to a special case of the counting queries considered in [Kostylev and Reutter, 2015; Bienvenu *et al.*, 2020; Calvanese *et al.*, 2020a].

Observe that since DL-Lite$^{\mathcal{H}}_{core}$ cannot restrict the size of models, the value M in a certain answer $[m, M]$ must be $+\infty$ whenever the query predicate F is satisfiable w.r.t. $\mathcal{T}$ (i.e. there is a model $\mathcal{I}$ of $\mathcal{T}$ such that $\mathrm{F}^{\mathcal{I}} \neq \emptyset$). For this reason, we assume the latter condition holds and focus on identifying certain answers of the form $[m, +\infty]$.

Complexity. We will be interested in classifying the complexity of the following problem:

> OMQA$(q, \mathcal{T})$: Given $\mathcal{A}$ and integer $m \geq 1$ (in binary), decide whether $[m, +\infty]$ is a certain answer to q w.r.t. $\mathcal{K}$.

where $(q, \mathcal{T})$ is an *ontology-mediated query (OMQ)* based upon a cardinality query q and a TBox $\mathcal{T}$ formulated in DL-Lite$^{\mathcal{H}}_{core}$ or one of its sublogics. Note that we are adopting the *data complexity* measure as $(q, \mathcal{T})$ is fixed.

Beyond well-known complexity classes such as P and coNP, we will refer to the following classes: TC^0 is the class of problems solvable by families of constant-depth polynomial-size circuits based upon AND, OR, NOT, and threshold gates, and L (resp. NL) is the class of problems solvable in deterministic (resp. nondeterministic) logarithmic space. It is known that: $\mathsf{TC}^0 \subseteq \mathsf{L} \subseteq \mathsf{NL} \subseteq \ldots \subseteq \mathsf{P} \subseteq \mathsf{coNP}$.

3 Tractable Cases

In this section, we identify two settings in which cardinality queries can be answered with the lowest possible complexity:

Theorem 1. *$\mathrm{OMQA}(q,\mathcal{T})$ is TC^0-complete if either (i) q is a role cardinality query and $\mathcal{T}$ a DL-Lite$_{core}$ TBox, or (ii) q is a concept cardinality query and $\mathcal{T}$ is a DL-Lite$^{\mathcal{H}}_{core}$ TBox without negative role inclusions.*

The remainder of this section is devoted to establishing TC^0 membership for case (i) where our query is $q_{\mathrm{S}} = \exists z_1, z_2.\mathrm{S}(z_1, z_2)$. A similar but simpler argument can be used for the membership half of case (ii), while TC^0-hardness is easily shown by reduction from the TC^0-complete NumOnes problem [Aehlig *et al.*, 2007] asking, given a binary string X and $k \geq 1$, whether X contains at least k 1-bits.

Existing proofs of sub-polynomial data complexity for restricted classes of counting queries rely on the canonical model minimizing the number of matches [Bienvenu *et al.*, 2020; Calvanese *et al.*, 2020a]. However, for the class of cardinality queries, the canonical model may not yield the minimum value. Therefore, we develop a different approach based upon a systematic exploration of a set of models that is guaranteed to contain an optimal model and whose size depends only on the TBox. This special set of models will be induced from strategies that dictate how to merge elements of the canonical model. To show such models contain the optimal value, we show that if we extract a strategy σ from an arbitrary model $\mathcal{I}$ and consider any model $\mathcal{J}$ induced by σ, then $\mathcal{J}$ has at most as many matches as the initial model $\mathcal{I}$.

We now formalize this approach. In order to abstract from specific ABox individuals, we introduce types.

Definition 1. *A* type *for a TBox $\mathcal{T}$ is a subset of $\mathsf{sig}(\mathcal{T})^{\pm}_{\mathsf{C}}$. The set of all types is $\Theta_{\mathcal{T}} = 2^{\mathsf{sig}(\mathcal{T})^{\pm}_{\mathsf{C}}}$. We denote by $\theta_{\mathcal{K}}(d)$ the* type of a domain element d w.r.t. $\mathcal{K}$ *and define it by: $\theta_{\mathcal{K}}(d) = \{\mathrm{B} \in \mathsf{sig}(\mathcal{T})^{\pm}_{\mathsf{C}} \mid \mathcal{K} \models \mathrm{B}(d)\}$ if $d \in \mathsf{Ind}(\mathcal{A})$, else $\theta_{\mathcal{K}}(d) = \emptyset$.*

Example 4. *In our running example, $\theta_{\mathcal{K}_e}(\mathsf{a}_1) = \{\mathrm{A}_1, \exists\mathrm{R}_1, \exists\mathrm{T}_1\}$ and $\theta_{\mathcal{K}_e}(\alpha) = \emptyset$ (since $\alpha \notin \mathsf{Ind}(\mathcal{A}_e)$).*

Strategies indicate for each generated role R the type onto which all elements $w\mathrm{R}$ should merge. Several copies of a type might be required to comply with negative inclusions (e.g. R_1 and R_2 associated to the same type but $\mathcal{T} \models \exists\mathrm{R}_1^- \sqsubseteq \neg\exists\mathrm{R}_2^-$).

