On Finite Entailment of Non-Local Queries in Description Logics

Tomasz Gogacz[1] , **Víctor Gutiérrez-Basulto**[2] , **Albert Gutowski**[1] ,
Yazmín Ibáñez-García[2] , **Filip Murlak**[1]

[1]University of Warsaw, Poland
[2]Cardiff University, UK

{t.gogacz, a.gutowski, f.murlak}@mimuw.edu.pl, {gutierrezbasultov, ibanezgarciay}@cardiff.ac.uk

Abstract

We study the problem of *finite* entailment of ontology-mediated queries. Going beyond local queries, we allow transitive closure over roles. We focus on ontologies formulated in the description logics $\mathcal{ALCOI}$ and $\mathcal{ALCOQ}$, extended with transitive closure. For both logics, we show 2ExpTime upper bounds for finite entailment of unions of conjunctive queries with transitive closure. We also provide a matching lower bound by showing that finite entailment of conjunctive queries with transitive closure in $\mathcal{ALC}$ is 2ExpTime-hard.

1 Introduction

The use of ontologies to provide background knowledge for enriching answers to queries posed to databases is a major research topic in the fields of knowledge representation and reasoning. In this data-centric setting, various options for the formalisms used to express ontologies and queries exist, but popular choices are description logics (DLs) and either *local* queries—typically unions of conjunctive queries (UCQs)—or *navigational* queries, like regular path queries (RPQs) and their extensions. The main reasoning problem in this scenario is query entailment, which has been extensively investigated for different combinations of DLs and query languages. An important assumption in data-centric applications is that both database instances and the models they represent are finite. The study of *finite* query entailment, where one is interested in reasoning over finite models only, is thus paramount. Even so, the finite model semantics has received far less attention than the unrestricted one.

Prior work on finite entailment in DLs has focused on local queries, UCQs in particular (Rosati 2008; Ibáñez-García, Lutz, and Schneider 2014; Amarilli and Benedikt 2015; Gogacz, Ibáñez-García, and Murlak 2018; Danielski and Kieronski 2019; Gogacz et al. 2019). The single work studying extensions of UCQs (Rudolph 2016), brings only undecidability results for navigational queries and various expressive DLs. In contrast, for unrestricted entailment a large body of literature is available, providing multiple positive results for navigational queries and both expressive and lightweight DLs (Calvanese, Eiter, and Ortiz 2014; Stefanoni et al. 2014; Bienvenu, Ortiz, and Simkus 2015; Jung et al. 2017; Gutiérrez-Basulto, Ibáñez-García, and Jung 2018; Gogacz et al. 2019; Bednarczyk and Rudolph 2019). The lack of research on finite entailment of non-local queries

(e.g. variants of RPQs) comes as a surprise since, as noted in the above works, they are necessary in data-centric applications dealing with graph databases.

In this paper, we close the distance to the undecidability frontier for finite entailment, delineated by Rudolph (2016), by identifying some decidable classes of non-local ontology-mediated queries. We focus on UCQs with transitive closure over roles; from the viewpoint of DL applications, transitive closure is arguably one of the most useful features of regular expressions over roles. As the ontology component, we consider extensions of the DL $\mathcal{ALC}$, allowing for transitive closure of roles. The study of finite entailment is relevant for this combination because, unlike for plain CQs, query entailment of CQs with transitive closure is not finitely controllable even for $\mathcal{ALC}$, and thus finite and unrestricted entailment do not coincide. As a consequence, dedicated algorithmic methods and lower bounds need to be developed.

Our main finding is that finite entailment of UCQs with transitive closure in $\mathcal{ALCOI}^+$ and $\mathcal{ALCOQ}^+$ is 2ExpTime-complete. As unrestricted entailment of positive RPQs is 2ExpTime-complete for both logics (Calvanese, Eiter, and Ortiz 2014; Bednarczyk and Rudolph 2019), and the hardness already holds for CQs and both $\mathcal{ALCI}$ and $\mathcal{ALCO}$ (Lutz 2008; Ngo, Ortiz, and Simkus 2016), we conclude that the complexity of our setting is the same over arbitrary and finite models. As finite entailment of two-way RPQs in $\mathcal{ALCIOF}$ is undecidable (Rudolph 2016), our results are quite close to the undecidability frontier.

While there are positive results on finite entailment in expressive DLs related to $\mathcal{ALCOI}^+$ or $\mathcal{ALCOQ}^+$, they all concern local queries. For example, the 2ExpTime algorithms for finite entailment of UCQs in $\mathcal{SOI}$, $\mathcal{SIF}$, and $\mathcal{SOQ}$ (Gogacz, Ibáñez-García, and Murlak 2018; Gogacz et al. 2019) are considerably easier to obtain owing to the relatively simple structure of transitive roles. For the more expressive $\mathcal{SHOIF}$, the problem is undecidable (Rudolph 2016). Indeed, to the best of our knowledge, this paper offers the first positive results on finite entailment of non-local queries in description logics.

In order to show our main result, we provide a series of intermediate reductions allowing us to work with simpler variants of the input logics and query formalisms. The reductions apply to both logics, requiring sometimes conceptually different proofs for each of them. Our proofs encom-

pass different techniques. We use unravelling operations to establish a tree-like model property; i.e., to show that if a query is not entailed by a knowledge base, then there is a tree-like counter-model. This in turn serves as the basis for automata-based approaches to finite entailment. We also use the coloured blocking principle to construct appropriate finite counter-models out of infinite tree-like counter-models.

A technical report with missing proofs can be found at https://arxiv.org/abs/2006.16869.

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* $\mathsf{N_I}$. A *role* is a role name or an *inverse role* r^-. The *(transitive-reflexive) closure* of a role r is r^*. $(\mathcal{ALCOIQ}^-)^+$-*concepts* C, D are defined by the grammar

$$C, D ::= A \mid \neg C \mid C \sqcap D \mid \{a\} \mid \exists s.C \mid (\leqslant n \ s' \ C)$$

where $A \in \mathsf{N_C}$, s is a role or the closure of a role, $n \geq 0$ is a natural number given in binary, and s' is a role name or the closure of a role name. We will use $(\geqslant n \ s' \ C)$ as abbreviation for $\neg(\leqslant n{-}1 \ s' \ C)$, together with standard abbreviations $\bot, \top, C \sqcup D, \forall s.C$. Concepts of the form $(\leqslant n \ s' \ C)$, $(\geqslant n \ s' \ C)$, and $\{a\}$ are called *at-most restrictions*, *at-least restrictions*, and *nominals*, respectively. Note that in $(\mathcal{ALCOIQ}^-)^+$-concepts, *inverse roles* are not allowed in at-most and at-least restrictions.

An $(\mathcal{ALCOIQ}^-)^+$-*TBox* $\mathcal{T}$ is a finite set of *concept inclusions (CIs)* $C \sqsubseteq D$, where C, D are $(\mathcal{ALCOIQ}^-)^+$-concepts. An *ABox* $\mathcal{A}$ is a finite non-empty set of *concept* and *role assertions* of the form $A(a), r(a, b), r^*(a, b)$ where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$ and $\{a, b\} \subseteq \mathsf{N_I}$. A *knowledge base (KB)* is a pair $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. We write $\mathsf{CN}(\mathcal{K})$, $\mathsf{rol}(\mathcal{K})$, $\mathsf{nom}(\mathcal{K})$, and $\mathsf{ind}(\mathcal{K})$ for, respectively, the set of *all concept names, role names, nominals, and individuals occurring in* $\mathcal{K}$. The *counting threshold* of $\mathcal{K}$ is one plus the greatest number used in $\mathcal{K}$. We let $\|\mathcal{K}\|$ be the total size of the representation of $\mathcal{K}$.

Sublogics of $(\mathcal{ALCOIQ}^-)^+$ are defined by forbidding any subset of the following features: nominals ($\mathcal{O}$), inverse roles ($\mathcal{I}$), counting ($\mathcal{Q}$), and closure of roles ($^+$); this is indicated by dropping corresponding letter or decoration from the name of the logic. The symbol $^-$ indicates restricted interaction between inverse roles and counting restrictions, so if either is forbidden, $^-$ is also dropped. The focus of this paper is on the logics $\mathcal{ALCOI}^+$ and $\mathcal{ALCOQ}^+$, obtained by forbidding counting and inverse roles, respectively. Other sublogics of $(\mathcal{ALCOIQ}^-)^+$, in particular $(\mathcal{ALCIQ}^-)^+$, are used to uniformize and share fragments of arguments.

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 $a \in A^\mathcal{I}$ for all $A(a) \in \mathcal{A}$, $(a, b) \in r^\mathcal{I}$ for all $r(a, b) \in \mathcal{A}$, and $(a, b) \in (r^\mathcal{I})^*$ for all $r^*(a, b) \in \mathcal{A}$, where $(r^\mathcal{I})^*$ is the usual transitive-reflexive closure of the binary relation $r^\mathcal{I}$. Finally, $\mathcal{I}$ is a *model of a KB* $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, written $\mathcal{I} \models \mathcal{K}$, if $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \models \mathcal{A}$.

An interpretation $\mathcal{I}$ is *finite* if $\Delta^\mathcal{I}$ is finite. 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, written $\mathcal{I} \upharpoonright \Delta$, 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 $r \in \mathsf{N_R}$.

