Finite Query Answering in Expressive Description Logics with Transitive Roles

Tomasz Gogacz
University of Warsaw, Poland
t.gogacz@mimuw.edu.pl

Yazmín Ibáñez-García
TU Wien, Austria
yazmin.garcia@tuwien.ac.at

Filip Murlak
University of Warsaw, Poland
fmurlak@mimuw.edu.pl

Abstract

We study the problem of *finite* ontology mediated query answering (FOMQA), the variant of OMQA where the represented world is assumed to be finite, and thus only finite models of the ontology are considered. We adopt the most typical setting with unions of conjunctive queries and ontologies expressed in description logics (DLs). The study of FOMQA is relevant in settings that are not finitely controllable. This is the case not only for DLs without the finite model property, but also for those allowing transitive role declarations. When transitive roles are allowed, evaluating queries is challenging: FOMQA is undecidable for $\mathcal{SHOIF}$ and only known to be decidable for the Horn fragment of $\mathcal{ALCIF}$. We show decidability of FOMQA for three proper fragments of $\mathcal{SOIF}$: $\mathcal{SOI}$, $\mathcal{SOF}$, and $\mathcal{SIF}$. Our approach is to characterise models relevant for deciding finite query entailment. Relying on a certain regularity of these models, we develop automata-based decision procedures with optimal complexity bounds.

Introduction

Evaluating queries in the presence of background knowledge has been extensively studied in several communities. A particularly prominent take on this problem is ontology mediated query answering (OMQA) where background knowledge represented by an ontology is leveraged to infer more complete answers to queries (Bienvenu and Ortiz 2015). A widely accepted family of ontology languages with varying expressive power is offered by Description Logics (DLs) (Baader et al. 2010), while the most commonly studied query language is that of (unions of) conjunctive queries.

Often, the intended models of the ontology are finite and this additional assumption allows to infer more properties: *finite ontology mediated query answering (FOMQA)* is the variant of OMQA restricted to finite models. For some logics the finite variant and the unrestricted variant of the problem coincide; we then say that OMQA is *finitely controllable*. Studying FOMQA is interesting in settings lacking finite controllability. This is the case not only for DLs lacking the *finite model property* (e.g., DLs allowing both inverse roles and number restrictions), but also for logics allowing *transitive role declarations*. Indeed, it has been recently proved that FOMQA is undecidable for $\mathcal{SHOIF}$ ontologies (Rudolph 2016), whereas the only fragment known to be decidable is Horn-$\mathcal{ALCIF}$ (Ibáñez-García, Lutz, and Schneider 2014); more expressive fragments of $\mathcal{SHOIF}$ are entirely uncharted. In this paper, we establish decidability for three of them: $\mathcal{SOI}$, $\mathcal{SOF}$, and $\mathcal{SIF}$.

OMQA is closely related to *query answering under integrity constraints* in database theory: given a finite database instance and a set of constraints, determine answers to a query that are certain to hold over any extension of the given instance that satisfies the constraints. Among important classes of constraints are inclusion dependencies (IDs) and functional dependencies (FDs). This problem, often called *open-world query answering (OWQA)*, has also been studied in the variant considering only finite extensions of the given database instance (finite OWQA), which is directly relevant for our work. OWQA over IDs is known to be finitely controllable (Johnson and Klug 1984; Rosati 2011). Rosati's techniques were extended to show finite controllability for the guarded fragment of first order logic (Bárány, Gottlob, and Otto 2014). Under combinations of IDs and FDs, OWQA is undecidable, both unrestricted and finite, but multiple decidable fragments have been isolated. For instance, for *non-conflicting* IDs and FDs (Calì, Lembo, and Rosati 2003), unrestricted OWQA is decidable. However, finite OWQA is undecidable already for non-conflicting IDs and keys, which are less expressive than FDs (Rosati 2011). The work of (Amarilli and Benedikt 2015) investigates finite OWQA for unary IDs and FDs over arbitrary signatures.

Combinations of unary IDs and unary FDs can be expressed in relatively simple DLs. This relationship and the techniques developed by (Cosmadakis, Kanellakis, and Vardi 1990) have been exploited in the study of finite satisfiability for simple DLs (Rosati 2008). Indeed, finite satisfiability has been studied extensively (Calvanese 1996; Lutz, Sattler, and Tendera 2005; Kazakov 2008; Pratt-Hartmann 2007), but FOMQA has received limited attention in the DL community. The mentioned results on the guarded fragment give finite controllability for DLs up to $\mathcal{ALCHOI}b$. For non-finitely-controllable DLs, only the already mentioned results about $\mathcal{SHOIF}$ and Horn-$\mathcal{ALCIF}$ are known. For Datalog$^{\pm}$, finite controllability holds for several fragments (Gogacz and Marcinkowski 2017; Amendola, Leone, and Manna 2017; Baget et al. 2011; Civili and Rosati 2012). Finally, (Pratt-Hartmann 2009) studies finite query answering for expressive fragments of first order logic and establishes undecidability for the two variable fragment with counting quantifiers ($\mathcal{C}^2$), and decidability for its guarded fragment, $\mathcal{GC}^2$. Decidability of $\mathcal{GC}^2$ has no direct implications for DLs with nominals or transitive roles, but it proves useful in the study of $\mathcal{SIF}$.

Contributions. We show that the combined complexity of FOMQA is in 2EXPTIME for $\mathcal{SOI}$, $\mathcal{SOF}$ and $\mathcal{SIF}$. These bounds are tight by existing matching lower bounds for OMQA for less expressive logics enjoying finite controllability (Ngo, Ortiz, and Simkus 2016; Lutz 2008). We present a direct construction of finite counter-models from arbitrary tree-like counter models for $\mathcal{ALCOI}$, thus re-proving finite controllability. An extension of this construction builds finite counter-models from special tree-like models of $\mathcal{SOI}$ and $\mathcal{SOF}$, which are guaranteed to exist whenever finite counter-models exist. This way finite query entailment reduces to entailment over a certain class of tree-like models recognisable by tree automata. For $\mathcal{SIF}$, we show that to some extent one can separate the reasoning about transitive and non-transitive (possibly functional) roles, and design a procedure that uses the decidability results for $\mathcal{SOI}$ and $\mathcal{ALCIF}$ as black boxes. The latter is derived from the work of (Pratt-Hartmann 2009). All missing arguments can be found in the full version of the paper, available at arxiv.org/abs/?????????.

Preliminaries

The DL $\mathcal{SOIF}$ extends the classical DL $\mathcal{ALC}$ with transitivity declarations on roles ($\mathcal{S}$), nominals ($\mathcal{O}$), inverses ($\mathcal{I}$), and role functionality declarations ($\mathcal{F}$) (Baader et al. 2010). We assume a signature of countably infinite disjoint sets of *concept names* $\mathsf{N_C} = \{A_1, A_2, \dots\}$, *role names* $\mathsf{N_R} = \{r_1, r_2, \dots\}$ and *individual names* $\mathsf{N_I} = \{a_1, a_2, \dots\}$. $\mathcal{SOIF}$*-concepts* C, D are defined by the grammar:

$$C, D ::= \top \mid A \mid \neg C \mid C \sqcap D \mid \{a\} \mid \exists r.C\,,$$

where $r \in \mathsf{N_R} \cup \{r^- \mid r \in \mathsf{N_R}\}$ is a *role*. Roles of the form r^- are called *inverse roles*. A $\mathcal{SOIF}$ *TBox* $\mathcal{T}$ is a finite set of *concept inclusions (CIs)* $C \sqsubseteq D$, *transitivity declarations* $\mathsf{Tr}(r)$, *functionality declarations* $\mathsf{Fn}(r)$, where C, D are $\mathcal{SOIF}$-concepts and r is a role. We assume that if the TBox contains $\mathsf{Tr}(r)$, then it contains neither $\mathsf{Fn}(r)$ nor $\mathsf{Fn}(r^-)$. With an appropriate extension of the signature, each $\mathcal{SOIF}$ TBox can be transformed into an equivalent TBox whose each CI has one of the following normal forms:

$$\textstyle\bigsqcap A_i \sqsubseteq \bigsqcup B_j\,, \quad A \equiv \{a\}\,, \quad A \sqsubseteq \forall r.B\,, \quad A \sqsubseteq \exists r.B\,,$$

where empty conjunction is equivalent to $\top$ and empty disjunction to $\bot$. We also assume that for each concept name A used in $\mathcal{T}$ there is a *complementary* concept name $\bar{A}$ axiomatised with CIs $\top \sqsubseteq A \sqcup \bar{A}$ and $A \sqcap \bar{A} \sqsubseteq \bot$.