Definition 2. *A* strategy *σ for the KB $\mathcal{K}$ is a function from $\mathsf{gen}_{\mathcal{K}}$ to $\Theta_{\mathcal{T}} \times \{1, \ldots, |\mathsf{sig}(\mathcal{T})^{\pm}_{\mathsf{R}}|\}$, that satisfies:*

1. $\forall \mathrm{R} \in \mathsf{gen}_{\mathcal{K}} : \sigma(\mathrm{R}) = (\mathfrak{t}, i) \wedge \mathrm{B} \in \mathfrak{t} \Rightarrow \mathcal{T} \not\models \exists\mathrm{R}^- \sqsubseteq \neg\mathrm{B}$.

2. $\forall \mathrm{R}_1, \mathrm{R}_2 \in \mathsf{gen}_{\mathcal{K}} : \sigma(\mathrm{R}_1) = \sigma(\mathrm{R}_2) \Rightarrow \mathcal{T} \not\models \exists\mathrm{R}_1^- \sqsubseteq \neg\exists\mathrm{R}_2^-$.

3. $\forall \mathfrak{t} \in \Theta_{\mathcal{T}}$, *if* $\mathfrak{t} \neq \emptyset$, *then* $|\{i \mid \exists\mathrm{R} \in \mathsf{gen}_{\mathcal{K}}, \sigma(\mathrm{R}) = (\mathfrak{t}, i)\}| \leq |\{\mathsf{a} \mid \mathsf{a} \in \mathsf{Ind}(\mathcal{A}) \wedge \theta_{\mathcal{K}}(\mathsf{a}) = \mathfrak{t}\}|$.

Conditions 1 and 2 ensure that merging will not violate any negative inclusions. Condition 3 ensures the ABox provides at least as many individuals of a non-empty type as the strategy requires copies of this type.

Example 5. *The following mapping σ_e is a strategy for $\mathcal{K}_e$:*

$$\begin{array}{lcllcl}
\mathrm{T}_1 & \mapsto & (\emptyset, 1) & \mathrm{R}_2 & \mapsto & (\{\mathrm{B}_1, \exists\mathrm{R}_1, \exists\mathrm{S}, \exists\mathrm{S}^-\}, 1) \\
\mathrm{T}_2 & \mapsto & (\emptyset, 2) & \mathrm{S} & \mapsto & (\emptyset, 2) \\
\mathrm{R}_1 & \mapsto & (\emptyset, 2) & \mathrm{S}^- & \mapsto & (\{\mathrm{A}_1, \exists\mathrm{R}_1, \exists\mathrm{T}_1\}, 1)
\end{array}$$

To construct a model from a strategy σ, the basic idea is to merge elements $w\mathrm{R}$ with an element of type $\sigma(\mathrm{R})$, with the latter selected according to a *choice of well-typed elements*:

Definition 3. *A mapping* $\mathsf{ch} : \mathsf{gen}_{\mathcal{K}} \to \mathsf{Ind}(\mathcal{A}) \uplus \{\bot_i \mid i = 1, \ldots, |\mathsf{sig}(\mathcal{T})^{\pm}_{\mathsf{R}}|\}$, *is a* choice of well-typed elements for σ over $\mathcal{K}$ *if it satisfies the following conditions:*

1. $\forall \mathrm{R} \in \mathsf{gen}_{\mathcal{K}}, \exists i$ *such that* $\sigma(\mathrm{R}) = (\theta_{\mathcal{K}}(\mathsf{ch}(\mathrm{R})), i)$

2. $\forall \mathrm{R}_1, \mathrm{R}_2 \in \mathsf{gen}_{\mathcal{K}},\ \mathsf{ch}(\mathrm{R}_1) = \mathsf{ch}(\mathrm{R}_2) \Leftrightarrow \sigma(\mathrm{R}_1) = \sigma(\mathrm{R}_2)$.

Example 6. *The function ch_e, defined as below, is a choice of well-typed elements for σ_e over $\mathcal{K}_e$:*

$$\begin{array}{lclclclclcl}
\mathrm{T}_1 & \mapsto & \bot_1 & \mathrm{T}_2 & \mapsto & \bot_2 & \mathrm{R}_1 & \mapsto & \bot_2 \\
\mathrm{R}_2 & \mapsto & \mathsf{b}_1 & \mathrm{S} & \mapsto & \bot_2 & \mathrm{S}^- & \mapsto & \mathsf{a}_1
\end{array}$$

It turns out however that when $\mathrm{R} = \mathrm{S}$ or $\mathrm{R} = \mathrm{S}^-$, it is useful to depart from this guideline in order to reduce the number of query matches, as this stand-alone example illustrates:

Example 7. *Consider $\mathcal{T} = \{\mathrm{A} \sqsubseteq \exists\mathrm{S}, \mathrm{B} \sqsubseteq \exists\mathrm{S}^-\}$ and $\mathcal{A} = \{\mathrm{A}(\mathsf{a}_1), \mathrm{A}(\mathsf{a}_2), \mathrm{B}(\mathsf{b}_1), \mathrm{B}(\mathsf{b}_2)\}$. If we merge $\mathsf{a}_1\mathrm{S}$ with $\mathsf{a}_2\mathrm{S}$, and $\mathsf{b}_1\mathrm{S}^-$ with $\mathsf{b}_2\mathrm{S}^-$, then there will be at least three matches of q_{S}, no matter which further merges are performed. However, by 'pairing' a_1 with b_1 and a_2 with b_2, we can obtain a model with only two matches: $(\mathsf{a}_1, \mathsf{b}_1), (\mathsf{a}_2, \mathsf{b}_2)$.*

The next three definitions serve to identify the *critical elements* for which such a pairing operation is useful.