A *homomorphism* from interpretation $\mathcal{I}$ to interpretation $\mathcal{J}$, written as $h : \mathcal{I} \to \mathcal{J}$ is a function $h : \Delta^\mathcal{I} \to \Delta^\mathcal{J}$ that preserves roles, concepts, and individual names: that is, for all $r \in \mathsf{N_R}$, $(h(d), h(e)) \in r^\mathcal{J}$ whenever $(d, e) \in r^\mathcal{I}$, for all $A \in \mathsf{N_C}$, $h(d) \in A^\mathcal{J}$ whenever $d \in A^\mathcal{I}$, and $h(a) = a$ for all $a \in \mathsf{ind}(\mathcal{K})$.

Queries and Finite Entailment

Let $\mathsf{N_V}$ be a countably infinite set of *variables*. An *atom* is an expression of the form $A(t)$, $t = t'$, $r(t, t')$, or $r^*(t, t')$ with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $t, t' \in \mathsf{N_V} \cup \mathsf{N_I}$, referred to as *concept*, *equality*, *role*, and *transitive atoms*, respectively. A *conjunctive query (with transitive atoms)* is a set of atoms, understood as the conjunction thereof. We write $\mathcal{CQ}^+$ for conjunctive queries (with transitive atoms), and $\mathcal{CQ}$ for conjunctive queries without transitive atoms. Let $var(q)$ be the set of variables occurring in the atoms of $q \in \mathcal{CQ}^+$. A *match for q in $\mathcal{I}$* is a function $\eta : var(q) \to \Delta^\mathcal{I}$ such that $\mathcal{I}, \eta \models q$ under the standard semantics of first-order logic, assuming that the extension of r^* is the reflexive-transitive closure of the extension of r. An interpretation $\mathcal{I}$ *satisfies* q, written $\mathcal{I} \models q$, if there exists a match for q in $\mathcal{I}$.

Fix $p, q \in \mathcal{CQ}^+$ and a function $\eta : var(p) \to \mathsf{N_V} \cup \mathsf{N_I}$. Let $\eta(p)$ be obtained from p by substituting each $x \in var(p)$ with $\eta(x)$. We call η a *homomorphism* if $\eta(p) \subseteq q^*$, where q^* is obtained by saturating q as follows for all r and all $s, s' \in \{r, r^*\}$: if $\{x = y, y = z\} \subseteq q$, add $x = z$, if $\{s(x, y), y = y'\} \subseteq q$, add $s(x, y')$, if $\{s(x, y), x = x'\} \subseteq q$, add $s(x', y)$, if $\{s(x, y), s'(y, z)\} \subseteq q$, add $r^*(x, z)$. Let $\mathcal{I} \models q$. If there is a homomorphism from p to q, then $\mathcal{I} \models p$. If there is a homomorphism from $\mathcal{I}$ to $\mathcal{J}$, then $\mathcal{J} \models q$.

A union of conjunctive queries is a finite set of $\mathcal{CQ}^+$s. We write $\mathcal{UCQ}^+$ for unions of conjunctive queries, and $\mathcal{UCQ}$ for unions of conjunctive queries without transitive atoms. An interpretation $\mathcal{I}$ satisfies $Q \in \mathcal{UCQ}^+$, written as $\mathcal{I} \models Q$, if $\mathcal{I} \models q$ for some $q \in Q$. A *fragment* of Q is a $\mathcal{CQ}^+$ formed by a connected subset of atoms of some $q \in Q$. We let $\|Q\|$ denote the number of all fragments of Q; note that it is exponential in $\max_{q \in Q} |q|$.

We say that $\mathcal{K}$ *finitely entails* $\mathcal{Q}$, written $\mathcal{K} \models_{\mathsf{fin}} \mathcal{Q}$, if each finite model of $\mathcal{K}$ satisfies $\mathcal{Q}$. A model of $\mathcal{K}$ that does not

satisfy Q is a *counter-model*. The *finite entailment problem* asks if a given KB $\mathcal{K}$ finitely entails a given query Q.

We also consider *finite entailment modulo types*, which allows more precise complexity bounds. A *unary $\mathcal{K}$-type* is a subset of $\mathsf{CN}(\mathcal{K})$ including either A or $\bar{A}$ for each $A \in \mathsf{CN}(\mathcal{K})$. Let $\mathsf{Tp}(\mathcal{K})$ be the set of all unary $\mathcal{K}$-types. For an interpretation $\mathcal{I}$ and an element $d \in \Delta^{\mathcal{I}}$, the *unary $\mathcal{K}$-type of d in $\mathcal{I}$* is $\mathsf{tp}^{\mathcal{I}}(d) = \left\{ A \in \mathsf{CN}(\mathcal{K}) \mid d \in A^{\mathcal{I}} \right\}$. We say that $\mathcal{I}$ *realizes* a unary $\mathcal{K}$-type τ if $\tau = \mathsf{tp}^{\mathcal{I}}(d)$ for some $d \in \Delta^{\mathcal{I}}$. For a KB $\mathcal{K}$, a query Q, and a set of unary types $\Theta \subseteq \mathsf{Tp}(\mathcal{K})$ we write $\mathcal{K} \models_{\mathsf{fin}}^{\Theta} Q$ if for each finite interpretation $\mathcal{I}$ that only realizes types from Θ, if $\mathcal{I} \models \mathcal{K}$ then $\mathcal{I} \models Q$. In this context, a counter-model is a model of $\mathcal{K}$ that only realizes types from Θ and does not satisfy Q.

Normal Form and Additional Assumptions

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

$$\bigsqcap_i A_i \sqsubseteq \bigsqcup_j B_j, \quad A \sqsubseteq (\leqslant n \, r \, B), \quad A \sqsubseteq (\geqslant n \, r \, B),$$

$$A \sqsubseteq (\leqslant n \, r^* \, B), \quad A \sqsubseteq (\geqslant n \, r^* \, B),$$

$$A \sqsubseteq \forall r^-.B, \quad A \sqsubseteq \exists r^-.B, \quad A \sqsubseteq \exists (r^-)^*.B,$$

where A, A_i, B, B_j are concept names or nominals, $r \in \mathsf{N_R}$, and empty disjunction and conjunction are equivalent to $\bot$ and $\top$, respectively. In logics without counting, the number n in at-most restrictions must be 0, and in at-least restrictions it must be 1. We also assume that for each concept name A used in $\mathcal{K}$ there is a *complementary* concept name $\bar{A}$ axiomatized with CIs $\top \sqsubseteq A \sqcup \bar{A}$ and $A \sqcap \bar{A} \sqsubseteq \bot$.

A concept name $B \in \mathsf{CN}(\mathcal{K})$ is *relevant* if $\mathcal{K}$ contains a CI of the form $A \sqsubseteq (\leqslant n \, r^* \, B)$ with $n > 0$. We let $\mathsf{RCN}(\mathcal{K})$ denote the set of relevant concept names in $\mathcal{K}$. A concept name B is *relevant in $\mathcal{I}$ for $d \in \Delta^{\mathcal{I}}$ with respect to $r \in \mathsf{N_R}$* if $d \in A^{\mathcal{I}}$ for some CI $A \sqsubseteq (\leqslant n \, r^* \, B)$ in $\mathcal{K}$ with $n > 0$. We call $\mathcal{K}$ *sticky* if for each model $\mathcal{I}$ of $\mathcal{K}$, each $r \in \mathsf{N_R}$, and each $d \in \Delta^{\mathcal{I}}$, all concept names relevant for d with respect to r are also relevant (with respect to r) for each r-successor of d. Stickiness of $\mathcal{K}$ can be assumed without loss of generality: for each $B \in \mathsf{RCN}(\mathcal{K})$ and $r \in \mathsf{rol}(\mathcal{K})$ introduce fresh concept names B_r and $\overline{B}_r$ axiomatized with $\top \sqsubseteq B_r \sqcup \overline{B}_r$, $B_r \sqcap \overline{B}_r \sqsubseteq \bot$, $B_r \sqsubseteq (\leqslant 0 \, r \, \overline{B}_r)$, $B_r \sqsubseteq (\leqslant N \, r^* \, B)$, where N is the counting threshold in $\mathcal{K}$, and add $A \sqsubseteq B_r$ for each CI $A \sqsubseteq (\leqslant n \, r^* \, B)$ in $\mathcal{K}$.

A variable y is *linking* in $q \in \mathcal{CQ}^+$ if the only atoms in q using y are $r^*(x, y), r^*(y, z)$ for some $x, z \in var(q)$. In a match η for q, $\eta(y)$ can be any node on a path from $\eta(x)$ to $\eta(z)$. We call q *normalized* if in every atom over r^* at least one variable is linking, and for every two atoms $r^*(x, y), r^*(y, z)$ with y linking, exactly one of the variables x and z is linking too. Each query can be normalized by first eliminating all linking variables, and then subdividing each r^* atom into three r^* atoms using two fresh linking variables. Without loss of generality we can assume that the input $\mathcal{UCQ}^+$s consist of normalized connected $\mathcal{CQ}^+$s.

3 Plan of Attack

Our main technical contribution are the following results.

Theorem 1. *Finite entailment of $\mathcal{UCQ}^+$s over $\mathcal{ALCOI}^+$ or $\mathcal{ALCOQ}^+$ knowledge bases is* 2EXPTIME-*complete.*

To prove upper bounds, we will show a series of reductions that will allow us to deal at the end with a base case where the query is a plain UCQ and the knowledge base is 'single-role ABox-trivial' either in $\mathcal{ALCI}^+$ or in $\mathcal{ALCQ}^+$ without at-most restrictions over closures of roles. These reductions can be seen as different phases in our decision procedure, each of them depending on the previous one.

