On Finite and Unrestricted Query Entailment beyond $\mathcal{SQ}$ with Number Restrictions on Transitive Roles

Tomasz Gogacz[1] , **Víctor Gutiérrez-Basulto**[2] , **Yazmín Ibáñez-García**[2] ,
Jean Christoph Jung[3] and **Filip Murlak**[1]
[1]University of Warsaw, Poland
[2]Cardiff University, UK
[3]University of Bremen, Germany
t.gogacz@mimuw.edu.pl, {gutierrezbasultov, ibanezgarciay}@cardiff.ac.uk,
jeanjung@uni-bremen.de, fmurlak@mimuw.edu.pl

Abstract

We study the description logic $\mathcal{SQ}$ with number restrictions applicable to transitive roles, extended with either *nominals* or *inverse roles*. We show tight 2EXPTIME upper bounds for *unrestricted* entailment of *regular path queries* for both extensions and *finite* entailment of *positive existential queries* for nominals. For inverses, we establish 2EXPTIME-completeness for *unrestricted* and *finite* entailment of *instance queries* (the latter under restriction to a single, transitive role).

1 Introduction

A prominent line of research in knowledge representation and database theory has focused on the evaluation of queries over incomplete data enriched by ontologies providing background knowledge. In this paradigm, ontologies are commonly formulated using description logics (DLs), believed to offer a good balance between expressivity and complexity. This is supported, for instance, by the good understanding of 'data-tractable' DLs [Kontchakov and Zakharyaschev, 2014; Bienvenu and Ortiz, 2015]. Yet, for some expressive DLs the complexity of query entailment is less understood.

In this paper, we study query entailment in extensions of the description logic (DL) $\mathcal{SQ}$ allowing number restrictions ($\mathcal{Q}$) to be applied to transitive roles ($\mathcal{S}$). Most previous work on query entailment in expressive DLs, such as $\mathcal{SHIQ}$ or $\mathcal{SHOQ}$, forbid the interaction of number restrictions and transitive roles [Glimm *et al.*, 2008b; Glimm *et al.*, 2008a; Calvanese *et al.*, 2014], but it is required in areas like biomedicine, e.g., to restrict the number of certain parts an organ has. For instance, one can express that the human heart has exactly one mitral valve, which has to be shared by its left and right atrium [Gutiérrez-Basulto *et al.*, 2018]. Allowing for the interaction of $\mathcal{S}$ and $\mathcal{Q}$ is dangerous in the sense that even modest extensions of $\mathcal{SQ}$, such as with role inclusions or inverse roles, lead to an undecidable satisfiability problem [Kazakov *et al.*, 2007]. Decidability of satisfiability in $\mathcal{SQ}$ and in its extension with nominals was shown several years ago [Kazakov *et al.*, 2007; Kaminski and Smolka, 2010], but only recently tight computational complexity bounds were established [Gutiérrez-Basulto *et al.*, 2017]. Even more recently, decidability for entailment of regular path queries over $\mathcal{SQ}$ knowledge bases was established. More precisely, based on a novel *tree-like model property* of $\mathcal{SQ}$ it was possible to devise an automata-based decision procedure yielding a tight 2EXPTIME upper bound [Gutiérrez-Basulto *et al.*, 2018].

The objective of this paper is to provide a more complete picture of query entailment in DLs with number restrictions on transitive roles. We pursue two specific goals.

First, we aim at understanding the limits of decidability of query entailment for such DLs. To this end, we investigate the extensions of $\mathcal{SQ}$ by *nominals ($\mathcal{SOQ}$)* and *controlled inverse roles ($\mathcal{SIQ}^-$)*, where we allow number restrictions on inverse non-transitive roles and only existential restrictions on inverse transitive roles. As query language, we consider *positive existential regular path queries*, thus capturing the common languages of conjunctive and regular path queries.

Our second aim is to initiate the study of *finite* query entailment for $\mathcal{SIQ}^-$ and $\mathcal{SOQ}$, where one is interested in reasoning only over finite models. This distinction is crucial because in database applications, both database instances and the models they represent are commonly assumed to be finite. The study of finite query entailment in $\mathcal{SQ}$ is interesting since, due to the presence of transitivity, $\mathcal{SQ}$ lacks *finite controllability*, and therefore unrestricted and finite entailment do not coincide. Interestingly, most previous works on finite query entailment consider logics lacking finite controllability because of number restrictions and inverse roles [Rosati, 2008; Pratt-Hartmann, 2009; Ibáñez-García *et al.*, 2014; Amarilli and Benedikt, 2015]. The study of finite query entailment in logics with transitivity (without number restrictions on transitive roles) started only recently [Rudolph, 2016; Gogacz *et al.*, 2018; Danielski and Kieronski, 2018]. Here, we focus on finite entailment of positive existential queries in $\mathcal{SOQ}$ and of instance queries in $\mathcal{SIQ}^-$.

Our main contributions are as follows. In Sect. 3, we start by showing a *tree-like model property* for both $\mathcal{SOQ}$ and $\mathcal{SIQ}^-$. More specifically, we carefully extend and adapt the *canonical* tree decompositions that were introduced for $\mathcal{SQ}$ in previous work [Gutiérrez-Basulto *et al.*, 2018] to also incorporate the presence of controlled inverses and nominals. Next, we prove that if a query is not entailed by a knowledge base (KB), then there is a counter-model with a canonical tree decomposition of small width. This tree-like model property is the basis for automata-based approaches to unrestricted and finite query entailment in the remainder of the paper. First, in Sect. 4, we construct tree automata to optimally decide entailment of regular path queries over $\mathcal{SOQ}$ and $\mathcal{SIQ}^-$ KBs in 2ExpTime. We move then, in Sect. 5, to finite entailment of positive existential queries over $\mathcal{SOQ}$ KBs, showing again an optimal 2ExpTime upper bound. To this end, we look at more refined canonical tree decompositions, which ensure the existence of a finite counter model. In other words, we reduce finite query entailment to entailment over models with this special canonical tree decomposition. Finally, in Sect. 6, we investigate the complexity for unrestricted and finite instance query (IQ) entailment in $\mathcal{SIQ}^-$. In particular, we show that IQ entailment is 2ExpTime-hard both in the finite and in the unrestricted case. We found this surprising since it is rarely the case that IQ entailment becomes more difficult when inverses are added to the logic. Moreover, the result provides an orthogonal reason for 2ExpTime-hardness for conjunctive query entailment in $\mathcal{SIQ}^-$ [Lutz, 2008]. We complement this lower bound with matching upper bounds in the unrestricted case, thus confirming the conjecture that satisfiability in $\mathcal{SIQ}^-$ is decidable [Kazakov *et al.*, 2007]. In the finite case, we show a 2ExpTime-upper bound for KBs using a single transitive role. Note that $\mathcal{SIQ}^-$ with a single transitive role is a notational variant of the *graded modal logic with converse* $\mathbf{K4}(\Diamond_{\geq}, \Diamond^-)$. Thus, our result entails 2ExpTime-completeness for global consequence in $\mathbf{K4}(\Diamond_{\geq}, \Diamond^-)$, which was only known to be decidable [Bednarczyk *et al.*, 2019].

A long version with appendix can be found under http://www.informatik.uni-bremen.de/tdki/research/papers.html.