Definition 4. *We set $\mathcal{D}^+_{\mathcal{K}} = \{\mathsf{a} \mid \mathsf{a} \in \mathsf{Ind}(\mathcal{A}) \wedge \mathsf{a}\mathrm{S} \in \Delta^{\mathcal{C}_{\mathcal{K}}}\}$ and $\mathcal{D}^-_{\mathcal{K}} = \{\mathsf{a} \mid \mathsf{a} \in \mathsf{Ind}(\mathcal{A}) \wedge \mathsf{a}\mathrm{S}^- \in \Delta^{\mathcal{C}_{\mathcal{K}}}\}$.*

Definition 5. *Given a strategy σ, we set $\mathcal{D}^+_{\sigma} = \{\mathrm{R} \mid \mathrm{R} \in \mathsf{dom}(\sigma) \setminus \{\mathrm{S}, \mathrm{S}^-\} \wedge \mathcal{T} \models \exists\mathrm{R}^- \sqsubseteq \exists\mathrm{S} \wedge \exists\mathrm{S} \notin \mathfrak{t}$ if $\sigma(\mathrm{R}) = (\mathfrak{t}, k)\}$ and $\mathcal{D}^-_{\sigma} = \{\mathrm{R} \mid \mathrm{R} \in \mathsf{dom}(\sigma) \setminus \{\mathrm{S}, \mathrm{S}^-\} \wedge \mathcal{T} \models \exists\mathrm{R}^- \sqsubseteq \exists\mathrm{S}^- \wedge \exists\mathrm{S}^- \notin \mathfrak{t}$ if $\sigma(\mathrm{R}) = (\mathfrak{t}, k)\}$.*

Definition 6. *Let ch be a choice of well-typed elements for σ. We set $\mathsf{crit}^+ = \mathcal{D}^+_{\mathcal{K}} \cup \mathsf{ch}(\mathcal{D}^+_{\sigma})$ and $\mathsf{crit}^- = \mathcal{D}^-_{\mathcal{K}} \cup \mathsf{ch}(\mathcal{D}^-_{\sigma})$ and use* critical elements *to refer to the elements of these sets.*

Example 8. *For σ_e and ch_e as defined in Examples 5 and 6, we have $\mathsf{crit}^+ = \{\mathsf{a}_2, \mathsf{b}_1, \bot_1, \bot_2\}$ and $\mathsf{crit}^- = \{\mathsf{a}_2, \bot_2\}$.*

Intuitively, a pairing matches critical elements from crit^+ (which require an outgoing S) with those from crit^- (which require an incoming S).

Definition 7. *A* pairing for ch and σ *consists of two partial functions $\mathsf{p}^+ : \mathsf{crit}^+ \to \mathsf{crit}^-$ and $\mathsf{p}^- : \mathsf{crit}^- \to \mathsf{crit}^+$ such that one of the functions is total and injective, and the other is its partial inverse.*

Example 9. *A pairing for ch_e and σ_e is given by $\mathsf{p}^+_e = \{\mathsf{a}_2 \mapsto \mathsf{a}_2, \mathsf{b}_1 \mapsto \bot_2\}$ and $\mathsf{p}^-_e = \{\mathsf{a}_2 \mapsto \mathsf{a}_2, \bot_2 \mapsto \mathsf{b}_1\}$.*

We are now ready to define the interpretation of a strategy.

Definition 8. *Consider a strategy σ, choice of well-typed elements ch, and pairing $(\mathsf{p}^+, \mathsf{p}^-)$ for ch. For every $\mathrm{R} \in \mathsf{sig}(\mathcal{T})^{\pm}_{\mathsf{R}}$, pick a function s_{R} that maps every individual in*

$\{\mathsf{a} \mid \mathcal{K} \models \mathrm{R}(\mathsf{a}, \mathsf{b})$ *for some* $\mathsf{b} \in \mathsf{N_I}\}$ *to an individual* $\mathsf{s}_{\mathrm{R}}(\mathsf{a})$ *such that* $\mathcal{K} \models \mathrm{R}(\mathsf{a}, \mathsf{s}_{\mathrm{R}}(\mathsf{a}))$. *Define function* χ *as follows:*