I. We start by showing in Section 4 that one can eliminate nominals from the input knowledge bases, i.e. roughly, that one can reduce finite entailment of $\mathcal{UCQ}^+$s over $\mathcal{ALCOI}^+$ or $\mathcal{ALCOQ}^+$ KBs to $\mathcal{ALCI}^+$ or $\mathcal{ALCQ}^+$ KBs, respectively. We further show that the input ABox can be assumed to be 'trivial' in the sense that it contains no role assertions and that only one individual name occurs in it.

II. In Section 5 we show that the variants obtained in Step I above can be further reduced to the case where the knowledge base contains a single role name. This is in line with seeing $\mathcal{ALCQ}^+$ and $\mathcal{ALCI}^+$ as a fusion of logics (Baader et al. 2002), where the interaction between different roles is limited. In a nutshell, we show that there exists a finite counter-model if and only if there exists a tree-like counter-model accepting a decomposition into components interpreting a single role name.

III. In Section 6, assuming single-role ABox-trivial $\mathcal{ALCI}^+$ or $\mathcal{ALCQ}^+$ KBs, we eliminate transitive closure from queries and from at-most restrictions. This step is the one requiring the most technical effort. We will develop the notion of hybrid decompositions, which are tree decompositions that associate with each node arbitrarily large interpretations, but have a certain more subtle parameter bounded. We will observe special characteristics of these interpretations that allow us to establish structural restrictions between neighbouring nodes in the decomposition. Roughly, the key result will be that finite counter-models can be unravelled into counter-models admitting a hybrid decomposition with the above features. This tree-like model property will then be the basis for automata-based approaches to finite entailment in this setting.

IV. Section 7 provides a decision procedure for the base case described above. We will rely again on tree unravelling and automata-based techniques.

For the matching lower bound, we show the following.

Theorem 2. *Finite entailment of $\mathcal{CQ}^+$s over $\mathcal{ALC}$ knowledge bases is* 2EXPTIME-*hard.*

The proof is by reducing the word problem of exponentially space bounded alternating Turing machines, which is known to be 2EXPTIME-hard (Chandra, Kozen, and Stockmeyer 1981). The encoding is a direct adaptation of the one used to show that (arbitrary) entailment of CQs over $\mathcal{SH}$ knowledge bases is 2EXPTIME-hard (Eiter et al. 2009), but use $\mathcal{CQ}^+$s to compensate for the weaker logic.

4 Eliminating Nominals, Trivializing ABoxes

Finite entailment of a $\mathcal{UCQ}^+$ over an $\mathcal{ALCOI}^+$ or $\mathcal{ALCOQ}^+$ KB can be reduced to multiple instances of finite entailment

of $\mathcal{UCQ}^+$s over $\mathcal{ALCI}^+$ or $\mathcal{ALCQ}^+$ KBs, respectively, with *trivial* ABoxes; the latter means that ABoxes mention only one individual (with a fully specified unary type) and contain no binary assertions. Concluding complexity bounds for $\mathcal{ALCOI}^+$ or $\mathcal{ALCOQ}^+$ from bounds for $\mathcal{ALCI}^+$ or $\mathcal{ALCQ}^+$ requires precise estimations of the number of these instances and their parameters. The key parameters are the size of the KB, the number of $\mathcal{CQ}^+$s in the query, and their size.

Consider an $\mathcal{ALCOI}^+$ or $\mathcal{ALCOQ}^+$ KB $\mathcal{K}$ and $Q \in \mathcal{UCQ}^+$ with $\mathcal{CQ}^+$s of size at most m. Then, the reduction gives an at most doubly exponential number of instances, with KBs of size at most $\text{poly}(\|\mathcal{K}\|, N^{1+|\text{RCN}(\mathcal{K})|})$ and queries with at most $|Q| \cdot (\|\mathcal{K}\| \cdot N^{|\text{RCN}(\mathcal{K})|})^{O(m)}$ $\mathcal{CQ}^+$s of size $O(m)$. Importantly, the reduction preserves the counting threshold N and the set $\text{RCN}(\mathcal{K})$ of relevant concept names.

Let us sketch the argument. Using routine one-step unravelling one can show that for $\mathcal{ALCOI}^+$ it suffices to search for counter-models $\mathcal{I}$ that can be decomposed into domain-disjoint interpretations $\mathcal{I} \restriction \text{ind}(\mathcal{K})$ and $\mathcal{I}_1, \ldots, \mathcal{I}_n$ for some $n \leq |\mathcal{K}| \cdot |\text{ind}(\mathcal{K})|$ with two kinds of additional edges: arbitrary edges connecting elements from $\bigcup_i \Delta^{\mathcal{I}_i}$ with elements from $\text{nom}(\mathcal{K})$, and a single edge connecting a distinguished element $d_i \in \Delta^{\mathcal{I}_i}$ with a corresponding element $e_i \in \text{ind}(\mathcal{K})$ for each $i \leq n$. Iterating over possible $\mathcal{I} \restriction \text{ind}(\mathcal{K})$, we check if there exist $\mathcal{I}_i$ that are compatible with $\mathcal{K}$, provide witnesses for elements from $\text{ind}(\mathcal{K})$, and avoid satisfying the query. The properties required for a single $\mathcal{I}_i$ can be encoded as a finite entailment problem for a modified KB whose ABox describes the unary type of d_i, and a union of selected fragments of Q with some variables substituted by elements of $\{d_i\} \cup \text{nom}(\mathcal{K})$. By representing the existence of an r-edge to $a \in \text{nom}(\mathcal{K})$ with a fresh concept name $A_{\exists r.\{a\}}$, suitably axiomatized, we can eliminate nominals from the KB and from the query.

For $\mathcal{ALCOQ}^+$, unravelling is made harder by at-most restrictions over closures of roles. We shall treat in a special way all successors of elements from $\text{ind}(\mathcal{K})$ that are affected by such at-most restrictions. Let us call an r-successor e of $d \in \Delta^{\mathcal{I}}$ *directly r-relevant* if $e \in A^{\mathcal{I}}$ for some A relevant for d with respect to r, and let $\text{rel}_r^{\mathcal{I}}(d)$ be the set of all r-*relevant successors* of d; that is, the least set containing all directly r-relevant successors of d and closed under directly r-relevant successors. The key insight is that this set has bounded size.

Fact 1 (Gutiérrez-Basulto, Ibáñez-García, and Jung 2018).
If N is the counting threshold in $\mathcal{K}$ and $\mathcal{I} \models \mathcal{K}$, then $|\text{rel}_r^{\mathcal{I}}(d)| \leq N^{|\text{RCN}(\mathcal{K})|}$ for all $d \in \Delta^{\mathcal{I}}$ and $r \in \text{rol}(\mathcal{K})$.

In the argument for $\mathcal{ALCOI}^+$, we replace $\text{ind}(\mathcal{K})$ with

$$\text{ind}(\mathcal{K}) \cup \bigcup \{\text{rel}_r^{\mathcal{I}}(a) \mid r \in \text{rol}(\mathcal{K}), a \in \text{ind}(\mathcal{K})\},$$

and similarly for $\text{nom}(\mathcal{K})$. Appendix A has full details.

In the following sections we shall focus on the decision procedures for $\mathcal{ALCI}^+$ and $\mathcal{ALCQ}^+$, but to ensure that they give 2EXPTIME upper bounds for $\mathcal{ALCOI}^+$ and $\mathcal{ALCOQ}^+$ when combined with the reductions above, we will need more careful complexity analysis.

5 Eliminating Multiple Roles

Moving on with simplifications, we reduce finite entailment of $\mathcal{UCQ}^+$s over $\mathcal{ALCI}^+$ or $\mathcal{ALCQ}^+$ KBs to the single-role case. More precisely, we show how to solve one instance of finite entailment by solving many instances of finite entailment modulo types with single-role KBs.

Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ALCI}^+$ or $\mathcal{ALCQ}^+$ KB, $Q \in \mathcal{UCQ}^+$ and $m = \max_{q \in Q} |q|$. One can decide if $\mathcal{K} \models_{\text{fin}} Q$ in time $O\big(\text{poly}(\|\mathcal{K}\|, 2^{|\text{CN}(\mathcal{K})| \cdot |Q| \cdot 4^m})\big)$ using oracle calls to finite entailment modulo types with single-role KBs of size at most $\text{poly}(\|\mathcal{K}\|, \|Q\|)$, unions of at most $|Q| \cdot 2^{\text{poly}(m)}$ $\mathcal{CQ}^+$s of size $\mathcal{O}(m)$, and type sets of size at most $2^{|\text{CN}(\mathcal{K})| \cdot |Q| \cdot 4^m}$; the KBs inherit their counting threshold and set of relevant concept names from $\mathcal{K}$.

The argument again relies on unravelling, which allows focusing on *tree-like* counter-models; that is, ones that can be decomposed into multiple finite Σ_r-subinterpretations with $\Sigma_r = \{r\} \cup \text{CN}(\mathcal{K})$ and $r \in \text{rol}(\mathcal{K})$, called *bags*, arranged into a (possibly infinite) tree such that: (1) two bags share a single element if they are neighbours, and are disjoint if they are not neighbours; (2) each element d occurs in exactly one Σ_r-bag for each $r \in \text{rol}(\mathcal{K})$ and some bag containing d is the parent of all other bags containing d; (3) the root bag satisfies $\mathcal{A}$. A tree-like interpretation is a model of $\mathcal{K}$ iff each Σ_r-bag is a model of the TBox $\mathcal{T}_r$, collecting all CIs from $\mathcal{T}$ that mention only role name r.