$\mathcal{SOI}$, $\mathcal{SOF}$ and $\mathcal{SIF}$ TBoxes are restrictions of $\mathcal{SOIF}$ TBoxes. $\mathcal{SOI}$ TBoxes do not contain functionality declarations, whereas concept inclusions in $\mathcal{SOF}$ and $\mathcal{SIF}$ do not contain inverse roles and nominals, respectively. Because the inverse of a transitive role is transitive anyway, for $\mathcal{SOI}$, $\mathcal{SIF}$, and $\mathcal{SOIF}$ we shall assume that if $\mathsf{Tr}(r)$ is present in the TBox, then so is $\mathsf{Tr}(r^-)$.

An *ABox* is a finite set of *concept* and *role* assertions of the form $A(a)$ and $r(a, b)$, 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 $|\mathcal{K}|$ for $|\mathcal{A}| + |\mathcal{T}|$. We use $\mathsf{CN}(\mathcal{K})$, $\mathsf{Rol}(\mathcal{K})$, $\mathsf{Nom}(\mathcal{K})$, and $\mathsf{Ind}(\mathcal{K})$ to denote, respectively, the set of *all concept names, roles, nominals, and individuals occurring in* $\mathcal{K}$. We stress that if r occurs in $\mathcal{K}$, but r^- does not, then $r^- \notin \mathsf{Rol}(\mathcal{K})$.

A *unary type* is a subset of $\mathsf{CN}(\mathcal{K})$ that contains exactly one of the concept names A, $\bar{A}$ for each $A \in \mathsf{CN}(\mathcal{K})$. We write $\mathsf{Tp}(\mathcal{K})$ for the set of all unary types.

The semantics is defined via interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ with a non-empty *domain* $\Delta^{\mathcal{I}}$ and an *interpretation function* $\cdot^{\mathcal{I}}$ assigning to each $A \in \mathsf{CN}(\mathcal{K})$ a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and to each role name r with $r \in \mathsf{Rol}(\mathcal{K})$ or $r^- \in \mathsf{Rol}(\mathcal{K})$, a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The interpretation of complex concepts and roles is defined as usual (Baader et al. 2010). We only consider interpretations complying with the *standard name assumption* in the sense that $a^{\mathcal{I}} = a$ for every $a \in \mathsf{N_I}$.

An interpretation $\mathcal{I}$ *satisfies* $\alpha \in \mathcal{T} \cup \mathcal{A}$, written as $\mathcal{I} \models \alpha$, if the following holds: if α is a CI $C \sqsubseteq D$ then $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, if α is a transitivity declaration $\mathsf{Tr}(r)$ then $r^{\mathcal{I}}$is transitive, if α is a functionality declaration $\mathsf{Fn}(r)$ then $r^{\mathcal{I}}$ is a partial function, if α is an assertion $A(a)$ then $a \in A^{\mathcal{I}}$, and if α is an assertion $r(a, b)$ then $(a, b) \in r^{\mathcal{I}}$.

Finally, $\mathcal{I}$ is a *model* of: a TBox $\mathcal{T}$, denoted $\mathcal{I} \models \mathcal{T}$, if $\mathcal{I} \models \alpha$ for all $\alpha \in \mathcal{T}$; an ABox $\mathcal{A}$, denoted $\mathcal{I} \models \mathcal{A}$, if $\mathcal{I} \models \alpha$ for all $\alpha \in \mathcal{A}$; and a KB $\mathcal{K}$ if $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \models \mathcal{A}$.

Interpretation $\mathcal{I}$ is a *subinterpretation* of interpretation $\mathcal{J}$, written as $\mathcal{I} \subseteq \mathcal{J}$, if $\Delta^{\mathcal{I}} \subseteq \Delta^{\mathcal{J}}$, $A^{\mathcal{I}} \subseteq A^{\mathcal{J}}$, and $r^{\mathcal{I}} \subseteq r^{\mathcal{J}}$ for all $A \in \mathsf{CN}(\mathcal{K})$, $r \in \mathsf{Rol}(\mathcal{K})$. An interpretation $\mathcal{I}$ is a subinterpretation of $\mathcal{J}$ *induced* by $\Delta_0 \subseteq \Delta^{\mathcal{J}}$, written as $\mathcal{I} = \mathcal{J} \upharpoonright \Delta_0$, if $\Delta^{\mathcal{I}} = \Delta_0$, $A^{\mathcal{I}} = A^{\mathcal{J}} \cap \Delta_0$, and $r^{\mathcal{I}} = r^{\mathcal{J}} \cap \Delta_0 \times \Delta_0$ for all $A \in \mathsf{CN}(\mathcal{K})$, $r \in \mathsf{Rol}(\mathcal{K})$. We write $\mathcal{J} \setminus X$ for the subinterpretation of $\mathcal{J}$ induced by $\Delta^{\mathcal{J}} \setminus X$.

Let $\mathcal{I}$ and $\mathcal{J}$ be interpretations of $\mathcal{K}$. A *homomorphism* from $\mathcal{I}$ to $\mathcal{J}$, 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, $(h(d), h(d')) \in r^{\mathcal{J}}$ whenever $(d, d') \in r^{\mathcal{I}}$, $r \in \mathsf{Rol}(\mathcal{K})$, $h(d) \in A^{\mathcal{J}}$ whenever $d \in A^{\mathcal{I}}$, $A \in \mathsf{CN}(\mathcal{K})$, and $h(a) = a$ for all $a \in \mathsf{Ind}(\mathcal{K})$. Note that $\mathcal{I} \subseteq \mathcal{J}$ iff the identity mapping id is a homomorphism $\mathrm{id} : \mathcal{I} \to \mathcal{J}$.

Let $\mathsf{N_V}$ be a countably infinite set of *variables*. An *atom* is an expression of the form $A(x)$ or $r(x, y)$ with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $x, y \in \mathsf{N_V}$, referred to as *concept atoms* and *role atoms*, respectively. A *conjunctive query (CQ)* Q is an existentially quantified conjunction q of atoms, $\exists x_1 \cdots \exists x_n\, q$. For simplicity we restrict it to be Boolean; that is, $\mathit{var}(Q) = \{x_1, \dots, x_n\}$. This is without loss of generality since the case of non-Boolean CQs can be reduced to the case of Boolean queries; see e.g. (Rudolph and Glimm 2010).

A *match for* Q *in* $\mathcal{I}$ is a total function $\pi : \mathit{var}(Q) \to \Delta^{\mathcal{I}}$ such that $\mathcal{I}, \pi \models q$ under the standard semantics of first-order logic. An interpretation $\mathcal{I}$ satisfies Q, written as $\mathcal{I} \models Q$ if there exists a match for Q in $\mathcal{I}$. Note that we do not consider queries with constants (i.e., individual names); such queries can be viewed as non-boolean queries with a fixed valuation of free variables, and thus are covered by the reduction to the Boolean case. We do consider *unions of con-*

junctive queries (UCQs), which are disjunctions of CQs. An interpretation $\mathcal{I}$ satisfies a UCQ Q if it satisfies one of its disjuncts. It follows immediately that UCQs are *preserved under homomorphisms*; that is, if $\mathcal{I} \models Q$ and there is a homomorphism from $\mathcal{I}$ to $\mathcal{J}$, then also $\mathcal{J} \models Q$.

A query Q is *entailed by a KB* $\mathcal{K}$, denoted as $\mathcal{K} \models Q$, if every model of $\mathcal{K}$ satisfies Q. A model of $\mathcal{K}$ that does not satisfy Q is called a *counter-model*. The *query entailment problem* asks whether a KB $\mathcal{K}$ entails a (U)CQ Q. Moreover, this problem is equivalent to that of finding a counter-model. It is well known that the *query answering problem* can be reduced to query entailment.