2 Preliminaries

Description Logics

We consider a vocabulary consisting of countably infinite disjoint sets of *concept names* $\mathsf{N_C}$, *role names* $\mathsf{N_R}$, and *individual names* $\mathsf{N_I}$, and assume that $\mathsf{N_R}$ is partitioned into two infinite sets of *non-transitive role names* $\mathsf{N_R^{nt}}$ and *transitive role names* $\mathsf{N_R^{t}}$. A *role* is a role name or an *inverse role* r^-; a *transitive role* is a transitive role name or the inverse of one. $\mathcal{SIQ}^-$*-concepts* C, D are defined by the grammar

$$C, D ::= A \mid \neg C \mid C \sqcap D \mid \exists r.C \mid (\leqslant n\ s\ C)$$

where $A \in \mathsf{N_C}$, r is a role, $n \geq 0$ is a natural number given in binary, and s is either a non-transitive role or a transitive role name. $\mathcal{SOQ}$*-concepts* C, D are defined by the grammar

$$C, D ::= A \mid \neg C \mid C \sqcap D \mid \{a\} \mid (\leqslant n\ r\ C)$$

where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, $a \in \mathsf{N_I}$ and n is as above. We will use $(\geqslant n\ r\ C)$ as abbreviation for $\neg(\leqslant n{-}1\ r\ C)$, together with standard abbreviations $\bot, \top, C \sqcup D, \forall r.C$. Concepts of the form $(\leqslant n\ r\ C)$, $(\geqslant n\ r\ C)$, and $\{a\}$ are called *at-most restrictions*, *at-least restrictions*, and *nominals*, respectively. Note that in $\mathcal{SIQ}^-$ concepts, *inverse transitive roles* are not allowed in at-most and at-least restrictions.

A $\mathcal{SIQ}^-$*-TBox (respectively,* $\mathcal{SOQ}$*-TBox)* $\mathcal{T}$ is a finite set of *concept inclusions (CIs)* $C \sqsubseteq D$, where C, D are $\mathcal{SIQ}^-$-concepts (respectively, $\mathcal{SOQ}$-concepts). An *ABox* $\mathcal{A}$ is a finite non-empty set of *concept* and *role assertions* of the form $A(a)$, $r(a, b)$ where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$ and $\{a, b\} \subseteq \mathsf{N_I}$; $\mathsf{ind}(\mathcal{A})$ is the set of individual names occurring in $\mathcal{A}$. A *knowledge base (KB)* is a pair $\mathcal{K} = (\mathcal{T}, \mathcal{A})$; $\mathsf{nom}(\mathcal{K})$ is the set of nominals occurring in $\mathcal{K}$ and $\mathsf{ind}(\mathcal{K}) = \mathsf{ind}(\mathcal{A}) \cup \mathsf{nom}(\mathcal{K})$.

Without loss of generality, we assume throughout the paper that all CIs are in one of the following *normal forms*:

$$\textstyle\bigsqcap_i A_i \sqsubseteq \bigsqcup_j B_j, \quad A \sqsubseteq \forall r^-.B, \quad A \sqsubseteq \exists r^-.B,$$
$$A \sqsubseteq (\leqslant n\ s\ B), \quad A \sqsubseteq (\geqslant n\ s\ B),$$

where A, A_i, B, B_j are concept names or nominals, $r \in \mathsf{N_R}$, s is a non-transitive role or a transitive role name, and empty disjunction and conjunction are equivalent to $\bot$ and $\top$, respectively. We further assume that for every at-most and at-least restriction, $\mathcal{T}$ contains an equivalent concept name.

Interpretations

The semantics is given as usual via *interpretations* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ consisting of a non-empty *domain* $\Delta^\mathcal{I}$ and an *interpretation function* $\cdot^\mathcal{I}$ mapping concept names to subsets of the domain and role names to binary relations over the domain. Further, we adopt the *standard name assumption*, i.e., $a^\mathcal{I} = a$ for all $a \in \mathsf{N_I}$. The interpretation of complex concepts C is defined in the usual way [Baader *et al.*, 2017]. An interpretation $\mathcal{I}$ is a *model of a TBox* $\mathcal{T}$, written $\mathcal{I} \models \mathcal{T}$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$ for all CIs $C \sqsubseteq D \in \mathcal{T}$. It is a *model of an ABox* $\mathcal{A}$, written $\mathcal{I} \models \mathcal{A}$, if $(a, b) \in r^\mathcal{I}$ for all $r(a, b) \in \mathcal{A}$ and $a \in A^\mathcal{I}$ for all $A(a) \in \mathcal{A}$. Finally, $\mathcal{I}$ is a *model of a KB* $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, written $\mathcal{I} \models \mathcal{K}$, if $\mathcal{I} \models \mathcal{T}$, $\mathcal{I} \models \mathcal{A}$, and $r^\mathcal{I}$ is transitive for all $r \in \mathsf{N_R^t}$ occurring in $\mathcal{K}$. If $\mathcal{K}$ has a model, we say that it is *satisfiable*.

An interpretation $\mathcal{I}'$ is a *sub-interpretation* of $\mathcal{I}$, written as $\mathcal{I}' \subseteq \mathcal{I}$, if $\Delta^{\mathcal{I}'} \subseteq \Delta^\mathcal{I}$, $A^{\mathcal{I}'} \subseteq A^\mathcal{I}$, and $r^{\mathcal{I}'} \subseteq r^\mathcal{I}$ for all $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. For $\Sigma \subseteq \mathsf{N_C} \cup \mathsf{N_R}$, $\mathcal{I}$ is a Σ*-interpretation* if $A^\mathcal{I} = \emptyset$ and $r^\mathcal{I} = \emptyset$ for all $A \in \mathsf{N_C} \setminus \Sigma$ and $r \in \mathsf{N_R} \setminus \Sigma$. The *restriction of* $\mathcal{I}$ *to signature* Σ is the maximal Σ-interpretation $\mathcal{I}'$ with $\mathcal{I}' \subseteq \mathcal{I}$. The *restriction of* $\mathcal{I}$ *to domain* Δ is the maximal sub-interpretation of $\mathcal{I}$ with domain Δ. The union $\mathcal{I} \cup \mathcal{J}$ of $\mathcal{I}$ and $\mathcal{J}$ is an interpretation such that $\Delta^{\mathcal{I}\cup\mathcal{J}} = \Delta^\mathcal{I} \cup \Delta^\mathcal{J}$, $A^{\mathcal{I}\cup\mathcal{J}} = A^\mathcal{I} \cup A^\mathcal{J}$, and $r^{\mathcal{I}\cup\mathcal{J}} = r^\mathcal{I} \cup r^\mathcal{J}$ for all $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. The *transitive closure* $\mathcal{I}^*$ of $\mathcal{I}$ is an interpretation such that $\Delta^{\mathcal{I}^*} = \Delta^\mathcal{I}$, $A^{\mathcal{I}^*} = A^\mathcal{I}$ for all $A \in \mathsf{N_C}$, $r^{\mathcal{I}^*} = r^\mathcal{I}$ for all $r \in \mathsf{N_R^{nt}}$, and $r^{\mathcal{I}^*} = (r^\mathcal{I})^+$ for all $r \in \mathsf{N_R^t}$.

A *tree decomposition* $\mathfrak{T}$ *of an interpretation* $\mathcal{I}$ is a pair $(T, \mathfrak{I})$ where T is a tree and $\mathfrak{I}$ is a function that assigns an interpretation $\mathfrak{I}(w) = (\Delta_w, \cdot^{\mathfrak{I}(w)})$ to each $w \in T$ such that $\mathcal{I} = \bigcup_{w \in T} \mathfrak{I}(w)$ and for every $d \in \Delta^\mathcal{I}$, the set $\{w \in T \mid d \in \Delta_w\}$ is connected in T. We often blur the distinction between a node w of T and the associated interpretation $\mathfrak{I}(w)$, using the term *bag* for both. The *width* of $\mathfrak{T}$ is $\sup_{w \in T} |\Delta_w| - 1$; the *outdegree* of $\mathfrak{T}$ is the outdegree of T. For each $d \in \Delta^\mathcal{I}$, there is a unique bag w closest to the *root* ε such that $d \in \Delta_w$. We say that d is *fresh* in this bag, and write $F(w)$ for the set of all elements fresh in w.