$$\begin{array}{rcl}
\Delta^{\mathcal{C}_{\mathcal{K}}} & \rightarrow & \mathsf{Ind}(\mathcal{A}) \cup \{\bot_i \mid i = 1, \ldots, |\mathsf{sig}(\mathcal{T})^{\pm}_{\mathsf{R}}|\} \\
\mathsf{a} & \mapsto & \mathsf{a} \\
w\mathrm{S} & \mapsto & \begin{cases} \mathsf{s}_{\mathrm{S}}(\chi(w)) & \textit{if } \mathsf{s}_{\mathrm{S}}(\chi(w)) \textit{ is defined} \\ \mathsf{p}^+(\chi(w)) & \textit{else if } \mathsf{p}^+(\chi(w)) \textit{ is defined} \\ \mathsf{ch}(\mathrm{S}) & \textit{otherwise} \end{cases} \\
w\mathrm{S}^- & \mapsto & \begin{cases} \mathsf{s}_{\mathrm{S}^-}(\chi(w)) & \textit{if } \mathsf{s}_{\mathrm{S}^-}(\chi(w)) \textit{ is defined} \\ \mathsf{p}^-(\chi(w)) & \textit{else if } \mathsf{p}^-(\chi(w)) \textit{ is defined} \\ \mathsf{ch}(\mathrm{S}^-) & \textit{otherwise} \end{cases} \\
w\mathrm{R} & \mapsto & \mathsf{ch}(\mathrm{R})
\end{array}$$

The interpretation of σ *(according to* ch, $(\mathsf{p}^+, \mathsf{p}^-)$ *and the* s_{R}*) has domain* $\chi(\Delta^{\mathcal{C}_{\mathcal{K}}})$ *and interpretation function* $\chi \circ \cdot^{\mathcal{C}_{\mathcal{K}}}$.

Example 10. *With choice* ch_e *and pairing* $(\mathsf{p}^+_e, \mathsf{p}^-_e)$*, we get* $\chi(\mathsf{b}_2\mathrm{R}_2) = \mathsf{ch}(\mathrm{R}_2) = \mathsf{b}_1$*,* $\chi(\mathsf{b}_2\mathrm{R}_2\mathrm{S}) = \mathsf{p}^+_e(\mathsf{b}_1) = \bot_2$*, and* $\chi(\mathsf{b}_2\mathrm{R}_2\mathrm{S}^-) = \mathsf{s}_{\mathrm{S}^-}(\mathsf{b}_1) = \mathsf{b}_2$ *(observe that on our example, the function* $\mathsf{s}_{\mathrm{S}^-}$ *is uniquely defined, and the same is true for the other roles). Figure 1c displays the interpretation of* σ_e.

Observe that the interpretation of a strategy σ depends not only on σ but also on the functions $\mathsf{ch}, \mathsf{p}^+, \mathsf{p}^-, \mathsf{s}_{\mathrm{R}}$. Importantly, however, the key property of such interpretations (stated in Lemma 1 later in this section) holds for *any* particular choice of these functions.

It remains to prove that a model minimizing the number of matches can be found among the interpretations of strategies. The first step is to to extract a strategy from a model.

Definition 9. *Let* $\mathcal{I}$ *be a model of* $\mathcal{K}$*,* $f : \mathcal{C}_{\mathcal{K}} \rightarrow \mathcal{I}$ *be a homomorphism, and* repr *be a function mapping each role* $\mathrm{R} \in \mathsf{gen}_{\mathcal{K}}$ *to an element with shape* $w\mathrm{R}$ *from* $\Delta^{\mathcal{C}_{\mathcal{K}}}$*. Then* $\mathcal{P} = \{P_1, \ldots, P_k\}$*, defined by*

$$\{P_1, \ldots, P_k\} = \{(f \circ \mathsf{repr})^{-1}(w) \mid w \in \Delta^{\mathcal{I}}\} \setminus \{\emptyset\}$$

is a partition of $\mathsf{gen}_{\mathcal{K}}$*. The strategy extracted from* $\mathcal{I}$ *(for* f *and* repr*) is defined as:*

$$\begin{array}{rcl}
\mathsf{gen}_{\mathcal{K}} & \rightarrow & \Theta_{\mathcal{T}} \times \{1, \ldots, |\mathsf{sig}(\mathcal{T})^{\pm}_{\mathsf{R}}|\} \\
\mathrm{R} & \mapsto & ((\theta_{\mathcal{K}} \circ f \circ \mathsf{repr})(\mathrm{R}), i) \textit{ with } \mathrm{R} \in P_i
\end{array}$$

Example 11. *In our running example, there is a unique homomorphism* f_e *from* $\mathcal{C}_{\mathcal{K}_e}$ *to the model displayed in Figure 1b. Let* repr_e *be:*

$$\begin{array}{lll lll lll}
\mathrm{T}_1 & \mapsto & \mathsf{a}_1\mathrm{T}_1 & \mathrm{R}_2 & \mapsto & \mathsf{b}_2\mathrm{R}_2 & \mathrm{T}_2 & \mapsto & \mathsf{a}_2\mathrm{T}_2 \\
\mathrm{S} & \mapsto & \mathsf{b}_1\mathrm{SSS} & \mathrm{R}_1 & \mapsto & \mathsf{b}_1\mathrm{R}_1 & \mathrm{S}^- & \mapsto & \mathsf{a}_2\mathrm{S}^-
\end{array}$$

The strategy extracted from this model (for f_e *and* repr_e*) is the strategy provided in Example 5.*

By applying the next lemma to a model $\mathcal{I}$ having the fewest possible number of matches, we obtain the desired conclusion: there is a model minimizing the number of matches among the models obtained by interpreting a strategy.