In this paper, we address the problem of *finite query entailment*, which is a variant of query entailment where only finite interpretations are considered: an interpretation $\mathcal{I}$ is *finite* if $\Delta^{\mathcal{I}}$ is finite, and a query Q is *finitely entailed by* $\mathcal{K}$, denoted as $\mathcal{K} \models_{\mathsf{fin}} Q$, if every finite model of $\mathcal{K}$ satisfies Q.

From tree-shaped to finite counter-models

Let us fix an $\mathcal{ALCOI}$ knowledge base $\mathcal{K}$ and a union of conjunctive queries Q. Because we have nominals in our logic, we can assume without loss of generality that $\mathcal{K}$'s ABox does not contain role assertions.

The construction of a finite counter-model begins from a tree-shaped counter-model. An interpretation $\mathcal{I}$ is *tree-shaped* if the interpretation $\mathcal{I} \setminus \mathsf{Nom}(\mathcal{K})$ is a finite collection of trees of bounded degree, with elements of $\mathsf{Ind}(\mathcal{K}) \setminus \mathsf{Nom}(\mathcal{K})$ occurring only in the roots. It is well known that a tree-shaped counter-model can be obtained from an arbitrary counter-model $\mathcal{M}$ by the standard unravelling procedure. To turn a tree-shaped counter-model into a finite counter-model we use a variant of the *blocking principle*: a systematic policy of reusing elements. For example, rather than adding a fresh r-successor of unary type τ, one could add an r-edge to some previously added element of unary type τ (if there is one). This would give a finite model for $\mathcal{K}$, but not necessarily a counter-model for Q: a query asking for a cycle of length 42 might be unsatisfied in the original model, but the blocking principle introduces many new cycles, possibly one of length 42 among them. This is in fact the key difficulty to overcome: we need a blocking principle that does not introduce cycles shorter than the size of the query.

The first step is to look at sufficiently large neighbourhoods, rather than just unary types.

Definition 1. *For $d \in \Delta^{\mathcal{I}} \setminus \mathsf{Nom}(\mathcal{K})$, the n-neighbourhood $N_n^{\mathcal{I}}(d)$ is the subinterpretation of $\mathcal{I}$ induced by $\mathsf{Nom}(\mathcal{K})$ and all elements $e \in \Delta^{\mathcal{I}} \setminus \mathsf{Nom}(\mathcal{K})$ within distance n from d in $\mathcal{I} \setminus \mathsf{Nom}(\mathcal{K})$, enriched with a fresh concept interpreted as $\{d\}$. For $a \in \mathsf{Nom}(\mathcal{K})$, $N_n^{\mathcal{I}}(a)$ is the subinterpretation induced by $\mathsf{Nom}(\mathcal{K})$, enriched similarly.*

Replacing unary types with large neighbourhoods is not enough, because nearby elements can have arbitrary large isomorphic neighbourhoods: in the integers with the successor relation all n-neighbourhoods are isomorphic. The next step is to enrich the initial counter-model in such a way that overlapping neighbourhoods are not isomorphic, following an idea from (Gogacz and Marcinkowski 2013).

Definition 2. *A* colouring with k colours *of an interpretation $\mathcal{I}$ is an extension $\mathcal{J}$ of $\mathcal{I}$ with $\Delta^{\mathcal{J}} = \Delta^{\mathcal{I}}$, such that $\mathcal{J}$ coincides with $\mathcal{I}$ in every element in the signature of $\mathcal{I}$, and interprets fresh k concept names $B_1, \ldots, B_k$ such that $B_1^{\mathcal{J}}, \ldots, B_k^{\mathcal{J}}$ is a partition of $\Delta^{\mathcal{J}}$. We say that $d \in B_i^{\mathcal{J}}$ has colour B_i. A colouring $\mathcal{J}$ of $\mathcal{I}$ is* n-proper *if for each $d \in \Delta^{\mathcal{J}}$ all elements of $N_n^{\mathcal{J}}(d)$ have different colours.*

Because $\mathsf{Nom}(\mathcal{K})$ is contained in each neighbourhood, in n-proper colourings each nominal has a unique colour.

Lemma 1. *If $\mathcal{I} \setminus \mathsf{Nom}(\mathcal{K})$ has bounded degree, then for all $n \geq 0$ there exists an n-proper colouring of $\mathcal{I}$ with finitely many colours.*

We write $\mathcal{I}_n$ for an arbitrarily chosen n-proper colouring of $\mathcal{I}$. Because the neighbourhoods have bounded size and we used only finitely many colours, there are only finitely many n-neighbourhoods in $\mathcal{I}_n$ up to isomorphism. The blocking principle described below relies on this.

Let $\mathcal{I}$ be a tree-shaped counter-model for Q. We turn it into a finite counter-model for Q as follows. Because $\mathcal{I} \setminus \mathsf{Nom}(\mathcal{K})$ has bounded degree, we can consider an n-proper colouring $\mathcal{I}_n$ of $\mathcal{I}$. For each branch π in $\mathcal{I}_n \setminus \mathsf{Nom}(\mathcal{K})$, let d_π be the first node on π such that some earlier node e_π on π satisfies $N_n^{\mathcal{I}_n}(d_\pi) \simeq N_n^{\mathcal{I}_n}(e_\pi)$. The new interpretation $\mathcal{F}_n$ is obtained as follows. $\mathcal{F}_n \setminus \mathsf{Nom}(\mathcal{K})$ includes the branch π up to the predecessor of node d_π and the edge originally leading to d_π is redirected to e_π. Because the degree in $\mathcal{I}_n \setminus \mathsf{Nom}(\mathcal{K})$ is bounded, the domain of $\mathcal{F}_n \setminus \mathsf{Nom}(\mathcal{K})$ is a finite subset of the domain of $\mathcal{I}_n \setminus \mathsf{Nom}(\mathcal{K})$. The whole interpretation $\mathcal{F}_n$ is obtained by including $\mathsf{Nom}(\mathcal{K})$ into the domain and copying from $\mathcal{I}_n$ all edges connecting elements of $\mathsf{Nom}(\mathcal{K})$ with each other and with the elements of $\mathcal{F}_n \setminus \mathsf{Nom}(\mathcal{K})$.

Because we started from a model of $\mathcal{K}$, for all $n \geq 0$,

$$\mathcal{F}_n \models \mathcal{K}\,.$$

We claim that for sufficiently large n, $\mathcal{F}_n$ is a counter-model for Q. In order to prove this, we introduce yet another interpretation, containing $\mathcal{I}_n$ and $\mathcal{F}_n$ as subinterpretations.

Definition 3. *Let $i \leq n$ and let d, e be elements of $\mathcal{I}_n$. We say that (d, e) is an i-link along role r if either d has an r-successor e' in $\mathcal{I}_n$ such that $N_i^{\mathcal{I}_n}(e') \simeq N_i^{\mathcal{I}_n}(e)$, or e has an r-predecessor d' in $\mathcal{I}_n$ such that $N_i^{\mathcal{I}_n}(d') \simeq N_i^{\mathcal{I}_n}(d)$.*

Notice that for $i < j$, each j-link is also an i-link. Note also that (d, e) is an i-link along role r if and only if (e, d) is an i-link along r^-.

Definition 4. *For $i \leq n$, let $\mathcal{I}_n^i$ be the interpretation obtained from $\mathcal{I}_n$ by including into the interpretation of each role r all i-links along r; that is, for every role r and every i-link (d, e) along r, $(d, e) \in r^{\mathcal{I}_n^i}$.*

Clearly, we have

$$\mathcal{I}_n \subseteq \mathcal{I}_n^n \subseteq \mathcal{I}_n^{n-1} \subseteq \cdots \subseteq \mathcal{I}_n^1 \subseteq \mathcal{I}_n^0\,,$$

but the domains of all these interpretations coincide. We keep referring to the edges present in $\mathcal{I}_n^i$ but not in $\mathcal{I}_n$ as i-links, even though they are ordinary edges now.

Theorem 1. *Let P be a CQ with at most k binary atoms and let $n \geq k^2$. For each homomorphism $h : P \to \mathcal{I}_n^n$ there exists a homomorphism $h' : P \to \mathcal{I}_n$ such that*

$$N^{\mathcal{I}_n}_{n-k^2}(h(x)) \simeq N^{\mathcal{I}_n}_{n-k^2}(h'(x))$$

for all $x \in var(P)$.