Ontology-mediated Query Entailment

A *positive existential regular path query (PRPQ)* is a first-order formula $\varphi = \exists \mathbf{x}\, \psi(\mathbf{x})$ with $\psi(\mathbf{x})$ constructed using $\wedge$ and $\vee$ over atoms of the form $\mathcal{E}(t,t')$ where t, t' are variables from $\mathbf{x}$ or individual names from $\mathsf{N_I}$, and $\mathcal{E}$ is a *path expression* defined by the grammar

$$\mathcal{E}, \mathcal{E}' ::= r \mid r^- \mid A? \mid \mathcal{E}^* \mid \mathcal{E} \cup \mathcal{E}' \mid \mathcal{E} \circ \mathcal{E}',$$

where $r \in \mathsf{N_R}$ and $A \in \mathsf{N_C}$. A PEQ is a PRPQ that does not use the operators *, $\cup$, and $\circ$ in path expressions. Equivalently, it is an FO formula $\varphi = \exists \mathbf{x}\, \psi(\mathbf{x})$ where ψ is constructed using $\wedge$ and $\vee$ over atoms $r(t,t')$ and $A(t')$ with t, t' as above. An *instance query (IQ)* is just an expression of the shape $C(a)$ for some concept C and $a \in \mathsf{N_I}$.

The semantics of PRPQs is defined via matches. Let us fix a PRPQ $\varphi = \exists \mathbf{x}\, \psi(\mathbf{x})$ and an interpretation $\mathcal{I}$. Let $\mathsf{ind}(\varphi)$ be the set of individual names in φ. A *match for φ in $\mathcal{I}$* is a function $\pi : \mathbf{x} \cup \mathsf{ind}(\varphi) \to \Delta^\mathcal{I}$ such that $\pi(a) = a$, for all $a \in \mathsf{ind}(\varphi)$, and $\mathcal{I}, \pi \models \psi(\mathbf{x})$ under the standard semantics of first-order logic extended with a rule for atoms of the form $\mathcal{E}(t,t')$. An interpretation $\mathcal{I}$ satisfies φ, written as $\mathcal{I} \models \varphi$, if there is a match for φ in $\mathcal{I}$.

A PRPQ φ is *(finitely) entailed by a KB $\mathcal{K}$*, if $\mathcal{I} \models \varphi$ for every (finite) model $\mathcal{I}$ of $\mathcal{K}$; we write $\mathcal{K} \models \varphi$ and $\mathcal{K} \models_{\mathsf{fin}} \varphi$, respectively, in this case. Accordingly, we write $\mathcal{K} \models C(a)$ and $\mathcal{K} \models_{\mathsf{fin}} C(a)$ if $a \in C^\mathcal{I}$ in all (finite) models $\mathcal{I}$ of $\mathcal{K}$.

We study the corresponding *decision problem*—whether a given query is (finitely) entailed by a given KB—for different choices of knowledge base and query languages.

3 Tree-like Counter-Model Property

In this section we show a tree-like model property for $\mathcal{SIQ}^-$ and $\mathcal{SOQ}$: we show that if a query is not entailed by a KB, then there is a counter-model with a tree decomposition of bounded width and outdegree. For the automata-based decision procedure to yield optimal upper bounds, it is useful to consider *canonical* decompositions which we define next.

In canonical decompositions elements will be accompanied by certain key neighbors. Let us fix a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. For an interpretation $\mathcal{I}$, an element $d \in \Delta^\mathcal{I}$, and $r \in \mathsf{N}^t_\mathsf{R}$, the *$r$-cluster of d in $\mathcal{I}$*, denoted by $Q^\mathcal{I}_r(d)$, is the set containing d and each $e \in \Delta^\mathcal{I}$ such that both $(d,e) \in r^\mathcal{I}$ and $(e,d) \in r^\mathcal{I}$. This is the closest environment of d wrt. r. We also associate with d a larger set $\mathsf{rel}^\mathcal{I}_r(d)$ of r-successors *relevant for the at-most restrictions of $\mathcal{K}$*. We let $\mathsf{rel}^\mathcal{I}_r(d)$ be the least set X such that $Q^\mathcal{I}_r(d) \subseteq X$ and for all $e \in X$, $f \in \Delta^\mathcal{I}$, and $A \sqsubseteq (\leqslant n\ r\ B)$ in $\mathcal{T}$, if $e \in A^\mathcal{I}$, $f \in B^\mathcal{I}$, and $(e,f) \in r^{\mathcal{I}^*}$, then $Q^\mathcal{I}_r(f) \subseteq X$. The construction of canonical decompositions relies on the following properties of relevant successors.

Lemma 1. *For each $r \in \mathsf{N}^\mathsf{t}_\mathsf{R}$, the following hold:*

1. *for all $d, e \in \Delta^\mathcal{I}$, if $e \in \mathsf{rel}^\mathcal{I}_r(d)$ then $\mathsf{rel}^\mathcal{I}_r(e) \subseteq \mathsf{rel}^\mathcal{I}_r(d)$;*
2. *if each r-cluster in $\mathcal{I}$ has size at most N, then for each $d \in \Delta^\mathcal{I}$, $|\mathsf{rel}^\mathcal{I}_r(d)| \leq N \cdot 2^{\mathsf{poly}(|\mathcal{T}|)}$.*

In a canonical tree decomposition, formalized in Definition 1 below, each non-root bag keeps track of all concepts and a single role indicated by $\mathfrak{r}$. Nominals are captured within a finite subinterpretation $\mathcal{M}$ represented faithfully in all bags; in the absence of nominals, one can take empty $\mathcal{M}$ and drop (C_4). Conditions (C_0)–(C_3) ensure that apart from $\Delta^\mathcal{M}$, neighboring non-root bags share a single element, sometimes accompanied by its relevant successors.

Definition 1. *A tree decomposition $\mathfrak{T} = (T, \mathfrak{I})$ is* canonical *if there exists $\mathfrak{r} \colon T \to \mathsf{N_R} \cup \{\bot\}$ with $\mathfrak{r}^{-1}(\bot) = \{\varepsilon\}$ such that*

(B_0) for each $w \in T$, $\mathfrak{I}(w)$ is a Σ_w-interpretation where $\Sigma_\varepsilon = \mathsf{N_C} \cup \mathsf{N}^\mathsf{nt}_\mathsf{R}$ and $\Sigma_w = \mathsf{N_C} \cup \{\mathfrak{r}(w)\}$ for $w \neq \varepsilon$;

(B_1) for all $v, w \in T$, the restrictions of $\mathfrak{I}(v)$ and $\mathfrak{I}(w)$ to domain $\Delta_v \cap \Delta_w$ and signature $\Sigma_v \cap \Sigma_w$ coincide;

(B_2) for each $v \in T \setminus \{\varepsilon\}$, $d \in F(v)$, and $r \in \mathsf{N}^\mathsf{t}_\mathsf{R} \setminus \{\mathfrak{r}(v)\}$, a unique child w of v satisfies $\mathfrak{r}(w) = r$ and $d \in \Delta_w$;

and there is an interpretation $\mathcal{M}$ with $\mathsf{nom}(\mathcal{K}) \subseteq \Delta^\mathcal{M} \subseteq \Delta_\varepsilon$, such that for each $w \in T \setminus \{\varepsilon\}$ and its parent v, one has