Lemma 1. *Let* $\mathcal{I}$ *be a model of* $\mathcal{K}$*, and* $\mathcal{J}$ *an interpretation of a strategy extracted from* $\mathcal{I}$*.* $\mathcal{J}$ *is a model of* $\mathcal{K}$ *and* $q_{\mathrm{S}}^{\mathcal{J}} \leq q_{\mathrm{S}}^{\mathcal{I}}$.

We now sketch how to construct a family of TC^0 circuits (one for each size of ABox) to decide $\mathrm{OMQA}(q_{\mathrm{S}}, \mathcal{T})$. Each such circuit first computes the set $\mathsf{gen}_{\mathcal{K}}$ and the type of each ABox individual. Next, for each function $\varrho : \mathsf{gen}_{\mathcal{K}} \rightarrow \Theta_{\mathcal{T}} \times \{1, \ldots, |\mathsf{sig}(\mathcal{T})^{\pm}_{\mathsf{R}}|\}$ satisfying Conditions 1 and 2 of Definition 2, the circuit decides whether ϱ is a strategy for $\mathcal{K}$ (i.e. Condition 3 holds), and if so, computes the number of matches of q_{S} in interpretations induced by ϱ. Importantly, this can be done without actually building interpretations: in the appendix we give an explicit formula for this number and show it can be computed with a TC^0 circuit. Moreover, the number of strategies depends only on $|\mathcal{T}|$, so is constant w.r.t. data complexity. Finally, the circuit computes the minimum value across strategies and compares it with the input number.

4 Complexity Classification for DL-Lite$^{\mathcal{H}}_{pos}$

In this section, we consider DL-Lite$^{\mathcal{H}}_{pos}$ TBoxes. We show that coNP-hard OMQs exist and prove a complexity trichotomy which precisely delineates the tractability boundary.

We begin by exhibiting a coNP-complete[1] situation.

Example 12. $\mathrm{OMQA}(q_{\mathrm{S}}, \{\mathrm{B} \sqsubseteq \exists\mathrm{R}_1, \mathrm{R}_1 \sqsubseteq \mathrm{S}, \exists\mathrm{R}_1^- \sqsubseteq \exists\mathrm{R}_2, \mathrm{R}_2 \sqsubseteq \mathrm{S}\})$ *is* coNP*-complete. We consider the* NP*-complete* SET COVER *problem: given a set* $\mathcal{U}$*, set of subsets* $\mathcal{S} \subseteq 2^{\mathcal{U}}$ *whose union is* $\mathcal{U}$*, and number* k*, decide whether there exists a* k*-cover, i.e. a subset* $\mathcal{C}$ *of* $\mathcal{S}$ *with* $|\mathcal{C}| \leq k$ *whose union is* $\mathcal{U}$*. We prove that there exists a* k*-cover iff* $[\Sigma_{s \in \mathcal{S}}|s| + k + 1, +\infty]$ *is* not *a certain answer on the following ABox:* $\{\mathrm{B}(\mathsf{u}) \mid u \in \mathcal{U}\} \cup \{\mathrm{S}(\mathsf{u}, \mathsf{s}) \mid u \in s, s \in \mathcal{S}\}$*. Intuitively, from a* k*-cover* $\mathcal{C}$*, we obtain a countermodel in which role* R_1 *contains pairs* (u, s) *such that* $u \in s$ *and* $s \in \mathcal{C}$*, and there is one outgoing* R_2 *role from each* $s \in \mathcal{C}$*.*

The following definition abstracts the preceding example.

Definition 10. *A TBox* $\mathcal{T}$ *admits a* propagation of role W *by a concept* $\mathrm{B} \in \mathsf{sig}(\mathcal{T})^{\pm}_{\mathsf{C}}$ *and roles* $\mathrm{R}_1, \mathrm{R}_2$ *if* $\mathcal{T}$ *entails* $\{\mathrm{B} \sqsubseteq \exists\mathrm{R}_1, \mathrm{R}_1 \sqsubseteq \mathrm{W}, \exists\mathrm{R}_1^- \sqsubseteq \exists\mathrm{R}_2, \mathrm{R}_2 \sqsubseteq \mathrm{W}\}$.

A propagation of S (or S^-) is not sufficient to ensure coNP-hardness: the reduction sketched in Example 12 will fail in the presence of 'interferences', which can be of three types.