Theorem 1 holds for any interpretation $\mathcal{I}$ of any $\mathcal{SOIF}$ KB.

Before proving Theorem 1, let us see that it implies that $\mathcal{F}_{k^2} \not\models Q$, where k is a common upper bound on the number of binary atoms in the CQs constituting Q. Because $\mathcal{F}_{k^2}$ is obtained from $\mathcal{I}_{k^2}$ by adding some k^2-links and restricting the domain, it follows that $\mathcal{F}_{k^2} \subseteq \mathcal{I}_{k^2}^{k^2}$. Consequently, if there were a homomorphism $h : P \to \mathcal{F}_{k^2} \subseteq \mathcal{I}_{k^2}^{k^2}$ for some CQ P constituting Q, Theorem 1 would yield a homomorphism $h' : P \to \mathcal{I}_{k^2}$, contradicting $\mathcal{I} \not\models Q$. Thus, we have proved finite controllability for $\mathcal{ALCOI}$.

Corollary 1. *For each $\mathcal{ALCOI}$ KB $\mathcal{K}$ and UCQ Q,*

$$\mathcal{K} \models Q \text{ iff } \mathcal{K} \models_{\mathsf{fin}} Q\,.$$

Proof of Theorem 1. Let $h(P)$ denote the subinterpretation of $\mathcal{I}_n^n$ obtained by restricting the domain to $h(var(P))$, and only keeping in each role r edges $(h(x), h(y))$ such that $r(x, y)$ is an atom from P. We say that h *uses* an r-edge of $\mathcal{I}_n^n$ if this r-edge is present in $h(P)$.

Let ℓ be the number of links in $\mathcal{I}_n^n$ used by P. Then $\ell \leq k$, because P contains at most k binary atoms. The theorem follows by applying the following claim ℓ times: For each homomorphism $h : P \to \mathcal{I}_n^i$ with $k \leq i \leq n$ that uses at least one link, there exists a homomorphism $h' : P \to \mathcal{I}_n^{i-k}$ that uses strictly fewer links and satisfies

$$N^{\mathcal{I}_n}_{i-k}(h(x)) \simeq N^{\mathcal{I}_n}_{i-k}(h'(x))$$

for all $x \in var(P)$. Let us prove the claim.

Let (d, e) be a link used by h: an s-edge in $h(P) \subseteq \mathcal{I}_n^i$ that is not an s-edge in $\mathcal{I}_n$. Then (d, e) is an i-link in $\mathcal{I}_n$. By symmetry it suffices to consider the case when d has an s-successor e' in $\mathcal{I}_n$ such that $N_i^{\mathcal{I}_n}(e) \simeq N_i^{\mathcal{I}_n}(e')$. Let

$$g : N_i^{\mathcal{I}_n}(e) \to N_i^{\mathcal{I}_n}(e')$$

be the witnessing isomorphism. Because g is identity over $\mathsf{Nom}(\mathcal{K}) \subseteq \mathsf{Ind}(\mathcal{K})$, we have $e \notin \mathsf{Nom}(\mathcal{K})$; indeed, otherwise $e' = g(e) = e$ and (d, e) would be an s-edge in $\mathcal{I}_n$. Let E be the connected component of e in

$$h(P) \cap (\mathcal{I}_n \setminus \mathsf{Nom}(\mathcal{K}))\,,$$

where by $\mathcal{J}' \cap \mathcal{J}''$ we mean the interpretation $\mathcal{J}$ such that $\Delta^{\mathcal{J}} = \Delta^{\mathcal{J}'} \cap \Delta^{\mathcal{J}''}$, $A^{\mathcal{J}} = A^{\mathcal{J}'} \cap A^{\mathcal{J}''}$ for all concept names A, and $r^{\mathcal{J}} = r^{\mathcal{J}'} \cap r^{\mathcal{J}''}$ for all role names r. Because $h(P)$ has at most k edges and (d, e) is an s-edge in $h(P)$ but not in E, there are at most $k - 1$ edges in E. We shall bring E close to d in $\mathcal{I}_n$ by pulling it back by the i-link (d, e).

As E is a connected subinterpretation of $\mathcal{I}_n \setminus \mathsf{Nom}(\mathcal{K})$ and has at most $k - 1$ edges, each element of E lies within distance $k - 1$ from e. In particular, $E \subseteq N_i^{\mathcal{I}_n}(e)$. Hence, E is contained in the domain of g and we can define

$$h' : P \to \mathcal{I}_n^{i-k}$$

as follows. For each $x \in var(P)$, let $h'(x) = g(h(x))$ if $h(x) \in E$, and $h'(x) = h(x)$ otherwise. The additional claim of the theorem follows immediately because g preserves $(i - k)$-neighbourhoods of elements within distance k from e. We only need to verify that h' is indeed a homomorphism and that it uses fewer links than h.

Let $r(x, y)$ be an atom of the query P. There are three cases to consider. First, suppose that $h(x), h(y) \notin E$. Then

$$(h'(x), h'(y)) = (h(x), h(y))\,.$$

We have that $(h(x), h(y))$ is an r-edge in $\mathcal{I}_n^{i-k}$ because h is a homomorphism into $\mathcal{I}_n^i \subseteq \mathcal{I}_n^{i-k}$. Obviously, h' uses no new links for such atoms.

Next, suppose that $h(x), h(y) \in E$. Then

$$(h'(x), h'(y)) = (g(h(x), g(h(y)))\,.$$

Moreover, $(h(x), h(y))$ is an r-edge in $\mathcal{I}_n^i$ because h is a homomorphism. Suppose it is a link along r. Then, $h(x)$ has an r-successor in $\mathcal{I}_n$ with the same colour as $h(y)$, or $h(y)$ has an r-predecessor in $\mathcal{I}_n$ with the same colour as $h(x)$. Because both $h(x)$ and $h(y)$ lie within distance $k-1$ from e, this successor or predecessor belongs to $N_i^{\mathcal{I}_n}(e)$, along with $h(x)$ and $h(y)$. But this is impossible because all elements of $N_i^{\mathcal{I}_n}(e)$ have different colours. Hence, $(h(x), h(y))$ is an r-edge in $N_i^{\mathcal{I}_n}(e)$ and $(g(h(x)), g(h(y)))$ is an r-edge in $N_i^{\mathcal{I}_n}(e')$. That is, $(g(h(x)), g(h(y)))$ is an r-edge in $\mathcal{I}_n^{i-k}$, and is not a link along r.

Finally, suppose that $h(x) \notin E$ and $h(y) \in E$ (the symmetric case is analogous). Because h is a homomorphism, $(h(x), h(y))$ is an r-edge in $\mathcal{I}_n^i$. Now there are two subcases. Assume first that $(h(x), h(y))$ is also an r-edge in $\mathcal{I}_n$. By the definition of E it is not an r-edge in $\mathcal{I}_n \setminus \mathsf{Nom}(\mathcal{K})$, so it must be an r-edge between a nominal and an element of E. As such, it is also an r-edge in $N_i^{\mathcal{I}_n}(e)$. Consequently,

$$(h'(x), h'(y)) = (h(x), g(h(y))) = (g(h(x)), g(h(y)))$$

is an r-edge in $N_i^{\mathcal{I}_n}(e')$ and we conclude like previously.