Evaluating Q over a tree-like interpretation $\mathcal{I}$ can be distributed over bags by means of Q-*labellings*, which assign to each element its *(unary) Q-type*, summarizing information relevant for Q as a set of pairs (p, V) where p is a fragment of Q and $\emptyset \neq V \subseteq \text{var}(p)$. We shall think of a Q-labelling as an extension $\mathcal{I}'$ of $\mathcal{I}$ to fresh concept names $A_{p,V}$. We are interested in Q-labellings $\mathcal{I}'$ such that $A_{q,V}^{\mathcal{I}'} = \emptyset$ for all $q \in Q$ and $\emptyset \neq V \subseteq \text{var}(q)$; we call them Q-*refutations*. We call $\mathcal{I}'$ *correct* if $e \in A_{p,V}^{\mathcal{I}'}$ iff $\eta(V) = \{e\}$ for some match η for p in $\mathcal{I}'$. We call $\mathcal{I}'$ *consistent* if each bag satisfies the following: for each partition of a fragment p' into fragments $p, p_1, p_2, \ldots, p_k$ with $\text{var}(p_i) \cap \text{var}(p_j) = \emptyset$ for $i \neq j$, $V_i = \text{var}(p_i) \cap \text{var}(p)$, and $\emptyset \neq V \subseteq \text{var}(p)$, there is no match η for p in the bag such that $\eta(V_i) = \{e_i\} \subseteq A_{p_i,V_i}^{\mathcal{I}'}$ for all i but $\eta(V) = \{e\} \not\subseteq A_{p',V}^{\mathcal{I}'}$.

Lemma 1. *The correct Q-labelling of $\mathcal{I}$ is a Q-refutation iff $\mathcal{I}$ admits a consistent Q-refutation.*

It remains to show how to find consistent Q-refutations $\mathcal{J}$ such that $\mathcal{J} \models \mathcal{K}$ and prove that they can be turned into finite counter-models.

Consistent Q-refutations can be recognized by a tree automaton but there are two obstacles. First, bags are finite, but arbitrarily large, so tree-like interpretations do not naturally encode as finitely-labelled trees. Second, the information that needs to be passed between bags is the unary type of the shared element, and the unary type of the element shared with the parent bag needs to be related with the *set of unary types* of elements shared with child bags. As the number of such types is $2^{\text{CN}(\mathcal{K}) \cdot |Q| \cdot 4^m}$, this suggests a triply exponential construction. Our response is to generalize automata to trees

over infinite alphabets and avoid representing the transition relation explicitly.

Definition 1. *A* (Büchi) automaton $\mathcal{B} = (\Sigma, \Gamma, S, I, F, \delta)$ *consists of a* node alphabet Σ, *an* edge alphabet Γ, *a finite set S of* states, *sets $I, F \subseteq S$ of* initial *and* accepting *states, and a* transition function $\delta : S \times \Sigma \to 2^{(\Gamma \times S)^*}$ *mapping state-letter pairs to sets of words over the alphabet $\Gamma \times S$. A run ρ of $\mathcal{B}$ on a tree T maps nodes of T to S such that $\rho(\varepsilon) \in I$ and for each node v with children $v_1, v_2, \ldots, v_n$,*

$$(\gamma_1, \rho(v_1))(\gamma_2, \rho(v_2)) \ldots (\gamma_n, \rho(v_n)) \in \delta(\rho(v), \sigma),$$

where γ_i is the label of the edge (v, v_i) and σ is the label of the node v. A tree T is accepted *by $\mathcal{B}$ if for some run ρ, each branch of T contains infinitely many nodes from $\rho^{-1}(F)$. All trees accepted by $\mathcal{B}$ form the set* recognized *by $\mathcal{B}$.*

Let us build an automaton $\mathcal{B}_{\mathcal{K},Q}$ recognizing consistent Q-refutations. Because the state space is finite, the automaton cannot compute which elements are shared between bags: it must be given this information. We provide it by marking in each bag the element shared with the parent bag and displaying the shared element on the edge to the parent. With that, the actual elements used in the bags do not matter any more: the element on the edge between bags indicates the element of the parent bag that should be identified with the marked element of the child bag. Let Δ be a countably infinite set. The node alphabet is the set of finite Σ_r-interpretations for $r \in \text{rol}(\mathcal{K})$ with domains contained in Δ, and the edge alphabet is Δ. States consist of $r \in \text{rol}(\mathcal{K})$ and a unary type τ that includes no $A_{q,V}$ with $q \in Q$. A state (r, τ) is initial if $\tau \supseteq \{A \mid A(a) \in \mathcal{A}\}$ where $\{a\} = \text{ind}(\mathcal{K})$. All states are accepting. The transition function ensures that the input tree represents a tree-like interpretation and that each Σ_r-bag is a model of $\mathcal{T}_r$ and satisfies the consistency condition.

Our automata are infinite objects, but in the oracle model their emptiness can be tested in PTIME (see Appendix B).

Fact 2. *There is a polynomial-time algorithm with an oracle that, for each automaton $\mathcal{B} = (\Sigma, \Gamma, S, I, F, \delta)$, accepts input (S, I, F) with oracle $\text{step}_{\mathcal{B}}$ iff $\mathcal{B}$ accepts some tree, where $\text{step}_{\mathcal{B}}$ is the set of pairs $(P, q) \in 2^S \times S$ such that $\delta(q, \sigma) \cap (P \times \Gamma)^* \neq \emptyset$ for some $\sigma \in \Sigma$.*

Deciding if $(P, (r, \tau)) \in \text{step}_{\mathcal{B}_{\mathcal{K},Q}}$ reduces to a single instance of finite entailment modulo types, for $\mathcal{K}' = (\mathcal{T}_r, \mathcal{A}_\tau)$ where $\mathcal{A}_\tau$ encodes τ as a set of assertions on some distinguished individual, Q' obtained, informally speaking, by taking the union of all matches forbidden in the consistency condition, and $\Theta' = \{\tau \mid \forall r \in \text{rol}(\mathcal{K}) \ (r, \tau) \in P \cup (r, \tau)\}$. Hence, applied to $\mathcal{B}_{\mathcal{K},Q}$, Fact 2 yields the desired bounds.

To get finite counter-models we use coloured blocking. For $d \in \Delta^{\mathcal{I}}$, the *$n$-neighbourhood $N_n^{\mathcal{I}}(d)$ of d* is the interpretation obtained by restricting $\mathcal{I}$ to elements $e \in \Delta^{\mathcal{I}}$ within distance n from d in $\mathcal{I}$, enriched with a fresh concept interpreted as $\{d\}$. A *colouring of $\mathcal{I}$ with k colours* is an extension $\mathcal{I}'$ of $\mathcal{I}$ to k fresh concept names $B_1, \ldots, B_k$ such that $B_1^{\mathcal{I}'}, \ldots, B_k^{\mathcal{I}'}$ is a partition of $\Delta^{\mathcal{I}'} = \Delta^{\mathcal{I}}$. We say that $d \in B_i^{\mathcal{I}'}$ has colour B_i. We call $\mathcal{I}'$ *n-proper* if for each $d \in \Delta^{\mathcal{I}'}$ all elements of $N_n^{\mathcal{I}'}(d)$ have different colours.

Fact 3 (Gogacz, Ibáñez-García, and Murlak 2018). *If $\mathcal{I}$ has bounded degree, then for all $n \geq 0$ there exists an n-proper colouring $\mathcal{I}'$ of $\mathcal{I}$ with finitely many colours. Consider interpretation $\mathcal{J}$ obtained from $\mathcal{I}'$ by redirecting some edges such that the old target and the new target have isomorphic n-neighbourhoods in $\mathcal{I}'$. Then, for each $q \in \mathcal{CQ}$ with at most $\sqrt{n}$ binary atoms, if $\mathcal{I} \not\models q$, then $\mathcal{J} \not\models q$.*

From the proof of Fact 2 it also follows that if $\mathcal{B}$ accepts some tree, then it also accepts a *regular* tree; that is, a tree that has only finitely many non-isomorphic subtrees. Each regular tree has bounded branching and uses only finitely many different labels. Consequently, the tree-like counter-model $\mathcal{I}$ corresponding to a regular tree accepted by $\mathcal{B}_{\mathcal{K},Q}$ uses finitely many different bags (up to isomorphism), which means that their size is bounded by some k. Therefore, $\mathcal{I}$ has bounded degree and, because Σ_r-bags are disjoint, the length of simple directed r-paths in $\mathcal{I}$ is also bounded by k.

For $\ell = 2k$, let $Q^{(\ell)}$ be obtained from Q by replacing each transitive atom $r^*(x, y)$ with the disjunction

$$x = y \vee r(x, y) \vee \bigvee_{1 < i \leq \ell} r^i(x, y),$$

where $r^i(x, y)$ expresses the existence of an r-path of length i from x to y as $r(x, z_1) \wedge r(z_1, z_2) \wedge \cdots \wedge r(z_{i-1}, y)$ for fresh variables $z_1, \ldots, z_{i-1}$. Rewrite $Q^{(\ell)}$ as a $\mathcal{UCQ}$ and let t be the maximal number of binary atoms in one $\mathcal{CQ}$ in $Q^{(\ell)}$.