(C_0) if $\mathfrak{r}(w) \in \mathsf{N}^\mathsf{t}_\mathsf{R}$ and $v = \varepsilon$, then $\Delta_\varepsilon \subseteq \Delta_w$;

(C_1) if $\mathfrak{r}(w) \in \mathsf{N}^\mathsf{nt}_\mathsf{R}$, then $\Delta_v \cap \Delta_w = \{d\} \cup \Delta^\mathcal{M}$ for some $d \in F(v)$;

(C_2) if $\mathfrak{r}(v) \neq \mathfrak{r}(w) \in \mathsf{N}^\mathsf{t}_\mathsf{R}$ and $v \neq \varepsilon$, then $\Delta_w \cap \Delta_v = \{d\} \cup \Delta^\mathcal{M}$ for some $d \in F(v)$;

(C_3) if $\mathfrak{r}(w) = \mathfrak{r}(v) = r \in \mathsf{N}^\mathsf{t}_\mathsf{R}$, then $\Delta_v \cap \Delta_w = \mathsf{rel}^{\mathfrak{I}(v)}_r(d) \cup \Delta^\mathcal{M}$ and $\mathsf{rel}^{\mathfrak{I}(v)}_r(d) = \mathsf{rel}^{\mathfrak{I}(w)}_r(d)$ for some d such that either $d \in F(v)$ or $d \in F(u)$ and $\mathfrak{r}(u) \neq \mathfrak{r}(v)$ for the parent u of v; and

(C_4) if $\mathfrak{r}(w) = r \in \mathsf{N}^\mathsf{t}_\mathsf{R}$, then $\mathsf{rel}^{\mathfrak{I}(w)}_r(d) = \mathsf{rel}^\mathcal{M}_r(d)$ for all $d \in \Delta^\mathcal{M}$.

Theorem 1. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a KB in normal form and φ a PRPQ with $\mathcal{K} \not\models \varphi$. If $\mathcal{K}$ is a $\mathcal{SOQ}$ KB or a $\mathcal{SIQ}^-$ KB, then there exists a model $\mathcal{J}$ of $\mathcal{T}$ and $\mathcal{A}$ such that*

- *$\mathcal{J}$ has a canonical tree decomposition of width and outdegree $\mathsf{poly}(|\mathsf{ind}(\mathcal{K})|) \cdot 2^{\mathsf{poly}(|\mathcal{T}|)}$; and*
- *$\mathcal{J}^* \models \mathcal{K}$ and $\mathcal{J}^* \not\models \varphi$.[1]*

Proof. Let us fix a counter-model $\mathcal{I}$ for $\mathcal{K}$ and φ. We can assume that $|Q^\mathcal{I}_r(d)| \leq |\mathsf{ind}(\mathcal{K})| + 2^{\mathsf{poly}(|\mathcal{T}|)}$ for all $d \in \Delta^\mathcal{I}$ [Gutiérrez-Basulto *et al.*, 2018]. By Lemma 1, $|\mathsf{rel}^\mathcal{I}_r(d)| \leq |\mathsf{ind}(\mathcal{K})| \cdot 2^{\mathsf{poly}(|\mathcal{T}|)}$ for all $d \in \Delta^\mathcal{I}$.

To build a canonical tree decomposition $\mathfrak{T}$, we unravel $\mathcal{I}$ starting from the interpretation of the ABox and then applying the extension rules ($\mathbf{R}_0$)–($\mathbf{R}_3$) below, corresponding to conditions (C_0)–(C_3): ($\mathbf{R}_0$) collects relevant successors of the individuals in the ABox, ($\mathbf{R}_1$) performs standard unraveling of non-transitive roles, ($\mathbf{R}_2$) takes care of the change of roles, and ($\mathbf{R}_3$) realizes further unraveling of transitive roles.

More precisely, for the root bag, we take $\mathcal{I}$ restricted to the domain $\mathsf{ind}(\mathcal{A}) \cup \Delta$ and the signature $\mathsf{N_C} \cup \mathsf{N}^\mathsf{nt}_\mathsf{R}$, where Δ is the union of $\mathsf{rel}^\mathcal{I}_r(a)$ for all $a \in \mathsf{nom}(\mathcal{K})$ and $r \in \mathsf{N}^\mathsf{t}_\mathsf{R}$.

($\mathbf{R}_0$) For each $r \in \mathsf{N}^\mathsf{t}_\mathsf{R}$, we add as a child bag of ε the restriction of $\mathcal{I}$ to signature $\mathsf{N_C} \cup \{r\}$ and domain $\bigcup_{a \in \mathsf{ind}(\mathcal{A})} \mathsf{rel}^\mathcal{I}_r(a) \cup \Delta$ with each $e \notin \mathsf{ind}(\mathcal{A}) \cup \Delta$ replaced by a fresh copy e'. We call e the *original of e'*.

[1] Recall that in a model of the ABox or the TBox, the extensions of role names from $\mathsf{N}^\mathsf{t}_\mathsf{R}$ need not be transitive.

Then, we use the following rules $(\mathbf{R}_1)$–$(\mathbf{R}_3)$ *ad infinitum*, applying each rule only once to each previously added bag v.

$(\mathbf{R}_1)$ For each $r \in \mathsf{N}_\mathsf{R}^{\mathsf{nt}}$, and each $d' \in F(v)$, let $d \in \Delta^\mathcal{I}$ be the original of d' (possibly $d = d'$) and let W_0 be the set of originals of all r-successors and r-predecessors of d' in $\mathfrak{I}(v)$. Pick a minimal set $W \subseteq \Delta^\mathcal{I}$ containing $\{d\} \cup W_0 \cup \Delta$ such that for each $s \in \{r, r^-\}$ and $A \sqsubseteq (\geqslant n\ s\ B)$ in $\mathcal{T}$, if $d \in A^\mathcal{I}$, then d has at least n different s-successors in $B^\mathcal{I} \cap W$. For each $e \in W \setminus (W_0 \setminus \Delta)$, add as a child bag of v the restriction of $\mathcal{I}$ to signature $\mathsf{N}_\mathsf{C} \cup \{r\}$ and domain $\{d, e\} \cup \Delta$ with all r-edges from $\Delta \setminus \{d\}$ to $\{d, e\} \setminus \Delta$ removed, d replaced by d' and each $f \in \{e\} \setminus \Delta$, by a fresh copy f'.

$(\mathbf{R}_2)$ Assuming $v \neq \varepsilon$, for each $r \in \mathsf{N}_\mathsf{R}^\mathsf{t}$ with $r \neq \mathfrak{r}(v)$, and each $d' \in F(v)$, let d be the original of d'. Add as a child bag of v the restriction of $\mathcal{I}$ to signature $\mathsf{N}_\mathsf{C} \cup \{r\}$ and domain $\mathsf{rel}_r^\mathcal{I}(d) \cup \Delta$ where d is replaced by d' and each $e \in \mathsf{rel}_r^\mathcal{I}(d) \setminus (\{d\} \cup \Delta)$, by a fresh copy e'.