Assume now that $(h(x), h(y))$ is an i-link along r. We need to check that $(h(x), g(h(y)))$ is an r-edge in $\mathcal{I}_n^{i-k}$. Since $h(y)$ and $g(h(y))$ are in distance at most $k - 1$ from e and e', respectively, and $N_i^{\mathcal{I}_n}(e) \simeq N_i^{\mathcal{I}_n}(e')$, it follows that

$$N^{\mathcal{I}_n}_{i-k}(h(y)) \simeq N^{\mathcal{I}_n}_{i-k}(g(h(y)))\,.$$

Because $(h(x), h(y))$ is an i-link, it is also an $(i - k)$-link. If $h(x)$ has an r-successor f in $\mathcal{I}_n$ such that

$$N^{\mathcal{I}_n}_{i-k}(f) \simeq N^{\mathcal{I}_n}_{i-k}(h(y)) \simeq N^{\mathcal{I}_n}_{i-k}(g(h(y)))\,,$$

then $(h(x), g(h(y)))$ is an $(i - k)$-link along r, unless the successor f is $g(h(y))$ itself; in either case $(h(x), g(h(y)))$ is an r-edge in $\mathcal{I}_n^{i-k}$. The remaining possibility is that $h(y)$ has an r-predecessor f in $\mathcal{I}_n$ such that

$$N^{\mathcal{I}_n}_{i-k}(f) \simeq N^{\mathcal{I}_n}_{i-k}(h(x))\,.$$

Because $h(y)$ lies within distance $k - 1$ from e,

$$N^{\mathcal{I}_n}_{i-k}(f) \subseteq N_i^{\mathcal{I}_n}(e)\,.$$

Hence, $g(f)$ is an r-predecessor of $g(h(y))$ such that

$$N_{i-k}^{\mathcal{I}_n}(g(f)) \simeq N_{i-k}^{\mathcal{I}_n}(h(x)) .$$

Consequently, $(h(x), g(h(y)))$ is an $(i-k)$-link along r, unless $g(f)$ is $h(x)$ itself; in either case $(h(x), g(h(y)))$ is an r-edge in $\mathcal{I}_n^{i-k}$.

Thus h' is a homomorphism and uses links only for the atoms of P for which h uses links. To see that h' uses strictly fewer links than h, recall that instead of the i-link (d, e) along s, it uses the s-edge (d, e'), which is not a link. □

$\mathcal{SOI}$ and $\mathcal{SOF}$

The goal of this section is to prove the following theorem.

Theorem 2. *The finite query entailment problem for both $\mathcal{SOI}$ and $\mathcal{SOF}$ is* 2EXPTIME*-complete.*

The lower bounds follow immediately from the results on unrestricted query entailment for $\mathcal{ALCO}$ (Ngo, Ortiz, and Simkus 2016) and $\mathcal{ALCI}$ (Lutz 2008), and Corollary 1; the challenge is to prove the upper bounds. We develop our argument with $\mathcal{SOI}$ in mind, but it adapts easily to $\mathcal{SOF}$ (see the full version).

Let us fix a $\mathcal{SOI}$ knowledge base $\mathcal{K}$ and a union of conjunctive queries Q. Like for $\mathcal{ALCOI}$, we can assume that $\mathcal{K}$'s ABox contains no role assertions.

Because $\mathcal{K}$ is normalised, complete information about restrictions on the types of neighbours of a node is encoded in its unary type. Now, we would like the unary type to determine also the neighbouring nominals. This can be assumed without loss of generality, because one can always extend $\mathcal{K}$ by adding for each $a \in \mathsf{Nom}(\mathcal{K})$ and $r \in \mathsf{Rol}(\mathcal{K})$ fresh concept names $A_{r,a}$, $A_{r^-,a}$ axiomatised with $A_{r,a} \equiv \exists r.\{a\}$, $\{a\} \equiv \forall r.A_{r^-,a}$, and normalise the resulting KB.

Let $\mathcal{I}^*$ be the interpretation obtained from interpretation $\mathcal{I}$ by closing transitively the interpretation of each transitive role. Note that each existential restriction satisfied in $\mathcal{I}$ is also satisfied in $\mathcal{I}^*$. The same holds for quantifier-free CI, and for universal restrictions involving non-transitive roles. For universal restrictions involving transitive roles, we ensure this property by adding a fresh concept name B' for each $B \in \mathsf{CN}(\mathcal{K})$ and CIs $A \sqsubseteq \forall r.B'$, $B' \sqsubseteq \forall r.B'$, $B' \sqsubseteq B$ for each CI of the form $A \sqsubseteq \forall r.B$ with r transitive.

The last assumption we would like to make about $\mathcal{K}$ is that the unary type of each element of $\mathsf{Nom}(\mathcal{K})$ is fully specified in the ABox; that is, for all $a \in \mathsf{Nom}(\mathcal{K})$ and $A \in \mathsf{CN}(\mathcal{K})$, the ABox contains either $A(a)$ or $\bar{A}(a)$. This can be done without loss of generality, because $\mathcal{K} \models_{\mathsf{fin}} Q$ iff $\mathcal{K}' \models_{\mathsf{fin}} Q$ for each $\mathcal{K}'$ that can be obtained from $\mathcal{K}$ by completing assertions about nominals. This adds the factor $2^{|\mathsf{Nom}(\mathcal{K})|\cdot|\mathsf{CN}(\mathcal{K})|}$ to the running time of the decision procedure, but the overall complexity bound is not affected, because it is exponential in the size of $\mathcal{K}$ anyway.

Building on the results of the previous section, we show that the existence of a finite counter-model for Q is equivalent to the existence of a possibly infinite counter-model of a special form, which generalises tree-shaped models. The special form is based on the notion of clique-forests.

Definition 5. *A clique-forest for an interpretation $\mathcal{I}$ of $\mathcal{K}$ is a forest (a sequence of trees) whose each node v is labelled with a subinterpretation $\mathcal{I}_v$ of $\mathcal{I} \setminus \mathsf{Nom}(\mathcal{K})$ such that*

- *the sets $\Delta^{\mathcal{I}_v}$ are a partition of $\Delta^{\mathcal{I}\setminus\mathsf{Nom}(\mathcal{K})}$;*
- *each $\mathcal{I}_v$ is either a single element with all roles empty (element node) or a clique over some transitive role with all other roles empty and no repetitions of unary types (clique node);*
- *apart from edges within cliques, in $\mathcal{I} \setminus \mathsf{Nom}(\mathcal{K})$ there is exactly one edge between $\Delta^{\mathcal{I}_u}$ and $\Delta^{\mathcal{I}_v}$ for every two adjacent nodes u and v: assuming u is the parent of v, it is an r-edge from an element of $\Delta^{\mathcal{I}_u}$ to an element of $\Delta^{\mathcal{I}_v}$ for some $r \in \mathsf{Rol}(\mathcal{K})$.*

Definition 6. *An interpretation $\mathcal{I}$ of $\mathcal{K}$ is a $\mathcal{SOI}$-forest if it admits a clique-forest that consists of at most $|\mathcal{K}|^2$ trees of branching at most $|\mathcal{K}|^2$, such that each element of $\mathsf{Ind}(\mathcal{K}) \setminus \mathsf{Nom}(\mathcal{K})$ occurs in some root.*

Let $\mathcal{K}^*$ denote the KB obtained from $\mathcal{K}$ by dropping transitivity declarations.

Definition 7. *A* counter-example *for Q is a $\mathcal{SOI}$-forest $\mathcal{I}$ such that $\mathcal{I} \models \mathcal{K}^*$ and $\mathcal{I}^* \not\models Q$.*

If $\mathcal{I}$ is a counter-example for Q, thanks to the initial preprocessing, $\mathcal{I}^*$ is a counter-model for Q. One could also show that if there is a counter-model for Q, then there is a counter-example for Q. But we are interested in *finite* counter-models and for that we need an additional condition. Recall that a path is simple if it does not revisit elements.

Definition 8. *An interpretation $\mathcal{I}$ is* safe *if it does not contain an infinite simple r-path for any transitive role r.*

The whole argument now splits into two parts: equivalence of the existence of a *finite* counter-model and a *safe* counter-example, and effective regularity of the set of clique-forests of safe counter-examples. Together they show that finite query entailment can be solved by testing emptiness of an appropriate doubly-exponential automaton (with Büchi acceptance condition), which can be done in polynomial time. We begin from the second part, as it is needed to prove the first one.

Theorem 3. *Given a union Q of CQs, each of size at most m, one can compute (in time polynomial in the size of the output) an automaton of size $2^{|Q|\cdot|\mathcal{K}|^{\mathcal{O}(m)}}$ that recognises clique-forests of safe counter-examples for Q.*

The proof of Theorem 3 is a routine automata construction (detailed in the full version). Let us focus on the first part of the argument.