Definition 11. *A role* U interferes with the propagation of W by $\mathrm{B}, \mathrm{R}_1, \mathrm{R}_2$ *if it satisfies one of the following conditions:*

1. $\mathcal{T} \models \{\mathrm{B} \sqsubseteq \exists\mathrm{U}, \mathrm{U} \sqsubseteq \mathrm{W}, \mathrm{U} \sqsubseteq \mathrm{W}^-\}$*;*
2. $\mathcal{T} \models \{\exists\mathrm{W}^- \sqsubseteq \exists\mathrm{U}, \mathrm{U} \sqsubseteq \mathrm{W}\}$ *and either* $\mathcal{T} \models \mathrm{U} \sqsubseteq \mathrm{W}^-$ *or* $\mathcal{T} \not\models \mathrm{R}_2 \sqsubseteq \mathrm{W}^-$*;*
3. *if* $\mathrm{B} = \exists\mathrm{T}$ *and* $\mathrm{T} \sqsubseteq \mathrm{W}$*, then* $\mathcal{T} \models \{\exists\mathrm{T}^- \sqsubseteq \exists\mathrm{U}, \mathrm{U} \sqsubseteq \mathrm{W}\}$ *and either* $\mathcal{T} \models \mathrm{U} \sqsubseteq \mathrm{W}^-$ *or* $\mathcal{T} \not\models \mathrm{R}_2 \sqsubseteq \mathrm{W}^-$.

Remarkably, the existence of a propagation without any interfering role (which we call a *non-trivial propagation*) ensures coNP-hardness, while its absence ensures that $\mathrm{OMQA}(q_{\mathrm{S}}, \mathcal{T})$ is in P. We further distinguish two tractable cases, depending on the existence of a *non-trivial pairing*.

Definition 12. *A TBox* $\mathcal{T}$ *admits a* non-trivial pairing of S *if there exist* $\mathrm{B} \in \mathsf{sig}(\mathcal{T})^{\pm}_{\mathsf{C}}$ *and* $\mathrm{R} \in \mathsf{sig}(\mathcal{T})^{\pm}_{\mathsf{R}}$ *such that*

$$\mathcal{T} \models \mathrm{B} \sqsubseteq \exists\mathrm{R} \quad \mathcal{T} \models \mathrm{R} \sqsubseteq \mathrm{S} \quad \mathcal{T} \models \mathrm{R} \sqsubseteq \mathrm{S}^- \quad \mathcal{T} \not\models \mathrm{S} \sqsubseteq \mathrm{S}^-$$

and if $\mathrm{B} = \exists\mathrm{T}$*, then either* $\mathcal{T} \not\models \mathrm{T} \sqsubseteq \mathrm{S}$ *or* $\mathcal{T} \not\models \mathrm{T} \sqsubseteq \mathrm{S}^-$.

[1] A P upper bound for atomic counting queries in DL-Lite$^{\mathcal{H}}_{pos}$ erroneously appears in Table 1 of [Calvanese *et al.*, 2020a], but was corrected in a later arXiv version [Calvanese *et al.*, 2020b].

To formulate our trichotomy result, we recall that a *matching* in a graph $(\mathcal{V}, \mathcal{E})$ is a set of edges that are pairwise vertex-disjoint. The PERFECT MATCHING problem (abbreviated to PM) asks whether there exists a matching such that every vertex is incident to one of its edges. Despite being the focus of intensive research, its exact complexity remains open: in P [Edmonds, 1965] and NL-hard [Chandra *et al.*, 1984].

Theorem 2. *Let $\mathcal{T}$ be a DL-Lite$^{\mathcal{H}}_{pos}$ TBox.* $\mathrm{OMQA}(q_{\mathrm{S}}, \mathcal{T})$ *is* coNP*-complete if $\mathcal{T}$ admits a non-trivial propagation of either* S *or* S^-*, is* L*-equivalent to the complement of* PM *if it does not admit such a non-trivial propagation but admits a non-trivial pairing of* S*, and is in* TC^0 *otherwise.*

Proof sketch. The coNP-hardness proof generalizes the reduction sketched in Example 12. If there is a non-trivial pairing (but no non-trivial propagation), we show that, up to trivial cases solvable in TC^0, the existence of a model with few matches is equivalent to the existence of a large matching between critical individuals. This yields L-equivalence with the MAXIMUM MATCHING decision problem, which is L-equivalent to the better-known PM problem [Rabin and Vazirani, 1989]. TC^0 membership is proven by case analysis, where we exhibit for each case a model with an optimal (and easily computable) number of matches. □

5 First Look at DL-Lite$^{\mathcal{H}}_{core}$

We now turn to DL-Lite$^{\mathcal{H}}_{core}$ and exhibit new situations that are not captured by the preceding complexity classification.

First, we observe that negative concept and role inclusions introduce two new sources of coNP-hardness.

Theorem 3. *For* $\mathcal{T} = \{\mathrm{B} \sqsubseteq \exists\mathrm{U},\ \mathrm{U} \sqsubseteq \mathrm{S},\ \mathrm{C} \sqsubseteq \exists\mathrm{V},\ \mathrm{V} \sqsubseteq \mathrm{S},\ \exists\mathrm{U}^- \sqsubseteq \neg\exists\mathrm{V}^-\ \}$*,* $\mathrm{OMQA}(q_{\mathrm{S}}, \mathcal{T})$ *is* coNP*-complete.*