$(\mathbf{R}_3)$ Assuming $\mathfrak{r}(v) = r \in \mathsf{N}_\mathsf{R}^\mathsf{t}$, for each $d' \in \Delta^v$ fresh in v or in the parent u of v with $\mathfrak{r}(u) \neq r$, let d be the original of d'. Pick a minimal set $W \subseteq \Delta^\mathcal{I}$ containing $\mathsf{rel}_r^\mathcal{I}(d) \cup \Delta$ such that for each $A \sqsubseteq (\geqslant n\ r\ B)$ in $\mathcal{T}$, if $d \in A^\mathcal{I}$, then d has at least n different r-successors in $B^\mathcal{I} \cap W$, and for each $A \sqsubseteq \exists r^-.B$ in $\mathcal{T}$, if $d \in A^\mathcal{I}$, then d has an r^--successor in $B^\mathcal{I} \cap W$. For each $e \in W \setminus (\mathsf{rel}_r^\mathcal{I}(d) \cup \Delta)$, add as a child bag of v the restriction of $\mathcal{I}$ to the signature $\mathsf{N}_\mathsf{C} \cup \{r\}$ and domain $\mathsf{rel}_r^\mathcal{I}(e) \cup \mathsf{rel}_r^\mathcal{I}(d) \cup \Delta$ where each element $f \in \mathsf{rel}_r^\mathcal{I}(d) \setminus \Delta$ is replaced by its copy f' from $\mathfrak{I}(v)$, and each element $f \in \mathsf{rel}_r^\mathcal{I}(e) \setminus (\mathsf{rel}_r^\mathcal{I}(d) \cup \Delta)$ by a fresh copy f'.

Let $\mathcal{J}$ be the interpretation underlying the resulting decomposition $\mathfrak{T}$. The function mapping each $d' \in \Delta^\mathcal{J}$ to its original $d \in \Delta^\mathcal{I}$ gives a homomorphism from $\mathcal{J}$ to $\mathcal{I}$, and consequently also from $\mathcal{J}^*$ to $\mathcal{I}$. It follows that $\mathcal{J}^* \not\models \varphi$. Taking $\mathcal{I}$ restricted to Δ as $\mathcal{M}$, it is routine to check that $\mathfrak{T}$ and $\mathcal{J}$ satisfy the remaining postulated properties. Note that while the construction is described for any normalized $\mathcal{K}$, $(\mathbf{R}_1)$ is correct only if $\mathcal{K}$ is either a $\mathcal{SOQ}$ KB or a $\mathcal{SIQ}^-$ KB. Correctness of $(\mathbf{R}_2)$ and $(\mathbf{R}_3)$ follows from Lemma 1 (1). □

4 PRPQ Entailment for $\mathcal{SIQ}^-$ and $\mathcal{SOQ}$

We shall now exploit canonicity of tree decompositions in an automata-based decision procedure for query entailment in $\mathcal{SIQ}^-$ and $\mathcal{SOQ}$, yielding optimal complexity upper bounds.

Let us fix a ($\mathcal{SIQ}^-$ or $\mathcal{SOQ}$) KB $\mathcal{K}$ and a PRPQ φ, and denote with $\Sigma_\mathsf{C}, \Sigma_\mathsf{R}^t, \Sigma_\mathsf{R}^{nt}$ the concept names, transitive role names, and non-transitive role names used in $\mathcal{K}$. By Theorem 1, if φ is not entailed by $\mathcal{K}$, there exists a counter-model admitting a canonical tree decomposition of width and outdegree bounded by a constant N single exponential in $|\mathcal{K}|$. We effectively construct a non-deterministic tree automaton recognizing such decompositions of counter-models, and thus reduce query entailment to the emptiness problem.

Let us introduce the necessary notions for tree automata. A *k-ary Ω-labeled tree* is a pair (T, τ) where T is a tree each of whose nodes has at most k successors and $\tau : T \to \Omega$ assigns a letter from Ω to each node. A *non-deterministic tree automaton (NTA)* over k-ary Ω-labeled trees is a tuple $\mathfrak{A} = (Q, \Omega, q_0, \Lambda)$, where Q is a finite set of states, $q_0 \in Q$ is the initial state, $\Lambda \subseteq \bigcup_{i \leq k}(Q \times \Omega \times Q^i)$ is a set of transitions. A *run* r on a k-ary Ω-labeled tree (T, τ) is a Q-labeled tree (T, r) such that $r(\varepsilon) = q_0$ and, for every $x \in T$ with successors $x_1, \ldots, x_m$, there is a transition $(r(x), \tau(x), r(x_1) \cdots r(x_m)) \in \Lambda$. As usual, $\mathfrak{A}$ *recognizes* the set of all Ω-labeled trees admitting a run.

Since counter-models have a potentially infinite domain, we *encode* tree decompositions of width N using a domain D of $2N$ elements, similar to what has been done, e.g., in [Grädel and Walukiewicz, 1999]. Intuitively, if w is a successor node of v in the tree decomposition, then an element d occurring in (the bag at) w represents a fresh domain element iff d does not occur in v. More precisely, the alphabet Ω of the automaton is the set of all pairs $(x, \mathcal{I})$ such that either $x \in \Sigma_\mathsf{R}$ and $\mathcal{I}$ is a $\Sigma_\mathsf{C} \cup \{x\}$-interpretation with $\Delta^\mathcal{I} \subseteq D$, or $x = \bot$ and $\mathcal{I}$ is a $\Sigma_\mathsf{C} \cup \Sigma_\mathsf{R}^{nt}$-interpretation with $\Delta^\mathcal{I} \subseteq D$.

Lemma 2. *Given $\mathcal{K}$, φ, and N, one can compute in time $O(2^{\mathsf{poly}(N)})$ an NTA recognizing the set of encodings of canonical tree decompositions of width and outdegree at most N such that for the underlying interpretation $\mathcal{J}$ it holds that $\mathcal{J}^* \models \mathcal{K}$ and $\mathcal{J}^* \not\models \varphi$, as well as $\mathcal{J} \models \mathcal{A}$ and $\mathcal{J} \models \mathcal{T}$.*

Proof. The desired NTA is the intersection of an NTA $\mathfrak{A}_\mathcal{K}$ recognizing all canonical tree decompositions such that the underlying interpretation $\mathcal{J}$ satisfies $\mathcal{J} \models \mathcal{A}$, $\mathcal{J} \models \mathcal{T}$, and $\mathcal{J}^* \models \mathcal{K}$ and an NTA $\mathfrak{A}_{\neg\varphi}$ recognizing all tree decompositions of counter-models of φ. Since the latter is known from [Gutiérrez-Basulto *et al.*, 2018, Lemma 6], we concentrate on $\mathfrak{A}_\mathcal{K} = (Q, \Omega, q_0, \Lambda)$, working over N-ary trees.

Informally, its construction relies on the following ideas: *(i)* by (B_2) and (C_2), in every bag there is at most one d satisfying the condition '$d \in F(u) \ldots$' in Condition (C_3); thus, *(ii)* canonicity can be checked by initially guessing $\mathcal{M}$ and then comparing neighboring interpretations and remembering the mentioned d in the states; *(iii)* $\mathcal{J} \models \mathcal{A}$ can be verified by looking at labels of the root and its direct successors; *(iv)* due to canonicity and the TBox normal form, $\mathcal{J} \models \mathcal{T}$ can be verified by looking at the current label (this suffices for at-most restrictions over transitive roles, due to canonicity) and possibly at successor bags (at-least restrictions, and at-most restrictions over non-transitive roles); *(v)* $\mathcal{J}^* \models \mathcal{T}$ is a consequence of $\mathcal{J} \models \mathcal{T}$, by the normal form.