Theorem 4. *Q has a finite counter-model iff Q has a safe counter-example.*

Suppose first that there exists a finite counter-model $\mathcal{M}$ for Q. We build a $\mathcal{SOI}$ forest $\mathcal{I}$ out of it using a version of the standard unravelling. We begin by taking copies of all elements of $\mathsf{Ind}(\mathcal{K})$ with unary types copied accordingly. Then, recursively, for each added element d' and each CI $A \sqsubseteq \exists r.B$ that is not yet satisfied for d' in $\mathcal{I}$ proceed as follows. The element d' is a copy of some d from $\mathcal{M}$ of the same unary type. Therefore there exists an element e in $\mathcal{M}$

witnessing the CI. If $e \in \mathsf{Nom}(\mathcal{K})$, then it is already included in $\mathcal{I}$, and we just add an r edge from d' to e. Assume $e \notin \mathsf{Nom}(\mathcal{K})$. If r is not a transitive role, we just add a copy of e as an r-successor of d'. Assume that r is a transitive role. Let X be the strongly connected component of r that contains e and let X_0 be a minimal set that contains at least one element from each nonempty $C^{\mathcal{M}} \cap \big(X \setminus \mathsf{Nom}(\mathcal{K})\big)$, where C ranges over $\mathsf{CN}(\mathcal{K})$. By minimality, $|X_0| \leq |\mathcal{K}|$. We add to $\mathcal{I}$ an r-clique over a copy of X_0, with an r edge from d' to the copy of some element $f \in B^{\mathcal{M}} \cap X_0$; f exists because $e \in B^{\mathcal{M}} \cap \big(X \setminus \mathsf{Nom}(\mathcal{K})\big)$. Note that no other edges among newly added elements are present: existential restrictions for these nodes will be witnessed in the following steps of the construction. Let $\mathcal{I}$ be the interpretation obtained in the limit. By construction, $\mathcal{I}$ admits a clique-forest. For each element at most one successor per CI is added. Because each clique node contains up to $|\mathcal{K}|$ elements, the branching of the clique-forest is bounded by $|\mathcal{K}|^2$. The same bound holds for the number of trees in the clique-forest: we begin from $|\mathsf{Ind}(\mathcal{K})|$ nodes, but then the ones corresponding to elements of $\mathsf{Nom}(\mathcal{K})$ are removed and their children become roots. Hence, $\mathcal{I}$ is a $\mathcal{SOI}$ forest. Because we do not unravel cliques in transitive roles, it is safe.

Lemma 2. *$\mathcal{I}$ is a safe counter-example for Q.*

Assume now that there exists a safe counter-example $\mathcal{I}$ for Q. By Theorem 3, the set of clique-forests of safe counter-examples for Q can be recognised by an automaton. It is well known that the automaton then accepts a regular forest, which has only finitely many non-isomorphic subtrees. Hence, without loss of generality we can assume that the clique-forest of $\mathcal{I}$ has p non-isomorphic subtrees for some p. Using the methodology from the previous section we shall turn $\mathcal{I}$ into a finite counter-model for Q. The main obstacle is that Q uses transitive roles, which are not fully represented in $\mathcal{I}$. Our solution is to replace Q with a different query that can be evaluated directly over $\mathcal{I}$. This is done by exploiting a bound on the length of simple r-paths for transitive roles r, guaranteed by the regularity of the clique-forest of $\mathcal{I}$.

Definition 9. *An interpretation is* ℓ-bounded *if for each transitive role r, each simple r-path has length at most ℓ.*

Lemma 3. *$\mathcal{I} \setminus \mathsf{Nom}(\mathcal{K})$ is ℓ-bounded for $\ell = 2p \cdot |\mathcal{K}|$.*

Proof. Let r be a transitive role in $\mathcal{K}$. Each r-path going down the clique-forest of $\mathcal{I}$ contains at most p nodes. Indeed, if there were a longer r-path, then a subtree would occur twice on that path, which immediately leads to an infinite simple r-path in $\mathcal{I} \setminus \mathsf{Nom}(\mathcal{K})$, contradicting the safety of $\mathcal{I}$. Each simple path in the clique-forest can be split into an r-path going up and an r-path going down. Each of them has at most p nodes. Because each node contains at most $|\mathcal{K}|$ elements, it follows that each simple r-path in $\mathcal{I} \setminus \mathsf{Nom}(\mathcal{K})$ has length at most $2p \cdot |\mathcal{K}|$. □

Lemma 4. *For each $\mathcal{J}$, if $\mathcal{J} \setminus \mathsf{Nom}(\mathcal{K})$ is ℓ-bounded, then $\mathcal{J}$ is ℓ^*-bounded for $\ell^* = (\ell + 2) \cdot (|\mathsf{Nom}(\mathcal{K})| + 1)$.*

Let Q^* be obtained from Q by replacing each transitive atom $s(x, y)$ by the disjunction

$$\bigvee_{i \leq \ell^*} s^i(x, y)\,,$$

where $s^i(x, y)$ is the conjunctive query expressing the i-fold composition of s. Assuming that each disjunct of Q contains at most k binary atoms, Q^* can be rewritten as a union of conjunctive queries, each using at most $k \cdot \ell^*$ binary atoms.

Lemma 5. *For all ℓ^*-bounded $\mathcal{J}$, $\mathcal{J}^* \models Q$ iff $\mathcal{J} \models Q^*$.*

By Lemmas 3–5, we conclude that $\mathcal{I} \not\models Q^*$. Now we can use the blocking principle. Because clique nodes have at most $|\mathcal{K}|$ elements and each node has at most $|\mathcal{K}|^2$ children, $\mathcal{I} \setminus \mathsf{Nom}(\mathcal{K})$ has bounded degree and we can consider the n-properly coloured $\mathcal{I}_n$, for any n. On each branch π in $\mathcal{I}_n \setminus \mathsf{Nom}(\mathcal{K})$, let D_π be the first node for which some earlier node E_π satisfies $N_n^{\mathcal{I}_n}(d_\pi) \simeq N_n^{\mathcal{I}_n}(e_\pi)$, where $d_\pi \in D_\pi$ and $e_\pi \in E_\pi$ are the endpoints of the edges connecting D_π and E_π to their parent nodes. The new interpretation $\mathcal{F}_n$ is obtained as usual: we include the branch π up to the predecessor of node D_π and the edge originally leading to d_π is redirected to e_π; edges connecting the elements of $\mathsf{Nom}(\mathcal{K})$ with each other and with the elements of the included parts of the branches are copied from $\mathcal{I}_n$.

Because we started from $\mathcal{I} \models \mathcal{K}^*$, it is routine to check that $\mathcal{F}_n \models \mathcal{K}^*$ for all n. By the initial preprocessing, $(\mathcal{F}_n)^* \models \mathcal{K}$. Let us fix

$$n = \max((k \cdot \ell^*)^2, (\ell + 1)^2 + \ell)\,.$$

By Theorem 1, $\mathcal{F}_n \not\models Q^*$. We conclude $(\mathcal{F}_n)^* \not\models Q$ using Lemmas 4–5 and Theorem 5 below.

Definition 10. *A link (d, e) in $\mathcal{I}$ along r is* external *if either no r-path from the witnessing e' to d is disjoint from $\mathsf{Nom}(\mathcal{K})$ or dually no r-path from e to the witnessing d' is disjoint from $\mathsf{Nom}(\mathcal{K})$.*

By construction, all links in $\mathcal{I}_n$ along transitive roles included into $\mathcal{F}_n$ are external.

Theorem 5. *Assume that $\mathcal{I} \setminus \mathsf{Nom}(\mathcal{K})$ has bounded degree and is ℓ-bounded. Let $n > (\ell + 1)^2 + \ell$ and let $\mathcal{J}$ be a subinterpretation of $\mathcal{I}_n^n$ in which all links along transitive roles are external. Then, $\mathcal{J} \setminus \mathsf{Nom}(\mathcal{K})$ is also ℓ-bounded.*

Proof. Suppose there is a simple s-path π in $\mathcal{J} \setminus \mathsf{Nom}(\mathcal{K})$ of length $\ell + 1$, for some transitive role s. We can view π as a conjunctive query with $\ell + 1$ s-atoms. By applying Theorem 1 to π we lift the inclusion homomorphism $\pi \subseteq \mathcal{J} \subseteq \mathcal{I}_n^n$ to a homomorphism $h : \pi \to \mathcal{I}_n$, that preserves ℓ-neighbourhoods. Because π is disjoint from $\mathsf{Nom}(\mathcal{K})$, so is its image. Hence, we can view h as a homomorphism

$$h : \pi \to \mathcal{I}_n \setminus \mathsf{Nom}(\mathcal{K})\,.$$

Because $\mathcal{I}_n \setminus \mathsf{Nom}(\mathcal{K})$ is ℓ-bounded, it suffices to show that h is injective to obtain a contradiction.