Proof sketch. Let $(\mathcal{U}, \mathcal{S}, k)$ be an instance of SET COVER, and consider the ABox $\mathcal{A} = \{\mathrm{B}(\mathsf{u}) \mid u \in \mathcal{U}\} \cup \{\mathrm{S}(\mathsf{u}, \mathsf{s}^*) \mid u \in s, s \in \mathcal{S}\} \cup \{\mathrm{C}(\mathsf{s}) \mid s \in \mathcal{S}\} \cup \{\mathrm{S}(\mathsf{s}, \mathsf{s}^*) \mid s \in \mathcal{S}\}$. It can be shown that *no* k-cover exists iff every model of $(\mathcal{T}, \mathcal{A})$ has at least $|\mathcal{S}| + \sum_{s \in \mathcal{S}} |s| + k + 1$ matches. □

Theorem 4. *For* $\mathcal{T} = \{\ \mathrm{B} \sqsubseteq \exists\mathrm{U},\ \mathrm{U} \sqsubseteq \mathrm{S},\ \exists\mathrm{U}^- \sqsubseteq \exists\mathrm{V},\ \mathrm{V} \sqsubseteq \mathrm{S}^-,\ \mathrm{V} \sqsubseteq \neg\mathrm{W}\ \}$*,* $\mathrm{OMQA}(q_{\mathrm{S}}, \mathcal{T})$ *is* coNP*-complete.*

Perhaps more surprising, we show that there exist coNP-hard OMQs based upon concept cardinality queries.

Theorem 5. *For* $\mathcal{T} = \{\mathrm{A} \sqsubseteq \exists\mathrm{U},\ \exists\mathrm{U}^- \sqsubseteq \mathrm{C},\ \mathrm{U} \sqsubseteq \neg\mathrm{U}',\ \mathrm{B} \sqsubseteq \exists\mathrm{V},\ \exists\mathrm{V}^- \sqsubseteq \mathrm{C},\ \mathrm{V} \sqsubseteq \neg\mathrm{V}',\ \exists\mathrm{U}^- \sqsubseteq \neg\exists\mathrm{V}^-\}$*,* $\mathrm{OMQA}(q_{\mathrm{C}}, \mathcal{T})$ *is* coNP*-complete.*

Proof sketch. Hardness is shown by reducing the tautology problem. Three individuals are introduced per propositional variable (one for the variable itself with concept A, two for its possible truth values), as well as one individual per clause (with concept B). Each variable should have a truth value given by U (whose possible values in the ABox are restricted through the use of U$'$), and each clause should have a falsified literal given by V (whose possibles values in the ABox are restricted, according to the input formula, with V$'$). The input formula is a tautology iff every model introduces a new element marked C (as a witness for either $\exists$U or $\exists$V). □

Moreover, we further show that L-complete OMQs exist. The next result employs a role cardinality query, but a similar result can be obtained using a concept cardinality query.

Theorem 6. *For* $\mathcal{T} = \{\ \mathrm{B} \sqsubseteq \exists\mathrm{R},\ \mathrm{R} \sqsubseteq \mathrm{S},\ \mathrm{R} \sqsubseteq \neg\mathrm{R}^-\ \}$*,* $\mathrm{OMQA}(q_{\mathrm{S}}, \mathcal{T})$ *is* L*-complete.*

Proof. Hardness is by reduction from the L-complete problem UNDIRECTED FOREST ACCESSIBILITY (UFA) [Cook and McKenzie, 1987], which takes as input an undirected acyclic graph $(\mathcal{V}, \mathcal{E})$ with two connected components, vertices $s, t \in \mathcal{V}$, and asks if t is reachable from s. We set $\mathcal{A} = \{\mathrm{B}(\mathsf{u}) \mid u \in \mathcal{V}\} \cup \{\mathrm{S}(\mathsf{u}, \mathsf{v}) \mid \{u, v\} \in \mathcal{E}\} \cup \{\mathrm{S}(\mathsf{s}, \mathsf{v}^*), \mathrm{S}(\mathsf{t}, \mathsf{v}^*)\} \cup \{\mathrm{R}(\mathsf{s}, \mathsf{v}^*), \mathrm{R}(\mathsf{t}, \mathsf{v}^*)\}$ and observe that $((\mathcal{V}, \mathcal{E}), s, t) \in$ UFA iff $[2|\mathcal{E}| + 3, +\infty]$ is a certain answer. Indeed, there are $2|\mathcal{E}| + 2$ matches in the ABox, and a further match arises if we add R-atoms to satisfy $\mathrm{B} \sqsubseteq \exists\mathrm{R}$ in a connected component that contains neither s nor t (such a match can be avoided if it contains s or t). For the upper bound, we characterize the minimum number of matches based upon the graph structure of the ABox and show it can be computed in L, by using an oracle for undirected reachability. □

Our results imply that, under standard complexity-theoretic assumptions, at least four different complexities are possible for cardinality queries coupled with DL-Lite$^{\mathcal{H}}_{core}$ ontologies.