Formally, the set Q contains q_0 and all tuples of the shape

$$\langle (x, \mathcal{I}), F, \mathcal{M}, \mathcal{B}, \mathcal{C}, e, r, f \rangle,$$

where $(x, \mathcal{I}) \in \Omega$, $F \subseteq \Delta^\mathcal{I}$, $\mathcal{M}$ is a $\Sigma_\mathsf{C} \cup \Sigma_\mathsf{R}^{nt}$-interpretation with $\Delta^\mathcal{M} \subseteq D$, $\mathcal{B} \subseteq \mathcal{A}$, $\mathcal{C}$ is a set of assertions of the shape $(\geqslant n\ s\ B)(d)$, $(\leqslant n\ s\ B)(d)$, or $(\exists s.B)(d)$ with $d \in D$, $B \in \Sigma_\mathsf{C}$, s a role in $\mathcal{T}$, $n \leq N$, and $e, f \in D \cup \{\varepsilon\}$, $r \in \Sigma_\mathsf{R}^t$.

In state $q = \langle (x, \mathcal{I}), F, \mathcal{M}, \mathcal{B}, \mathcal{C}, e, r, f \rangle$ reading symbol $a = (x', \mathcal{I}')$, the automaton allows a transition only in case the following conditions are satisfied:

- Conditions (B_0)–(B_2) and (C_0)–(C_4) with $\mathcal{I}, \mathcal{I}', x, x', F$ taking the role of $\mathfrak{I}(v), \mathfrak{I}(w), \mathfrak{r}(v), \mathfrak{r}(w), F(v)$, respectively, and '$d \in F(u) \ldots$' in (C_3) replaced with '$d = f$';
- $x' \neq \bot$ and, if $x = \bot$, then $\mathcal{I}' \models \mathcal{B}$;
- either $e \neq \varepsilon$, $e \in \Delta^{\mathcal{I}'}$, and $r = x'$, or $e = \varepsilon$ and $x = x'$;
- $\mathcal{I}' \models C(d)$ for all $C(d) \in \mathcal{C}$ with $d \in \Delta^{\mathcal{I}'}$ or C of shape $(\geqslant n\ s\ B)$ or $\exists r^-.B$;

- $\mathcal{I}' \models \alpha$ for all $\alpha \in \mathcal{T}$ of the form $\bigsqcap_i A_i \sqsubseteq \bigsqcap_j B_j$, $A \sqsubseteq (\leqslant n\ r\ B)$, and $A \sqsubseteq \forall r^-.B$.

In this case, Λ allows all transitions $(q, a, q_1 \cdots q_m)$, $m \leq N$ where each q_i is of shape $\langle (x', \mathcal{I}'), F', \mathcal{M}, \emptyset, \mathcal{C}_i, e_i, r_i, f_i \rangle$ with $F' = \Delta^{\mathcal{I}'} \setminus \Delta^{\mathcal{I}}$ and:

- for each $d \in F'$ and each $r \in \Sigma^t_{\mathsf{R}} \setminus \{x'\}$, there is a unique i such that $e_i = d$ and $r_i = r$; conversely, if $e_i \neq \varepsilon$ for some i, then $e_i \in F'$ and $r_i \neq x'$;
- if $e \neq \varepsilon$, then $f_i = e$, for all i;
- for all $A \sqsubseteq \exists r^-.B \in \mathcal{T}$ and $d \in A^{\mathcal{I}'} \cap F'$ such that $d \notin (\exists r^-.B)^{\mathcal{I}'}$, we have $(\exists r^-.B)(d) \in \mathcal{B}_i$ for some i;
- for all $A \sqsubseteq (\leqslant n\ r\ B) \in \mathcal{T}$, $r \in \Sigma^{nt}_{\mathsf{R}}$, and $d \in A^{\mathcal{I}'} \cap F'$, there is a partition $n = n_0 + \ldots + n_m$, such that $d \in (\leqslant n_0\ r\ B)^{\mathcal{I}'}$, and $(\leqslant n_i\ r\ B)(d) \in \mathcal{C}_i$, for all i;
- for all $A \sqsubseteq (\geqslant n\ s\ B) \in \mathcal{T}$ and $d \in A^{\mathcal{I}'} \cap F'$, there is a partition $n = n_0 + \ldots + n_m$, such that $d \in (\geqslant n_0\ s\ B)^{\mathcal{I}'}$ and $(\geqslant n_i\ s\ B)(d) \in \mathcal{C}_i$, for all i with $n_i > 0$.

The transitions for q_0 are similar, but they additionally nondeterministically initialize $\mathcal{M}$ and check the non-transitive part of the ABox in the root, see the appendix. Correctness of the automaton is essentially a consequence of Points *(i)*–*(v)* mentioned above. It is routine to verify that $\mathfrak{A}_{\mathcal{K}}$ is of the required size and can be constructed in the required time. □

Recall that emptiness of NTAs can be checked in polynomial time. Thus, Lemma 2 together with the bounds on N from Theorem 1, yields a 2EXPTIME upper bound for PRPQ entailment in $\mathcal{SIQ}^-$ and $\mathcal{SOQ}$. A matching lower bound is inherited from positive existential query answering in $\mathcal{ALC}$ [Calvanese *et al.*, 2014].

Theorem 2. *PRPQ entailment over $\mathcal{SIQ}^-$ and $\mathcal{SOQ}$ knowledge bases is* 2EXPTIME*-complete.*

5 Finite PEQ Entailment for $\mathcal{SOQ}$

The goal of this section is to establish the following result.

Theorem 3. *Finite PEQ entailment over $\mathcal{SOQ}$ knowledge bases is* 2EXPTIME*-complete.*

The lower bound follows directly from the result on unrestricted query entailment for $\mathcal{ALCO}$ [Ngo *et al.*, 2016], as the latter logic enjoys finite controllability. For the upper bound, we carefully adapt an approach previously used for $\mathcal{SOF}$ [Gogacz *et al.*, 2018], which relies on the following additional condition imposed on tree-like counter-models.

Definition 2. *A canonical tree decomposition is* safe*, if it contains no infinite downward path such that for each node w in this path, $\mathfrak{r}(w)$ is the same transitive role name.*

In what follows, by a *counter-witness* we understand a model of the ABox and the TBox whose transitive closure is a counter-model. The approach requires two ingredients: (1) equivalence of the existence of a finite counter-model and the existence of a counter-witness that admits a safe canonical tree decomposition, and (2) effective regularity of the set of safe canonical tree decompositions (of given width and outdegree) of counter-witnesses. For (2), observe that safety can be easily checked by an automaton with Büchi acceptance condition [Grädel *et al.*, 2002] and the number of states quadratic in the number of transitive role names in $\mathcal{K}$: on each path the automaton remembers the role names associated with two most recently visited nodes; the state is accepting unless they are the same transitive role name. The product of this automaton and the one constructed in the previous section recognizes the desired language. Assuming (1) is also available, the upper bound follows like for the unrestricted case: the algorithm builds the automaton and tests its emptiness.

The reminder of this section provides (1). One implication is obtained via the following observation.

Lemma 3. *If $\mathcal{I}$ is a finite interpretation of a $\mathcal{SOQ}$ KB, then the unravelling procedure from the proof of Theorem 1 yields a safe tree decomposition.*

To prove the converse implication we begin from a carefully chosen counter-witness with a safe canonical tree decomposition. It is well known that each regular set of trees contains a *regular tree*, i.e., a tree with finitely many non-isomorphic subtrees. Hence, if there is a counter-witness with a safe canonical tree decomposition, there is also one with a regular safe canonical tree decomposition. Let $\mathfrak{T} = (T, \mathfrak{I})$ be such a tree decomposition of some counter-witness $\mathcal{I}$, and let $\mathcal{M}$ be the interpretation guaranteed by Definition 1.