Fact 4. *If each simple directed r-path in an interpretation $\mathcal{J}$ has length at most ℓ, then $\mathcal{J} \models Q$ iff $\mathcal{J} \models Q^{(\ell)}$.*

Fix $n = t^2$ and let $\mathcal{I}'$ be an n-proper colouring of $\mathcal{I}$. On each infinite branch, select the first Σ_r-bag $\mathcal{M}$ such that for some Σ_r-bag $\mathcal{M}'$ higher on this branch, the n-neighbourhood of the element e, shared by $\mathcal{M}$ and its parent, is isomorphic to the n-neighbourhood of the element e', shared by $\mathcal{M}'$ and its parent. The set of selected bags forms a maximal antichain, and by König's Lemma, it is finite. Let $\mathcal{F}$ be the interpretation obtained by taking the union of all strict ancestors of the selected bags, and for each element e shared by a selected Σ_r-bag and its parent, redirect each r-edge incident with e: instead of pointing at an r-neighbour of e, it should point at the corresponding r-neighbour of e'. If $\mathcal{K}$ is an $\mathcal{ALCQ}^+$ KB, redirect only outgoing r-edges, and drop the incoming ones. Multiple nodes e may be attached to the same bag $\mathcal{M}'$ in this way, but they are only r-reachable from each other via elements of $\mathcal{M}'$, so any simple r-path in $\mathcal{F}$ has length bounded by $2k = \ell$. It is routine to check that $\mathcal{F} \models \mathcal{K}$. By Fact 3, $\mathcal{F} \not\models Q^{(\ell)}$; by Fact 4, $\mathcal{F} \not\models Q$. Thus, $\mathcal{F}$ is a finite counter-model.

6 Eliminating Transitive Atoms

In this section we make the hardest step towards proving our main result: we reduce finite entailment of $\mathcal{UCQ}^+$s to finite entailment of $\mathcal{UCQ}$s, for single-role ABox-trivial KBs either in $\mathcal{ALCI}^+$ or in $\mathcal{ALCQ}^+$. Most of the argument is not only shared for the two DLs, but works for their common extension $(\mathcal{ALCIQ}^-)^+$. Thus, throughout this section we let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a single-role $(\mathcal{ALCIQ}^-)^+$ KB with a trivial ABox. We also fix some $Q \in \mathcal{UCQ}^+$ and a set Θ

of unary types. The overall strategy is to show that finite counter-models can be unravelled into safe counter-models admitting tree decompositions with bags of arbitrary size, but with a certain more subtle width measure bounded. The bags of these decompositions will be almost strongly connected, which will allow dropping transitive atoms when evaluating fragments of Q over a single bag. Moreover, each bag will contain only a bounded number of elements relevant for at-most restrictions over closures of roles, meaning that they can be replaced with nominals (for each bag separately). This way, the existence of bags that together form a safe counter-model can be reduced to multiple instances of finite entailment of $\mathcal{UCQ}$s in $\mathcal{ALCOI}^+$ or $\mathcal{ALCOQ}^+$ without at-most restrictions over closures of roles; eliminating nominals as explained in Section 4, one arrives at the base case, solved in the next section. As the last step, from the existence of such safe counter-models one infers the existence of finite counter-models.

Given that we are in the single-role case, we shall be using graph terminology without mentioning the role name. In particular, a *source* is an element without incoming edges, a *sink* is an element without outgoing edges, an *internal* element is one that has both incoming and outgoing edges, and an *isolated* element is one that has neither incoming nor outgoing edges.

Unravelling Finite Counter-Models

We relax the simplistic notion of decomposition used to eliminate multiple roles by allowing bags to share additional elements with their neighbours, provided that the total number of these additional elements is bounded for each bag.

Definition 2 (hybrid decomposition). *A* hybrid decomposition *of an interpretation $\mathcal{I}$ is a tree T in which each node $v \in T$ is labelled with a finite interpretation $\mathcal{I}_v = (\Delta_v, \cdot^{\mathcal{I}_v})$, called a* bag*, a set $\Gamma_v \subseteq \Delta_v$, and an element $f_v \in \Delta_v$, called the* root *of $\mathcal{I}_v$, such that*

1. *$\mathcal{I} = \bigcup_{v \in T} \mathcal{I}_v$;*
2. *for each $e \in \Delta^{\mathcal{I}}$, $\{v \in T \mid e \in \Delta_v\}$ is connected in T;*
3. *for each edge (u, v) in T, $\Delta_u \cap \Delta_v = \{f_v\} \cup (\Gamma_u \cap \Gamma_v)$.*

We let $\widetilde{\Gamma}_u = \{f_v \mid v \text{ is a child of } u\}$ and call $\max_{u \in T} |\Gamma_u|$ the width *of T. An element $d \in \Delta_u$ is fresh in u if u is the root of T or $d \notin \Delta_{u'}$ for the parent u' of u; d is local in u if $d = f_u$ or if $d \notin \Delta_v$ for all neighbours v of u.*

We often blur the distinction between the node v and the interpretation $\mathcal{I}_v$, using the term *bag* for both. Note that each element is fresh in exactly one bag, and the only local element in u that is not fresh in u is f_u (unless $u = \varepsilon$, when f_ε is fresh too).

While Definiton 2 captures the fundamental structural simplicity of our counter-models, the limitations of the target special case impose grittier structural requirements.

Definition 3 (well-formedness). *If $\mathcal{K}$ uses both inverses and counting, a hybrid decomposition T is* well-formed *for $\mathcal{K}$ if $\mathcal{I}_\varepsilon \models \mathcal{A}$ and for each $u \in T$,*

1. *for each concept name A, $A^{\mathcal{I}_u} = A^{\mathcal{I}} \cap \Delta_u$;*

2. *for each child v of u, f_v is a fresh sink or source in $\mathcal{I}_u$ (or a fresh element of Γ_u);*
3. *for each fresh sink or source d in $\mathcal{I}_u$ (or fresh $d \in \Gamma_u$), $d = f_v$ for exactly one child v of u;*
4. *for each $w \in T$, if $f_u \in \Gamma_w$, then $\Gamma_u \subseteq \Gamma_w$;*
5. *for each local element d in $\mathcal{I}_u$ and each concept name A relevant for d: if $d \in A^{\mathcal{I}_u}$ then $d \in \Gamma_u$, and $\Gamma_v \cap A^{\mathcal{I}_v} \subseteq \Gamma_u$ for all $v \in T$ such that f_v is a sink in $\mathcal{I}_u$ or f_u is a non-isolated source in $\mathcal{I}_v$.*

If $\mathcal{K}$ does not use inverses, "sink or source" in items 2 and 3 is replaced with "sink", and the case "f_u is a non-isolated source in $\mathcal{I}_v$" in item 5 is dropped, but additionally we require that each element of $\Gamma_u \setminus \{f_u\}$ is a sink in $\mathcal{I}_u$. If $\mathcal{K}$ does not use counting, we additionally require that $\Gamma_u = \emptyset$ for all $u \in T$.

Note that in a well-formed hybrid decomposition, each element is local in exactly one node.

We work with infinite counter-examples because they have simpler structure, but ultimately we need to make them finite. The following notion guarantees that this is possible.

Definition 4 (safety). *A hybrid decomposition T is* safe *if it admits no infinite sequence of nodes $u_0, u_1, \ldots$ such that for all i, either $f_{u_{i+1}}$ is a sink in $\mathcal{I}_{u_i}$ or f_{u_i} is a non-isolated source in $\mathcal{I}_{u_{i+1}}$.*

Finally, replacing $\mathcal{UCQ}^+$ with $\mathcal{UCQ}$ is possible thanks to strong connectedness guarantees on the bags.

Definition 5 (well-connectedness). *An interpretation $\mathcal{J}$ is* well-connected *if for all $d, e \in \Delta^{\mathcal{J}}$, $(d, e) \in (r^*)^{\mathcal{J}}$ iff either $d = e$ or d is not a sink and e is not a source in $\mathcal{J}$. A hybrid decomposition T is* well-connected *if $\mathcal{I}_v$ is well-connected for all $v \in T$.*

In the remainder, by a $\mathcal{K}$-decomposition we mean a hybrid decomposition well-formed for $\mathcal{K}$, of width at most $N^{|\mathsf{RCN}(\mathcal{K})|}$ where N is the counting threshold in $\mathcal{K}$.

Lemma 2. *If $\mathcal{K} \not\models_{\mathsf{fin}}^\Theta Q$, then there exists a counter-model admitting a safe and well-connected $\mathcal{K}$-decomposition.*

Proof. Let $\mathcal{J}$ be a finite counter-model. We construct the special counter-model $\mathcal{I}$ and a witnessing $\mathcal{K}$-decomposition T of $\mathcal{I}$ by unravelling $\mathcal{J}$ in a special way. We describe the construction for the case when $\mathcal{K}$ uses inverses; if it does not, simply replace "neighbours" with "direct successors".

Because the ABox of $\mathcal{K}$ is trivial, we have $\mathsf{ind}(\mathcal{K}) = \{d\}$. We begin from the interpretation $\mathcal{I}_\varepsilon$, obtained from $\mathcal{J} \restriction \{d\}$ by removing the only possible edge, $f_\varepsilon = d$, and $\Gamma_\varepsilon = \emptyset$. Now, roughly, for each previously added element e missing a neighbour, we will be adding a new bag containing e with all its missing neighbours, along with all other elements from $\mathsf{cmp}_{\mathcal{J}}(e)$—the strongly connected component of e in $\mathcal{J} \restriction (\Delta^{\mathcal{J}} \setminus \mathsf{rel}_{\mathcal{J}}(e))$, or $\{e\}$ if $e \in \mathsf{rel}_{\mathcal{J}}(e)$—with all their neighbours, and $\mathsf{rel}_{\mathcal{J}}(e)$. If an element f' in this bag corresponds to $f \in \Delta^{\mathcal{J}}$, we call f' a *copy of* f and f the *original of* f'. For convenience, we extend this nomenclature to the root bag: if $f' \in \Delta_\varepsilon$, then its original $f \in \Delta^{\mathcal{J}}$ is f' itself; if $f \in \Delta^{\mathcal{J}}$ belongs to Δ_ε, then its copy f' in Δ_ε is f itself. We proceed as follows, as long as there is something to do.