6 Conclusion

In this paper, we investigated the complexity of answering cardinality queries in the presence of DL-Lite ontologies. Our study provides several novel insights into the challenge of adopting counting queries in OMQA. On the one hand, we identified new sources of coNP-hardness, showing that even single-atom counting queries can be difficult to handle (which closes some questions about restricted forms of counting queries left open in [Calvanese *et al.*, 2020a]). On the other hand, we exhibited several settings in which cardinality queries can be answered with (sub-)polynomial data complexity; in particular, the problem is in TC^0 when the ontology is formulated in DL-Lite$_{core}$. Interestingly, our tractability results do not rely on the canonical model yielding the minimum number of matches, but instead involve a sophisticated analysis of how to best merge witnesses for existential axioms. Differently from [Kostylev and Reutter, 2015; Calvanese *et al.*, 2020a; Bienvenu *et al.*, 2020], we conducted our complexity analysis on the level of ontology-mediated queries, and notably obtained a full classification of the complexity of OMQs based upon DL-Lite$^{\mathcal{H}}_{pos}$ ontologies.

We find it promising that very low data complexity can be obtained even for settings in which non-trivial optimization is required, and we plan to explore how to extend and adapt our techniques to identify further tractability results for counting queries. Another important topic for future work is to transform our TC^0 procedures into more practical algorithms that are suitable for implementation on top of database systems.

Acknowledgements

This work was partially supported by ANR project CQFD (ANR-18-CE23-0003).

References

[Aehlig *et al.*, 2007] Klaus Aehlig, Stephen A. Cook, and Phuong Nguyen. Relativizing small complexity classes and their theories. In *Proc. of the 21st International Workshop on Computer Science Logic (CSL)*, pages 374–388, 2007.

[Artale *et al.*, 2009] Alessandro Artale, Diego Calvanese, Roman Kontchakov, and Michael Zakharyaschev. The DL-Lite family and relations. *Journal of Artificial Intelligence Research (JAIR)*, 36:1–69, 2009.

[Baader *et al.*, 2017] Franz Baader, Ian Horrocks, Carsten Lutz, and Ulrike Sattler. *An Introduction to Description Logic*. Cambridge University Press, 2017.

[Baget *et al.*, 2011] Jean-François Baget, Michel Leclère, Marie-Laure Mugnier, and Eric Salvat. On rules with existential variables: Walking the decidability line. *Journal of Artificial Intelligence (JAR)*, 175(9-10):1620–1654, 2011.

[Bienvenu and Ortiz, 2015] Meghyn Bienvenu and Magdalena Ortiz. Ontology-mediated query answering with data-tractable description logics. In *Tutorial Lectures of the 11th Reasoning Web International Summer School*, pages 218–307, 2015.

[Bienvenu *et al.*, 2020] Meghyn Bienvenu, Quentin Manière, and Michaël Thomazo. Answering counting queries over DL-Lite ontologies. In *Proc. of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1608–1614, 2020.

[Calì *et al.*, 2012] Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *Journal of Web Semantics (JWS)*, 14:57–83, 2012.

[Calvanese *et al.*, 2007] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *Journal of Automated Reasoning (JAR)*, 39(3):385–429, 2007.

[Calvanese *et al.*, 2008] Diego Calvanese, Evgeny Kharlamov, Werner Nutt, and Camilo Thorne. Aggregate queries over ontologies. In *Proc. of the 2nd International Workshop on Ontologies and Information Systems for the Semantic Web (ONISW)*, pages 97–104, 2008.

[Calvanese *et al.*, 2020a] Diego Calvanese, Julien Corman, Davide Lanti, and Simon Razniewski. Counting query answers over a DL-Lite knowledge base. In *Proc. of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1658–1666, 2020.

[Calvanese *et al.*, 2020b] Diego Calvanese, Julien Corman, Davide Lanti, and Simon Razniewski. Counting query answers over a DL-Lite knowledge base (extended version). *arXiv:2005.05886v3*, 2020.

[Chandra *et al.*, 1984] Ashok K. Chandra, Larry J. Stockmeyer, and Uzi Vishkin. Constant depth reducibility. *SIAM Journal on Computing*, 13(2):423–439, 1984.

[Cook and McKenzie, 1987] Stephen A. Cook and Pierre McKenzie. Problems complete for deterministic logarithmic space. *Journal of Algorithms*, 8(3):385–394, 1987.

[Edmonds, 1965] Jack Edmonds. Paths, trees and flowers. *Canadian Journal of Mathematics*, 17:449–467, 1965.

[Feier *et al.*, 2021] Cristina Feier, Carsten Lutz, and Marcin Przybylko. Answer counting under guarded TGDs. In *Proc. of the 24th International Conference on Database Theory (ICDT)*, 2021.

[Kostylev and Reutter, 2015] Egor V. Kostylev and Juan L. Reutter. Complexity of answering counting aggregate queries over DL-Lite. *Journal of Web Semantics (JWS)*, 33:94–111, 2015.

[Poggi *et al.*, 2008] Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Linking data to ontologies. *Journal of Data Semantics*, 10:133–173, 2008.

[Rabin and Vazirani, 1989] Michael O. Rabin and Vijay V. Vazirani. Maximum matchings in general graphs through randomization. *Journal of Algorithms*, 10(4):557–567, 1989.

[Xiao *et al.*, 2018] Guohui Xiao, Diego Calvanese, Roman Kontchakov, Domenico Lembo, Antonella Poggi, Riccardo Rosati, and Michael Zakharyaschev. Ontology-based data access: A survey. In *Proc. of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5511–5519, 2018.