Let us restructure $\mathfrak{T}$ by iteratively merging neighboring nodes associated to the same transitive role name: pick a node v with a child w such that $\mathfrak{r}(v) = \mathfrak{r}(w) \in \mathsf{N}^{\mathsf{t}}_{\mathsf{R}}$, redefine $\mathfrak{I}(v)$ as $\mathfrak{I}(v) \cup \mathfrak{I}(w)$, remove w from $\mathfrak{T}$, and promote all children of w to children of v. As a result we obtain a canonical tree decomposition $\mathfrak{S} = (S, \mathfrak{I})$ of $\mathcal{I}$. By construction, $\mathfrak{S}$ is *strongly canonical*: no neighboring nodes in $\mathfrak{S}$ are associated with the same transitive role name. Hence, for each node w with parent $v \neq \varepsilon$, $\Delta_v \cap \Delta_w \setminus \Delta^{\mathcal{M}} = \{d_w\}$ for some $d_w \in F(v)$.

Each regular safe tree decomposition has bounded length of downward paths of nodes associated with the same transitive role name. Consequently, the restructuring above keeps the outdegree and the width bounded.

Lemma 4. *$\mathfrak{S}$ has bounded degree and width.*

We can now easily turn $\mathcal{I}^*$ into a finite model of $\mathcal{K}$. Suppose that on each path of $\mathfrak{S}$, we fix a node v and its ancestor u (neither ε nor a child of ε) such that $\mathfrak{r}(v) = \mathfrak{r}(u)$, $\mathfrak{I}(v) \simeq \mathfrak{I}(u)$, and the witnessing isomorphism h maps d_v to d_u and is identity over $\Delta^{\mathcal{M}}$. Suppose also that for each element in $\Delta^{\mathcal{M}}$, all witnesses required by at-least restrictions can be found among elements of $\Delta^{\mathcal{M}}$ and elements that do not occur in the subtrees of $\mathfrak{S}$ rooted at the chosen nodes v. Note that for each path we can find such a pair of nodes, because the sizes of the bags are bounded in $\mathfrak{S}$. We shall modify $\mathcal{I}$ by removing parts of it and redirecting edges previously leading to the removed parts. Pick any path such that the corresponding d_v has not been processed yet and has not been removed. Remove from $\mathcal{I}$ the union of Δ_w with w ranging over descendents of v (including v), keeping only $\Delta^{\mathcal{M}}$ and d_v. Replace each $\mathfrak{r}(v)$-edge leading from d_v to a removed element $e \in \Delta_v$, with an $\mathfrak{r}(v)$-edge leading from d_v to $h(e)$. Repeat until no such path exists. The resulting interpretation $\mathcal{J}$ is obviously finite. Checking correctness is routine.

Lemma 5. $\mathcal{J}^* \models \mathcal{K}$.

To ensure that $\mathcal{J}^* \not\models \varphi$ we need to choose the nodes v and u more carefully. Relying on φ being a PEQ, not an arbitrary PRPQ, we apply the colored blocking principle [Gogacz *et al.*, 2018]: to keep u and v sufficiently similar and sufficiently far apart, we look at their neighborhoods of sufficiently large radius and use additional coloring to distinguish elements within each neighborhood (see Appendix for details).

6 IQ Entailment

We also get the following results on IQ entailment.

Theorem 4. *Finite and unrestricted IQ entailment is* coNExpTime*-complete over $\mathcal{SOQ}$ KBs. Unrestricted IQ entailment over $\mathcal{SIQ}^-$ KBs and finite IQ entailment over $\mathcal{SIQ}^-$ KBs restricted to a single transitive role and no non-transitive roles is* 2ExpTime*-complete.*

Proof. As is well-known, (finite) IQ entailment reduces to the complement of (finite) KB satisfiability; we focus on the latter. For $\mathcal{SOQ}$, we simply use the facts that (finite) KB satisfiability for $\mathcal{SOQ}$ is NExpTime-complete [Gutiérrez-Basulto *et al.*, 2017] and that the lower bound holds in the finite [Kazakov and Pratt-Hartmann, 2009].

The 2ExpTime upper bound for unrestricted KB satisfiability follows from Theorem 2, and for finite satisfiability in the above fragment of $\mathcal{SIQ}^-$ follows from Theorem 2 and a recent result by Bednarczyk *et al.* [2019], implying that satisfiability and finite satisfiability coincide for this fragment of $\mathcal{SIQ}^-$. This approach cannot be generalized to full $\mathcal{SIQ}^-$ since $\mathcal{SIQ}^-$ lacks the finite model property.

We next show that these upper bounds are tight by reducing the word problem for 2^n-space bounded *alternating Turing machines (ATMs)*, which is known to be 2ExpTime-hard [Chandra *et al.*, 1981]. An ATM $M = (Q, \Theta, \Gamma, q_0, \Delta)$ consists of the set Q of states partitioned into *existential states* $Q_\exists$ and *universal states* $Q_\forall$, the input alphabet Θ, the tape alphabet Γ, the *starting state* $q_0 \in Q_\exists$, and the *transition relation* Δ. Without loss of generality, we assume that each of M's configurations has exactly two successor configurations, universal and existential states alternate, and M accepts a word iff there is an infinite (alternating) run.

Given M, w, we construct in polynomial time a knowledge base $\mathcal{K} = (\mathcal{T}, \{I(a_0)\})$ using a single transitive role name r such that M accepts w iff $\mathcal{K}$ is satisfiable. We represent configurations of size 2^n in the leaves of binary trees of depth n. To this end, we use concept names $X_0, \dots, X_{n-1}$ (representing bits of an exponential counter) and $L_0, \dots, L_n$ (for the levels of the tree), and include the following CIs for all i, j with $0 \leq i < n$ and $i < j \leq n$:

$$L_i \sqsubseteq \exists r.(X_i \sqcap L_{i+1}) \sqcap \exists r.(\neg X_i \sqcap L_{i+1})$$
$$L_{i+1} \sqcap X_i \sqsubseteq \forall r.(L_j \to X_i)$$
$$L_{i+1} \sqcap \neg X_i \sqsubseteq \forall r.(L_j \to \neg X_i)$$

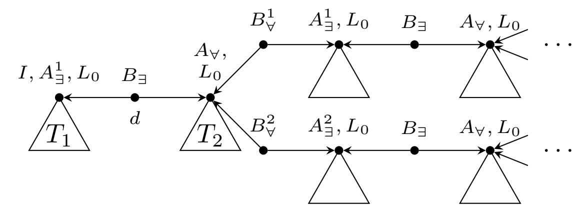
It should be clear that every model of L_0 and these CIs contains a full binary tree of depth n such that the *leaves*, that is, the elements satisfying L_n, correspond to the numbers $0, \dots, 2^n - 1$ in the natural way, via concept names X_i. Let $\Sigma = \Gamma \cup (Q \times \Gamma)$ be the set of possible labels of a cell in M's computation, and introduce concept names C^x_σ for every $\sigma \in \Sigma$, $x \in \{l, h, r\}$. Every leaf with number i is labeled with three concepts $C^l_{\sigma_1}, C^h_{\sigma_2}, C^r_{\sigma_3}$ representing the cells $i-1, i, i+1$ of a configuration using the CI:

$$L_n \sqsubseteq \textstyle\bigsqcap_{x\in\{l,h,r\}} \bigsqcup_{\sigma\in\Sigma} \left(C^x_\sigma \sqcap \bigsqcap_{\sigma'\neq\sigma} \neg C^x_{\sigma'}\right)$$