For each previously added node u and each sink or source $d' \in \Delta_u$ (or $d' \in \Gamma_u$) fresh in u, we add a new child v of u. Let $d \in \Delta^{\mathcal{J}}$ be the original of d' and let $X \subseteq \Delta^{\mathcal{J}}$ be the set of the originals of neighbours of d' in $\mathcal{I}_u$. The interpretation $\mathcal{I}_v$ is obtained by

- taking the restriction of $\mathcal{J}$ to the subdomain comprising $\mathsf{rel}_{\mathcal{J}}(d)$, $\mathsf{cmp}_{\mathcal{J}}(d)$, neighbours of $\mathsf{cmp}_{\mathcal{J}}(d) \setminus \{d\}$, and neighbours of d that do not belong to X;
- removing all edges that are not incident with $\mathsf{cmp}_{\mathcal{J}}(d)$ and all edges between d and elements from X (if $\mathcal{K}$ does not use inverses, drop edges outgoing from $\mathsf{rel}_{\mathcal{J}}(d) \setminus \{d\}$);
- replacing each element e with a fresh copy e', except that for elements $e \in \{d\} \cup \mathsf{rel}_{\mathcal{J}}(d)$ that already have a copy e' in $\mathcal{I}_u$, e' is reused in $\mathcal{I}_v$.

We let $f_v = d'$ and $\Gamma_v = \{e' \mid e \in \mathsf{rel}_{\mathcal{J}}(d)\}$.

By construction, T is a hybrid decomposition of width bounded by $\max_{d \in \Delta^{\mathcal{J}}} |\mathsf{rel}_{\mathcal{J}}(d)| \leq N^{|\mathsf{RCN}(\mathcal{K})|}$ (Fact 1). It is also not difficult to check that T is safe, well connected, and well formed, and that $\mathcal{I} \models \mathcal{K}$ (see Appendix C). Moreover, mapping each element of $\mathcal{I}$ to its original in $\mathcal{J}$ gives a homomorphism from $\mathcal{I}$ to $\mathcal{J}$, which implies that $\mathcal{I} \not\models Q$, and that $\mathcal{I}$ only realizes types from Θ. $\square$

Evaluating Queries over Unravellings

We aim at distributing query evaluation over bags, like when eliminating multiple roles. This is now harder because bags share more than one element, but it is possible because the total number of additional shared elements is bounded for each bag. These elements will be *parameters* of Q-types.

Definition 6 (nullary Q-types). *A (nullary) Q-type with parameters Γ is a set of pairs (p, η) where p is a fragment of Q and η is a partial function from $var(p)$ to Γ. The Q-type of $\mathcal{J}$ with parameters Γ is the set $\mathsf{tp}_Q^{\mathcal{J}}(\Gamma)$ of pairs (p, η) where p is a fragment of Q, and η is a partial function from $var(p)$ to Γ that can be extended to a matching for p in $\mathcal{J}$.*

The power of Q-types with parameters is *compositionality*. Consider interpretations $\mathcal{J}_1, \mathcal{J}_2$ and parameter sets Γ_1, Γ_2. If $\Gamma' \subseteq \Gamma_1$, then

$$\mathsf{tp}_Q^{\mathcal{J}_1}(\Gamma') = \mathsf{tp}_Q^{\mathcal{J}_1}(\Gamma_1) \restriction \Gamma',$$

where $\tau_1 \restriction \Gamma' = \{(p, \eta \restriction \Gamma') \mid (p, \eta) \in \tau_1\}$ is the *projection* of τ_1 on Γ'. If $\Delta^{\mathcal{J}_1} \cap \Delta^{\mathcal{J}_2} \subseteq \Gamma_1 \cap \Gamma_2$, then

$$\mathsf{tp}_Q^{\mathcal{J}_1 \cup \mathcal{J}_2}(\Gamma_1 \cup \Gamma_2) = \mathsf{tp}_Q^{\mathcal{J}_1}(\Gamma_1) \oplus \mathsf{tp}_Q^{\mathcal{J}_2}(\Gamma_2),$$

where $\tau_1 \oplus \tau_2$ is the *composition* of τ_1 and τ_2 defined as the set of tuples (p, η) such that there exist $\tau \subseteq \tau_1 \cup \tau_2$ with $var(p') \setminus \mathsf{dom}(\eta')$ pairwise disjoint for $(p', \eta') \in \tau$ and a homomorphism $\widetilde{\eta} : p \to \bigcup_{(p', \eta') \in \tau} \eta'(p')$ extending η.

Like before, we decorate elements with their Q-types.

Definition 7 (Q-labellings). *A Q-labelling for a $\mathcal{K}$-decomposition T of $\mathcal{I}$ is a function σ that maps each element d local in $u \in T$ to a Q-type $\sigma(d)$ with parameters $\{d\} \cup \Gamma_u$. We call σ correct if for each d local in u, $\sigma(d)$ is the Q-type of $\widehat{\mathcal{I}}_u$ with parameters $\{d\} \cup \Gamma_u$, where $\widehat{\mathcal{I}}_u$ is the interpretation represented by the subtree of T rooted at u.*

If an interpretation $\mathcal{I}$ admits a $\mathcal{K}$-decomposition, then $\mathcal{I} \not\models Q$ iff the correct Q-labelling for any $\mathcal{K}$-decomposition of $\mathcal{I}$ uses only Q-types in which no $q \in Q$ appears. Such Q-labellings, called Q-*refutations*, are what we need to find.

Given a $\mathcal{K}$-decomposition, we can (coinductively) compute the correct Q-labelling bottom-up, by composing and projecting Q-types of the current bag with the Q-types of its direct subtrees, as captured in the following lemma.

Lemma 3. *The correct Q-labelling for a $\mathcal{K}$-decomposition T of an interpretation $\mathcal{I}$ is the least (pointwise) Q-labelling σ for T such that*

$$\sigma(d) = \left(\mathsf{tp}_Q^{\mathcal{I}_u}(\{d\} \cup \Gamma_u \cup \widetilde{\Gamma}_u) \oplus \bigoplus_{e \in \widetilde{\Gamma}_u} \sigma(e) \right) \restriction \{d\} \cup \Gamma_u$$

for each node $u \in T$ and each element d local in u.

From $\mathcal{UCQ}^+$ to $\mathcal{UCQ}$

The condition in Lemma 3 provides the locality required in automata-based decision procedures, but it still relies on evaluating $\mathcal{UCQ}^+$s over bags. We now show how to avoid it.

Definition 8 (localization). *A localization of a $\mathcal{CQ}^+$ q is any $\mathcal{CQ}$ obtainable from q by replacing each atom $r^*(x, y)$ either with $x = y$ or with $r(x, y'), r(x', y)$ for some fresh variables x' and y' used only in these two atoms.*

Every matching for a $\mathcal{UCQ}^+$ q in an interpretation $\mathcal{J}$ extends to a matching for some localization of q. So, if $\mathcal{J}$ does not satisfy any localization of q, then $\mathcal{J} \not\models q$. Moreover, if $\mathcal{J}$ is well-connected, then each matching for a localization of q in $\mathcal{J}$ induces a matching for q. So, for well-connected $\mathcal{J}$, $\mathcal{J} \not\models q$ iff $\mathcal{J}$ does not satisfy any localization of q.

Definition 9 (weak Q-types). *The weak Q-type of $\mathcal{J}$ with parameters Γ is the set $\mathsf{wtp}_Q^{\mathcal{J}}(\Gamma)$ of pairs (p, η) such that p is a fragment of Q and η is a partial function from $var(p)$ to Γ that extends to a matching of some localization of p in $\mathcal{J}$.*

Weak Q-types are over-approximations of Q-types,

$$\mathsf{tp}_Q^{\mathcal{J}}(\Gamma) \subseteq \mathsf{wtp}_Q^{\mathcal{J}}(\Gamma),$$

and are exact for well-connected interpretations,

$$\mathcal{J} \text{ is well-connected} \implies \mathsf{tp}_Q^{\mathcal{J}}(\Gamma) = \mathsf{wtp}_Q^{\mathcal{J}}(\Gamma).$$

By replacing Q-types with weak Q-types in the condition of Lemma 3, we avoid evaluating fragments of $\mathcal{UCQ}^+$ by evaluating their localizations instead. We also relax the condition by replacing equality with inclusion, to facilitate the reduction to finite (non-)entailment.

Definition 10 (consistency). *A Q-labelling σ for a $\mathcal{K}$-decomposition T of an interpretation $\mathcal{I}$ is consistent if*

$$\sigma(d) \supseteq \left(\mathsf{wtp}_Q^{\mathcal{I}_u}(\{d\} \cup \Gamma_u \cup \widetilde{\Gamma}_u) \oplus \bigoplus_{e \in \widetilde{\Gamma}_u} \sigma(e) \right) \restriction \{d\} \cup \Gamma_u$$

for each node $u \in T$ and each element d local in u.

For any $\mathcal{K}$-decomposition T, the existence of a consistent Q-refutation implies the existence of a correct Q-refutation. Moreover, if T is well-connected, then the two conditions coincide, which eliminates false negatives in the finite entailment decision procedure.