Observe first that h is injective over segments of π that do not contain links. Indeed, because $\mathcal{I}_n$ is n-properly coloured and $n \geq |\pi|$, in each such segment all elements have different colours. Hence, it suffices to show that the images of the segments are disjoint. Suppose the images of some two different segments overlap on an element from a strongly connected component X of s in $\mathcal{I}_n \setminus \mathsf{Nom}(\mathcal{K})$. Hence, all segments between these two are entirely mapped to X. In particular, there exists an n-link (d,e) along s such that $h(d) \in X$ and $h(e) \in X$. We claim this is impossible.

By symmetry we can assume that d has an s-successor e' such that no s-path from e' to d is disjoint from $\mathsf{Nom}(\mathcal{K})$ and $N_n^{\mathcal{I}_n}(e') \simeq N_n^{\mathcal{I}_n}(e)$. In particular, e' and e have the same colour. Because $n > 1$, we have $e' \in N_n^{\mathcal{I}_n}(d)$. We obtain a contradiction by finding another element in $N_n^{\mathcal{I}_n}(d)$ of the same colour as e.

Let D be the strongly connected component of s in $\mathcal{I}_n \setminus \mathsf{Nom}(\mathcal{K})$ that contains d. Because $\mathcal{I}_n \setminus \mathsf{Nom}(\mathcal{K})$ is ℓ-bounded, all elements of D are within distance $\ell < n$ from d. Consequently, D is isomorphic to X, because h preserves ℓ-neighbourhoods. Hence, there exists an element $e'' \in D \subseteq N_n^{\mathcal{I}_n}(d)$ of the same colour as e. Because $e' \notin D$, we have $e' \neq e''$, as required for the contradiction. □

$\mathcal{SIF}$

For $\mathcal{ALCIF}$, a tight upper bound on the complexity of finite query entailment can be obtained by revisiting some known and implicitly proven results on the guarded fragment with two variables and counting (Pratt-Hartmann 2009; 2007). We consider a slightly more general problem of *finite entailment modulo types*, which will be useful later. For a KB $\mathcal{K}$, a query Q, and a set of unary types $T \subseteq \mathsf{Tp}(\mathcal{K})$ we write $\mathcal{K} \models_{\mathsf{fin}}^{T} Q$ if for each interpretation $\mathcal{I}$ that only realises types from T, if $\mathcal{I} \models \mathcal{K}$ then $\mathcal{I} \models Q$. This problem reduces to finite query entailment by including into Q one CQ for each type not listed in T, but this makes Q exponential in the size of $\mathsf{CN}(\mathcal{K})$ and leads to a worse complexity upper bound.

Theorem 6. *Given an $\mathcal{ALCIF}$ KB $\mathcal{K}$, a union Q of CQs, each of size at most m, and a set $T \subseteq \mathsf{Tp}(\mathcal{K})$, one can decide whether $\mathcal{K} \models_{\mathsf{fin}}^{T} Q$ in time $2^{\mathcal{O}(|\mathcal{K}|+|Q|\cdot m^m)}$.*

Corollary 2. *The finite query entailment problem for $\mathcal{ALCIF}$ is* 2EXPTIME*-complete.*

Relying on Theorem 6 and our previous results for $\mathcal{SOI}$, we extend the upper bound of Corollary 2 to $\mathcal{SIF}$.

Let us fix a UCQ Q and a $\mathcal{SIF}$ KB $\mathcal{K}$. We work again with counter-models of a special shape, this time based on tree partitions. We assume a proviso that the ABox of $\mathcal{K}$ does not contain transitive and non-transitive roles simultaneously; we lift it by the end of the section.

Definition 11. *A* tree partition *of an interpretation $\mathcal{I}$ is a tree T whose each node v is labelled with a finite subinterpretation $\mathcal{I}_v$ of $\mathcal{I}$, called a* bag, *such that $\bigcup_{v\in T} \mathcal{I}_v = \mathcal{I}$ and for each element some bag containing it is the parent of all other bags containing it. The maximal bag size is called the* width *of T.*

Definition 12. *An interpretation $\mathcal{I}$ is a $\mathcal{SIF}$-*tree *if it admits a tree partition such that*

- *the root bag contains* $\mathsf{Ind}(\mathcal{K})$,
- *each bag contains edges in transitive roles only (*TR *bag) or in non-transitive roles only (*NT *bag),*
- *each element is in exactly two bags, one* TR *and one* NT,
- *each two adjacent bags share exactly one element.*

Lemma 6. *There exists a finite counter-model for Q iff there exists a $\mathcal{SIF}$-tree counter-model for Q of finite width.*

Proof. Let $\mathcal{F}$ be a finite counter-model for Q. We turn it into a $\mathcal{SIF}$-tree counter-model $\mathcal{I}$ using a very simple unravelling procedure. For each $\mu \in \{\text{TR}, \text{NT}\}$, let $\mathcal{F}_\mu$ be the interpretation obtained from $\mathcal{F}$ by restricting the set of roles to μ roles. By the proviso, the ABox of $\mathcal{K}$ contains only μ_0 roles for some $\mu_0 \in \{\text{TR}, \text{NT}\}$. We construct the $\mathcal{SIF}$-tree top down. In the root we put $\mathcal{F}_{\mu_0}$ itself. Then, iteratively, for each element d that belongs only to a μ bag we add a child bag obtained by taking an isomorphic copy of $\mathcal{F}_\nu$ for $\nu \neq \mu$, in which all elements except d are replaced with their fresh copies; in particular, each individual different from d is replaced with an ordinary anonymous element of the same unary type. It is routine to verify that the resulting interpretation $\mathcal{I}$ is a model of $\mathcal{K}$. The natural homomorphism from $\mathcal{I}$ to $\mathcal{F}$ ensures that $\mathcal{I} \not\models Q$. The width of $\mathcal{I}$ is $|\mathcal{F}|$.

Let us now take a $\mathcal{SIF}$-tree $\mathcal{I}$ of width ℓ that is a counter-model for Q. We use the methodology developed for $\mathcal{SOI}$ to turn $\mathcal{I}$ into a finite counter-model. Because $|\mathcal{I}_v| \leq \ell$, $\mathcal{I}$ has degree at most $2 \cdot \ell \cdot |\mathcal{K}|$. Because each r-path for any transitive role r is contained within a single TR bag, it follows that $\mathcal{I}$ is $(\ell-1)$-bounded.

For the purpose of the coloured blocking principle, we need to ensure that each infinite branch of the tree partition of our interpretation contains infinitely many TR bags that consist of a single edge (pointing up or down the tree). We achieve this by performing an additional unravelling of $\mathcal{I}$. We start with a copy of the root bag in the tree partition of $\mathcal{I}$, where elements of $\mathsf{Ind}(\mathcal{K})$ are preserved and other elements are replaced with their fresh copies. Let d' be an element in the interpretation under construction that so far belongs to only one bag X'. By construction, d' is a copy of some element d of $\mathcal{I}$. If X' is a TR bag, add a copy of the NT bag that contains d, with d replaced with d' and other elements replaced with their fresh copies. Assume that X' is an NT bag. For each TR role r and each r-successor e of d, add three new bags. First, add a bag consisting of d', a fresh copy e' of e, and an r-edge from d' to e'. Then, for each $\mu \in \{\text{TR}, \text{NT}\}$, add a copy of the μ-bag containing e, with e replaced with e' and all other elements replaced with their fresh copies (different for each μ).

Let $\mathcal{J}$ be the interpretation obtained in the limit. Because in the tree partition of $\mathcal{I}$ TR bag and NT bags alternate, in the tree partition of $\mathcal{J}$ NT bags have only new single-edge TR bag children, new single-edge TR bags have one NT bag child and one TR bag child, and copies of original TR bags have only NT bag children. Consequently, on each infinite branch, there are infinitely many single-edge TR bags.