We use these trees as follows. The concept name I enforces a skeleton structure modeling an alternating computation using the following CIs, for $i \in \{1, 2\}$:

$$I \sqsubseteq A^1_\exists \qquad A^i_\exists \sqsubseteq L_0 \sqcap \exists r^-.(B_\exists \sqcap \exists r.A_\forall)$$
$$A_\forall \sqsubseteq L_0 \sqcap \exists r^-.(B^i_\forall \sqcap \exists r.A^i_\exists)$$

Thus, every model of I contains the following structure, where every triangle represents one of the described trees, and $A_\forall$ ($A^i_\exists$) marks universal (existential) configurations:



It remains to ensure that *(i)* the leaf labeling in every tree is actually a configuration, *(ii)* neighboring trees describe successor configurations, and *(iii)* the first tree is labeled with the initial configuration. We concentrate on *(ii)*, as *(i)* is similar and *(iii)* is straightforward. We illustrate the idea on T_1 and T_2 in the figure. In T_1, we enforce in every leaf an r-successor satisfying the label of that cell in the successor configuration (computable from the C^x_σ). In T_2, we enforce in every leaf an r-successor with the current label C^h_σ. Both in T_1 and T_2, these additional elements satisfy a fresh concept name S and have the same counter value as in the leaves. Observe that, by transitivity, all $2\cdot 2^n$ created nodes are 'visible' from d satisfying $B_\exists$ in the figure. By including the CI $B_\exists \sqsubseteq (\leqslant 2^n\ r\ S)$, S-elements with the same counter value from T_1 and T_2 are forced to identify, thus achieving the desired synchronization. Having *(i)*–*(iii)* in place, it is routine to show that $\mathcal{K}$ is satisfiable iff M accepts w. The lower bound applies to finite satisfiability, since $\mathcal{K}$ is satisfiable iff it is finitely satisfiable. ☐

7 Outlook

This paper makes a step towards a complete picture of query entailment in DLs with number restrictions on transitive roles. There are several natural next steps involving finite entailment. The first is to cover full $\mathcal{SIQ}^-$. A more challenging goal is to go beyond instance queries: an immediate obstacle is that the natural safety condition for $\mathcal{SIQ}^-$ does not guarantee strongly canonical decompositions. Covering full PRPQs even just for $\mathcal{SQ}$ seems to require generalizing the colored blocking principle, or finding an entirely different tool.

Acknowledgments

Gogacz and Murlak were funded by Poland's National Science Centre grant 2018/30/E/ST6/00042, Ibáñez-García by ERC Starting grant 637277 FLEXILOG, and Jung by ERC Consolidator grant 647289 CODA. This work was also supported by the OeAD WTZ project PL 15/2017.

References

[Amarilli and Benedikt, 2015] Antoine Amarilli and Michael Benedikt. Finite open-world query answering with number restrictions. In *LICS-15*, pages 305–316, 2015.

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

[Bednarczyk *et al.*, 2019] Bartosz Bednarczyk, Emanuel Kieronski, and Piotr Witkowski. On the complexity of graded modal logics with converse. In *JELIA*, volume 11468 of *LNCS*, pages 642–658, 2019.

[Bienvenu and Ortiz, 2015] Meghyn Bienvenu and Magdalena Ortiz. Ontology-mediated query answering with data-tractable description logics. In *Proc. of RW-15*, pages 218–307, 2015.

[Calvanese *et al.*, 2014] Diego Calvanese, Thomas Eiter, and Magdalena Ortiz. Answering regular path queries in expressive description logics via alternating tree-automata. *Inf. Comput.*, 237:12–55, 2014.

[Chandra *et al.*, 1981] Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. Alternation. *Journal of the ACM*, 28:114–133, 1981.

[Danielski and Kieronski, 2018] Daniel Danielski and Emanuel Kieronski. Finite satisfiability of unary negation fragment with transitivity. *arXiv:1809.03245*, 2018.

[Glimm *et al.*, 2008a] Birte Glimm, Ian Horrocks, and Ulrike Sattler. Unions of conjunctive queries in SHOQ. In *Proc. of KR-08*, pages 252–262, 2008.

[Glimm *et al.*, 2008b] Birte Glimm, Carsten Lutz, Ian Horrocks, and Ulrike Sattler. Conjunctive query answering for the description logic SHIQ. *J. Artif. Intell. Res. (JAIR)*, 31:157–204, 2008.

[Gogacz *et al.*, 2018] Tomasz Gogacz, Yazmin Ibáñez-García, and Filip Murlak. Finite query answering in expressive description logics with transitive roles. In *Proc. of KR-18*, 2018.

[Grädel and Walukiewicz, 1999] Erich Grädel and Igor Walukiewicz. Guarded fixed point logic. In *Proc. of LICS-99*, pages 45–54, 1999.

[Grädel *et al.*, 2002] Erich Grädel, Wolfgang Thomas, and Thomas Wilke, editors. *Automata, Logics, and Infinite Games: A Guide to Current Research*, volume 2500 of *LNCS*, 2002.

[Gutiérrez-Basulto *et al.*, 2017] Víctor Gutiérrez-Basulto, Yazmín Ibáñez-García, and Jean Christoph Jung. Number restrictions on transitive roles in description logics with nominals. In *Proc. of AAAI-17*, 2017.

[Gutiérrez-Basulto *et al.*, 2018] Víctor Gutiérrez-Basulto, Yazmín Ibáñez-García, and Jean Christoph Jung. Answering regular path queries over sq ontologies. In *Proc. of AAAI-18*. AAAI Press, 2018.

[Ibáñez-García *et al.*, 2014] Yazmin Ibáñez-García, Carsten Lutz, and Thomas Schneider. Finite model reasoning in horn description logics. In *Proc. of KR-14*, 2014.

[Kaminski and Smolka, 2010] Mark Kaminski and Gert Smolka. Terminating tableaux for $\mathcal{SOQ}$ with number restrictions on transitive roles. In *Proc. of the 6th IFIP TC*, pages 213–228, 2010.

[Kazakov and Pratt-Hartmann, 2009] Yevgeny Kazakov and Ian Pratt-Hartmann. A note on the complexity of the satisfiability problem for graded modal logics. In *Proc. of LICS-09*, pages 407–416, 2009.

[Kazakov *et al.*, 2007] Yevgeny Kazakov, Ulrike Sattler, and Evgeny Zolin. How many legs do I have? Non-simple roles in number restrictions revisited. In *Proc. of LPAR-07*, pages 303–317, 2007.

[Kontchakov and Zakharyaschev, 2014] Roman Kontchakov and Michael Zakharyaschev. An introduction to description logics and query rewriting. In *Proc. of RW-14*, pages 195–244, 2014.

[Lutz, 2008] Carsten Lutz. The complexity of conjunctive query answering in expressive description logics. In *Proceedings of IJCAR 2008*, pages 179–193, 2008.

[Ngo *et al.*, 2016] Nhung Ngo, Magdalena Ortiz, and Mantas Simkus. Closed predicates in description logics: Results on combined complexity. In *Proc. of KR-16*, pages 237–246, 2016.

[Pratt-Hartmann, 2009] Ian Pratt-Hartmann. Data-complexity of the two-variable fragment with counting quantifiers. *Inf. Comput.*, 207(8):867–888, 2009.

[Rosati, 2008] Riccardo Rosati. Finite model reasoning in dl-lite. In *Proc. of ESWC-19*, pages 215–229, 2008.

[Rudolph, 2016] Sebastian Rudolph. Undecidability results for database-inspired reasoning problems in very expressive description logics. In *Proc. of KR-16*, pages 247–257, 2016.