Lemma 4. *If a $\mathcal{K}$-decomposition T of $\mathcal{I}$ admits a consistent Q-refutation, then it admits a correct Q-refutation. If T is well-connected, then the correct Q-labelling is consistent.*

Proof. Let σ be the correct Q-labelling for T. By Lemma 3 and the relation between Q-types and weak Q-types, for each consistent Q-labelling σ' for T, $\sigma(v) \subseteq \sigma'(v)$ for all $v \in T$. Consequently, if some consistent Q-refutation σ' for T exists, then σ is a Q-refutation as well. As σ is correct, the first claim of the lemma follows. If T is well-connected, Q-types and weak Q-types coincide. Combined with Lemma 3, this implies that σ is consistent. $\square$

Lemma 5. *If $\mathcal{K} \not\models_{\mathsf{fin}}^{\Theta} Q$ then some model of $\mathcal{K}$ realizing only types from Θ has a safe $\mathcal{K}$-decomposition admitting a consistent Q-refutation.*

Proof. Assume $\mathcal{K} \not\models_{\mathsf{fin}}^{\Theta} Q$. By Lemma 2, there exists a counter-model $\mathcal{I}$ with a safe and well-connected $\mathcal{K}$-decomposition T. Let σ be the unique correct Q-labelling for T. Because $\mathcal{I} \not\models Q$, σ is a Q-refutation. By Lemma 4, σ is consistent. $\square$

Recognizing Safe Counter-Examples

We now construct an automaton $\mathcal{B}_{\mathcal{K},Q}$ recognizing safe $\mathcal{K}$-decompositions of models of $\mathcal{K}$, admitting consistent Q-refutations. Like in the case of multiple roles, we must give the automaton information about shared elements. For this purpose, we enrich each node v with a tuple $\mathbf{g}_v$ enumerating all elements of Γ_v that belong to the parent u of v, and on the edge between u and v we put f_v and $\mathbf{g}_v$. Again, the actual elements in Δ_v and Δ_u do not matter: f_v and $\mathbf{g}_v$ coupled with the label on the edge from u to v determine which elements represented in v and in u should be identified. Let $M = N^{|\mathsf{RCN}(\mathcal{K})|}$ where N is the counting threshold in $\mathcal{K}$. We fix a countably infinite set Δ and assume that the node alphabet of our automaton is the set of 4-tuples of the form $(\mathcal{J}, f, \Gamma, \mathbf{g})$ where $\mathcal{J}$ is a finite interpretation with $\Delta^{\mathcal{J}} \subseteq \Delta$, $f \in \Delta^{\mathcal{J}}$, $\Gamma \subseteq \Delta^{\mathcal{J}}$, $|\Gamma| \leq K$, and $\mathbf{g}$ is a tuple enumerating a subset of Γ. The edge alphabet is $\Delta \times \bigcup_{i=0}^{M} \Delta^i$.

The following lemma relies on $\mathcal{K}$ being expressed either in $\mathcal{ALCI}^{+}$ or in $\mathcal{ALCQ}^{+}$. By a base-case KB we understand a single-role KB in $\mathcal{ALCI}^{+}$ or $\mathcal{ALCQ}^{+}$, respectively, with a trivial ABox and no at-most restrictions over closures of roles. Let $m = \max_{q \in Q} |q|$.

Lemma 6. *There exists an automaton $\mathcal{B}_{\mathcal{K},Q,\Theta}$ recognizing those safe $\mathcal{K}$-decompositions of models of $\mathcal{K}$ realizing only types from Θ that admit consistent Q-refutations. The states, initial states, and accepting states of $\mathcal{B}_{\mathcal{K},Q,\Theta}$ can be computed in time $O\big(2^{\mathrm{poly}(\|Q\|,\|\mathcal{K}\|,M^m)}\big)$. The question if $(P, q) \in \mathsf{step}_{\mathcal{B}_{\mathcal{K},Q,\Theta}}$ can be reduced to an instance of finite entailment modulo types for a base case KB of size at most $\mathrm{poly}(\|\mathcal{K}\|, \|Q\|, N \cdot M^m)$ and counting threshold N, a $\mathcal{UCQ}$ consisting of at most $\|Q\|^2 \cdot M^m \cdot m^m$ CQs of size $O(m)$, and a type set of size at most $|\Theta| \cdot 2^{\mathrm{poly}(\|K\|, \|Q\|, N \cdot M^m)}$.*

Making Counter-Examples Finite

To complete the proof we need to derive finite non-entailment from the existence of a safe counter-example.

Lemma 7. *If $\mathcal{K}$ is in $\mathcal{ALCI}^{+}$ or $\mathcal{ALCQ}^{+}$ and some model of $\mathcal{K}$ realizing only types from Θ has a safe $\mathcal{K}$-decomposition admitting a consistent Q-refutation, then $\mathcal{K} \not\models_{\mathsf{fin}}^{\Theta} Q$.*

For $\mathcal{ALCQ}^{+}$, a finite counter-model comes for free. Indeed, each safe hybrid decomposition well-formed for a KB without inverses must be finite, so the interpretation it represents is finite too.

For $\mathcal{ALCI}^{+}$, we use the coloured blocking principle. By Lemma 6, the set of those $\mathcal{K}$-decompositions of models of $\mathcal{K}$ that admit a weakly-consistent Q-refutation is recognized by an automaton. Consequently, if nonempty, it contains a regular tree T; that is, T has only finitely many non-isomorphic subtrees. From the regularity of T it follows that $\mathcal{I}_v$ are chosen from a finite set, which means that their size is bounded. Similarly, the branching of T is bounded. Finally, safety and regularity of T together imply that there is a bound on the length of sequence of nodes $u_0, u_1, \ldots, u_k$ such that for all $i < k$, either f_{u_i} is a source in $\mathcal{I}_{u_{i+1}}$ or $f_{u_{i+1}}$ is a sink in $\mathcal{I}_{u_i}$. It follows that in the interpretation $\mathcal{I}$ represented by T, both the branching and the length of simple directed paths is bounded by some $\ell \in \mathbb{N}$.

By Fact 4, $\mathcal{I} \models Q$ iff $\mathcal{I} \models Q^{(\ell)}$, where $Q^{(\ell)} \in \mathcal{UCQ}$ is the query defined in Section 5. Fix $t = \max_{q \in Q^{(\ell)}} |q|$ and $n = \max(\ell^2, t^2)$. Let $\mathcal{I}'$ be an n-proper colouring of $\mathcal{I}$. Let V be the set of nodes $v \in T$ such that f_v is a sink in $\mathcal{I}_{v'}$ for the parent v' of v and $f_{v'}$ is a source in $\mathcal{I}_{v''}$ for the parent v'' of v'. By the boundedness properties of T, each infinite branch of T has infinitely many nodes in V. On each infinite branch, select the first node $v \in V$ with an ancestor $\hat{v} \in V$ such that the n-neighbourhood of f_v is isomorphic with the n-neighbourhood of $f_{\hat{v}}$. The set of selected nodes forms a maximal antichain in T and, by König's Lemma, it is finite. Let $\mathcal{F}$ be the interpretation obtained by taking the union of $\mathcal{I}_u$ for u ranging over the nodes of T that have a selected descendent, except that for each selected node v, f_v is removed from the domain and all incoming edges are redirected to $f_{\hat{v}}$.

Lemma 8. *The length of simple directed paths in $\mathcal{F}$ is bounded by ℓ and $\mathcal{F} \models \mathcal{K}$.*

By Fact 3, $\mathcal{F} \not\models Q^{(\ell)}$, and by Fact 4 and Lemma 8, $\mathcal{F} \not\models Q$. Thus, $\mathcal{F}$ is a finite counter-model.

Wrapping up

Assume $\mathcal{K}$ is in $\mathcal{ALCI}^{+}$ or $\mathcal{ALCQ}^{+}$. By Lemmas 5 and 7, $\mathcal{K} \not\models_{\mathsf{fin}}^{\Theta} Q$ iff some model of $\mathcal{K}$ realizing only types from Θ has a safe $\mathcal{K}$-decomposition admitting a weakly consistent Q-refutation. Combining Lemma 6 and Fact 2, we get a decision procedure for $\mathcal{K} \models_{\mathsf{fin}}^{\Theta} Q$ running in time $O\big(\mathrm{poly}(|\Theta|) \cdot 2^{\mathrm{poly}(\|Q\|, \|\mathcal{K}\|, M^m)}\big)$ and using oracle calls to finite entailment modulo types for instances with parameters bounded as in Lemma 6.

7 Base Case

We now solve finite entailment modulo types under the assumptions justified in previous sections; that is, for Q in $\mathcal{UCQ}$ and a single-role ABox-trivial $\mathcal{K}$ either in $\mathcal{ALCI}^+$ or in $\mathcal{ALCQ}^+$ without at-most restrictions over closures of roles. We show that it suffices to consider tree-shaped counter-models (Lemma 9), and that they consitute an effectively regular set (Lemma 10). The decision procedure then amounts to computing the automaton and testing its emptiness; it runs in time $\mathrm{poly}(|\Theta|, 2^{\|\mathcal{K}\|^2 \cdot \|Q\|})$ where Θ is the set of allowed types.

An interpretation $\mathcal{J}$ is *tree-shaped* if its elements can be arranged into a tree such that the unique $f \in \mathrm{ind}(\mathcal{K})$ is the root and r-edges in $\mathcal{J}$ are allowed only between parents and children; if $\mathcal{K}$ is an $\mathcal{ALCQ}^+$ KB, r-edges must point down.

Lemma 9. $\mathcal{K} \not\models_{\mathrm{fin}}^{\Theta} Q$ *iff there exists a tree-shaped counter-model of bounded degree realizing only types from Θ.*

Proof. Let $\mathcal{I}$ be a finite counter-model realizing only types from Θ. We get a tree-shaped counter-model $\mathcal{I}'$ by unravelling $\mathcal{I}$ from the unique $f \in \mathrm{ind}(\mathcal{K})$. If $\mathcal{K}$ is in $\mathcal{ALCQ}^+$, we unravel by adding fresh copies of all direct successors of previously copied elements from $\mathcal{I}$. If $\mathcal{K}$ is in $\mathcal{ALCI}^+$, we add fresh copies of all direct successors and fresh copies of all direct predecessors. It is clear that $\mathcal{I}'$ is a tree-shaped counter-model realizing only types from Θ, of degree bounded by twice the degree of $\mathcal{I}$.

Conversely, let $\mathcal{I}$ be a tree-shaped counter-model of bounded degree realizing only types from Θ.

If $\mathcal{K}$ is in $\mathcal{ALCQ}^+$, we enrich $\mathcal{I}$ to ensure that witnesses for at-least restrictions over r^* are never glued with each other. Note that in this case all r-edges in $\mathcal{I}$ point down. Let N be the counting threshold in $\mathcal{K}$. For each concept name A, introduce fresh concepts names $A_1, A_2, \ldots, A_N$, called the *shades* of A. Extend $\mathcal{I}$ to these concept names in such a way that $A_1^{\mathcal{I}}, A_2^{\mathcal{I}}, \ldots, A_N^{\mathcal{I}}$ form a partition of $A^{\mathcal{I}}$, and for each $n \leq N$, if $d \in \Delta^{\mathcal{I}}$ has at least n successors in A, then it has successors in at least n shades of A. This can be done greedily, by processing $\mathcal{I}$ top down, ensuring that subtrees rooted at unprocessed nodes use each shade of A at most once. Take a shallowest unprocessed node v. For $n = 1, 2, \ldots, N$, if v has at least n descendents in A but only $n - 1$ are painted with a shade of A, pick an unpainted one and paint it with an unused shade of A. After the whole tree is processed, paint each remaining element in A with an arbitrary shade of A.

For each concept name B (including the shades) enrich $\mathcal{I}$ further by introducing a fresh concept name B' with extension $(\exists r^*.B)^{\mathcal{I}}$ and add $B' \sqsubseteq \exists r^*.B$ to $\mathcal{K}$; if $\mathcal{K}$ is in $\mathcal{ALCI}^+$, introduce also B'' with extension $(\exists (r^-)^*.B)^{\mathcal{I}}$ and add $B'' \sqsubseteq \exists (r^-)^*.B$. Let $n = \max(2, |Q|^2)$ and let $\mathcal{I}'$ be an n-proper colouring of $\mathcal{I}$. Consider a level l in $\mathcal{I}'$ such that all n-neighbourhoods realized in $\mathcal{I}'$ are already realized above l. For each element above level l, choose a witness for each CI of the form $A \sqsubseteq \exists r^*.B$ or $A \sqsubseteq \exists (r^-)^*.B$ in $\mathcal{K}$. Restrict the domain to the nodes above level l and the chosen witnesses, together with the paths that lead to them.

Redirect each edge leaving the restricted domain to some element above level l, preserving the n-neighbourhood.

The only nontrivial thing to check is that at-least restrictions are not violated, in the case when $\mathcal{K}$ is in $\mathcal{ALCQ}^+$. For restrictions over r, this is because siblings have different colours in each n-proper colouring, so edges leading to different siblings are never redirected to the same element. For restrictions over r^*, this is because witnesses for the CI $A_i' \sqsubseteq \exists r^*.A_i$ are preserved for each shade A_i of A. $\qquad\square$

Lemma 10. *The set of tree-shaped counter-models realizing only types from Θ is recognized by an automaton computable in time $2^{O(\|\mathcal{K}\|^2 \cdot \|Q\|)}$.*

This is a routine construction. For $\mathcal{ALCQ}^+$, the node alphabet is Θ and the edge alphabet is trivial. To verify that the input tree is a model of $\mathcal{K}$, the automaton stores the unary type of the parent of the current node, and for each $A \in \mathsf{CN}(\mathcal{K})$, it stores in the state the minimum number of elements in A to be found in the current subtree, together with a binary flag indicating if progress has been made recently in finding them. This component of the automaton has $(4 \cdot N)^{|\mathsf{CN}(\mathcal{K})|}$ states, where N is the counting threshold in $\mathcal{K}$. Verifying that the input tree does not satisfy Q involves storing a set of subqueries of CQs from Q, which are not to be satisfied in the current subtree. This component has $2^{\|Q\|}$ states. The whole automaton is the product of the two components. It is easy to see that the step relation can be computed in time polynomial in the number of states.

The case of $\mathcal{ALCI}^+$ is similar, except that the edge alphabet is $\{r, r^-\}$ and the first component has separate information about elements reachable and backwards reachable from current node in the current subtree, and also outside (without progress flags). The first component thus has $2^{7 \cdot |\mathsf{CN}(\mathcal{K})|}$ states, because it only needs to count up to 1.

Note that if the automaton accepts any tree, it also accepts a regular one, and a regular tree has bounded degree.

8 Outlook

This paper provides first positive results on finite entailment of non-local queries over DLs knowledge bases. The main technical contribution is optimal 2ExpTime upper bounds for finite entailment of $\mathcal{UCQ}$'s over $\mathcal{ALCOI}^+$ and $\mathcal{ALCOQ}^+$ knowledge bases. To obtain these results, we have shown intermediate reductions that are interesting in their own, and could be applied to similar settings.

There are several directions to follow for future work. A first possibility is to vary the DL language. One could consider lightweight DLs from the $\mathcal{EL}$ and *DL-Lite* families or extensions of $\mathcal{ALCOI}^+$ with e.g. role inclusions. For the latter, a different approach to the one proposed here is needed because our techniques rely on the lack of interaction between different roles. Another option is to allow for controlled interaction of inverses and number restrictions as e.g. in $(\mathcal{ALCIQ}^-)^+$. A second possibility is to consider more expressive non-local queries, such as positive regular path queries. In this case new techniques seem to be needed, e.g. the coloured blocking principle does not work for PRPQs.

Acknowledgments

References

Amarilli, A., and Benedikt, M. 2015. Finite open-world query answering with number restrictions. In *LICS-15*, 305–316.

Baader, F.; Lutz, C.; Sturm, H.; and Wolter, F. 2002. Fusions of description logics and abstract description systems. *J. Artif. Intell. Res. (JAIR)* 16:1–58.

Baader, F.; Horrocks, I.; Lutz, C.; and Sattler, U. 2017. *An Introduction to Description Logic*. Cambridge University Press.

Bednarczyk, B., and Rudolph, S. 2019. Worst-case optimal querying of very expressive description logics with path expressions and succinct counting. In *IJCAI*.

Bienvenu, M.; Ortiz, M.; and Simkus, M. 2015. Regular path queries in lightweight description logics: Complexity and algorithms. *J. Artif. Intell. Res. (JAIR)* 53:315–374.

Calvanese, D.; Eiter, T.; and Ortiz, M. 2014. Answering regular path queries in expressive description logics via alternating tree-automata. *Inf. Comput.* 237:12–55.

Chandra, A. K.; Kozen, D. C.; and Stockmeyer, L. J. 1981. Alternation. *Journal of the ACM* 28:114–133.

Danielski, D., and Kieronski, E. 2019. Finite satisfiability of unary negation fragment with transitivity. In *In Proc. of MFCS-19*.

Eiter, T.; Lutz, C.; Ortiz, M.; and Simkus, M. 2009. Query answering in description logics with transitive roles. In *Proc. of IJCAI-09*.

Gogacz, T.; Gutiérrez-Basulto, V.; Ibáñez-García, Y.; Jung, J. C.; and Murlak, F. 2019. On finite and unrestricted query entailment beyond SQ with number restrictions on transitive roles. In *IJCAI-19*.

Gogacz, T.; Ibáñez-García, Y.; and Murlak, F. 2018. Finite query answering in expressive description logics with transitive roles. In *Proc. of KR-18*.

Gutiérrez-Basulto, V.; Ibáñez-García, Y.; and Jung, J. C. 2018. Answering regular path queries over SQ ontologies. In *Proc. of AAAI-18*.

Ibáñez-García, Y.; Lutz, C.; and Schneider, T. 2014. Finite model reasoning in horn description logics. In *Proc. of KR-14*.

Jung, J. C.; Lutz, C.; Martel, M.; and Schneider, T. 2017. Querying the unary negation fragment with regular path queries. In *Proc. of ICDT'17*.

Ladner, R.; Lynch, N.; and Selman, A. 1975. A comparison of polynomial time reducibilities. *Theoretical Computer Science* 1(2):103–123.

Lutz, C. 2008. The complexity of conjunctive query answering in expressive description logics. In *Proc. of IJCAR 2008*, 179–193.

Ngo, N.; Ortiz, M.; and Simkus, M. 2016. Closed predicates in description logics: Results on combined complexity. In *Proc. of KR-16*, 237–246.

Post, E. L. 1944. Recursively enumerable sets of positive integers and their decision problems. *Bull. Amer. Math. Soc.* 50(5):284–316.

Rosati, R. 2008. Finite model reasoning in DL-Lite. In *Proc. of ESWC-19*, 215–229.

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

Stefanoni, G.; Motik, B.; Krötzsch, M.; and Rudolph, S. 2014. The complexity of answering conjunctive and navigational queries over OWL 2 EL knowledge bases. *J. Artif. Intell. Res.* 51:645–705.