Interpretations of transitive roles in $\mathcal{J}$ need not be transitive relations, but it is straightforward to check that $\mathcal{J}$ is a model of $\mathcal{K}^*$; in particular, functionality declarations were not affected because the new single-edge bags involve only TR roles (non-functional). Moreover, $\mathcal{J}^* \not\models Q$ because $\mathcal{J}$ maps homomorphically to $\mathcal{I}$ and, consequently, so does $\mathcal{J}^*$. The degree in $\mathcal{J}$ is bounded by $2 \cdot \ell \cdot |\mathcal{K}| + 1$, because each element belongs to one TR bag and one NT bag of size at most ℓ, and possibly one single-edge bag. Finally, $\mathcal{J}$ is 2ℓ-bounded because in the worst case a simple r-path for any transitive role r goes first through a bag with at most ℓ elements, then two single-edge bags, and then another bag with at most ℓ elements.

We can now apply the coloured blocking principle. Suppose each disjunct of Q uses at most k binary atoms. Let $\ell^* = 2\ell$ and let Q^* be obtained from Q by replacing each transitive role atom S by the disjunction

$$\bigvee_{i \leq \ell^*} S^i(x, y)\,,$$

and rewriting the resulting query as a UCQ. Each CQ in Q^* has at most $k \cdot \ell^*$ binary atoms. Because $\mathcal{J}$ has bounded degree, we can consider its n-proper colouring $\mathcal{J}_n$ for any n. On each branch π of the tree partition of $\mathcal{J}_n$, let D_π be the first single-edge TR bag for which some earlier single-edge TR bag E_π satisfies $N_n^{\mathcal{J}_n}(d_\pi) \simeq N_n^{\mathcal{J}_n}(e_\pi)$, where $d_\pi \in D_\pi$ and $e_\pi \in E_\pi$ are the elements that D_π and E_π share with their respective parents. The new structure $\mathcal{F}_n$ is obtained like before: we include the branch π up to the predecessor of node D_π and the edge in D_π is redirected to the successor of e_π in E_π. Because $\mathcal{J}$ is a model of $\mathcal{K}^*$ and we only redirected edges in non-functional roles, it follows that $\mathcal{F}_n$ is a model of $\mathcal{K}^*$. Consequently, $\mathcal{F}_n^* \models \mathcal{K}$. Let us now fix

$$n = \max((k \cdot \ell^*)^2, (\ell^* + 1)^2 + \ell^*)\,.$$

By Theorem 1, we get $\mathcal{F}_n \not\models \mathcal{Q}^*$. Because $\mathcal{J}$ is ℓ^*-bounded and we clearly used only external links in the construction of $\mathcal{F}_n$, by Lemma 5 and Theorem 5 we obtain $\mathcal{F}_n^* \not\models \mathcal{Q}$. $\square$

Thus, it suffices to consider counter-models that are $\mathcal{SIF}$-trees of finite width, but there is a priori no bound on the width, which hinders direct application of automata. Instead, we show that one can test existence of $\mathcal{SIF}$-tree counter-models without manipulating $\mathcal{SIF}$-trees directly.

Our first step is to adjust the structure of Q's disjuncts to the structure of $\mathcal{SIF}$-trees. To keep this as simple as possible, we make a second proviso that each CQ constituting Q is connected. We eliminate it towards the end of the section. Let P be one of the CQs constituting Q. It is convenient to think P as an interpretation with the domain $\mathit{var}(P)$ and interpretations of concepts and roles given by the atoms of P. Whenever P is mapped homomorphically into a $\mathcal{SIF}$-tree $\mathcal{I}$, the image of P is a $\mathcal{SIF}$-tree as well. Indeed, because P is connected, a witnessing tree partition of the image of P is naturally induced by the tree partition of $\mathcal{I}$. Hence, if Q is a union of n CQs of size at most m, over $\mathcal{SIF}$-trees Q is equivalent to

$$Q_1 \cup Q_2 \cup \cdots \cup Q_p\,, \tag{*}$$

where the queries Q_i are all non-isomorphic $\mathcal{SIF}$-trees obtained as homomorphic images of the CQs of Q, each using a fresh set of variables, and $p \leq n \cdot m^m$.

For all $\mu \in \{\text{TR}, \text{NT}\}$ and $x \in \bigcup_i \mathit{var}(Q_i)$, let $Q_{\mu,x}$ be the query obtained by taking all bags that are reachable from the μ bag containing x without visiting the other bag containing x, as illustrated in Figure 1. For all $x \in \mathit{var}(Q_i)$ it holds that $Q_i = Q_{\text{TR},x} \wedge Q_{\text{NT},x}$.

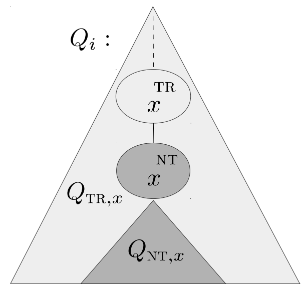

Figure 1: Queries $Q_{\text{TR},x}$ and $Q_{\text{NT},x}$ for $x \in \mathit{var}(Q_i)$.

Let $\mathcal{K}_Q$ be obtained from $\mathcal{K}$ by extending the TBox as follows: for each $\mu \in \{\text{TR}, \text{NT}\}$ and $x \in \bigcup_i \mathit{var}(Q_i)$, we add a fresh concept name $A_{\mu,x}$ and the complementary concept name $\bar{A}_{\mu,x}$, together with the usual axiomatisation. The interpretation of $A_{\mu,x}$ is intended to collect elements d such that $Q_{\mu,x}$ can be matched with x mapped to d.

A *specialisation* $\widetilde{Z}$ of a bag Z of query Q_i is obtained by including for each $x \in \mathit{var}(Z)$ and each $\mu \in \{\text{TR}, \text{NT}\}$ either the atom $A_{\mu,x}(x)$ or the atom $\bar{A}_{\mu,x}(x)$, where $\bar{A}_{\mu,x}$ is the concept name complementary to $A_{\mu,x}$. A specialisation $\widetilde{Z}$ of a μ-bag Z of Q_i is *consistent* if for all x it holds that: $\widetilde{Z}$ contains $A_{\mu,x}(x)$ iff for all $y \in \mathit{var}(\widetilde{Z}) \setminus \{x\}$, $\widetilde{Z}$ contains $A_{\nu,y}(y)$ with $\nu \neq \mu$. An interpretation $\mathcal{I}$ (with the extended set of concept names) is *consistent* if it does not match inconsistent specialisations of bags of queries $Q_1, Q_2, \ldots, Q_p$.

For a $\mathcal{SIF}$ KB $\mathcal{L}$ and $\mu \in \{\text{TR}, \text{NT}\}$ we write $\mathcal{L} \upharpoonright \mu$ for the KB obtained by dropping all ABox assertions, CIs, and declarations that involve ν-roles for $\nu \neq \mu$.

Definition 13. *$T \subseteq \mathsf{Tp}(\mathcal{K}_Q)$ is a* counter-witness *for Q if*

- *for all $x \in \bigcup_i \mathit{var}(Q_i)$, each $\tau \in T$ contains $\bar{A}_{\text{TR},x}$ or $\bar{A}_{\text{NT},x}$;*
- *assuming $\mathcal{K}$ uses only μ_0-roles in the ABox, there exists a consistent finite model of $\mathcal{K}_Q \upharpoonright \mu_0$ that realises only types from T; and*
- *for all $\tau \in T$ and $\mu \in \{\text{TR}, \text{NT}\}$ there exists a consistent finite model of the TBox of $\mathcal{K}_Q \upharpoonright \mu$ that realises type τ and realises only types from T.*

Lemma 7. *Q admits a $\mathcal{SIF}$ tree counter-model of finite width iff there exists a counter-witness for Q.*

Proof. Let $\mathcal{I}$ be a $\mathcal{SIF}$-tree counter-model for Q; we do not need to assume that $\mathcal{I}$ has finite width. Let $\mathcal{I}_Q$ be obtained by extending $\mathcal{I}$ with the unique interpretation of the concept names $A_{\mu,x}$ and $\bar{A}_{\mu,x}$ faithful to their intended meaning: if $Q_{\mu,x}$ can be matched in $\mathcal{I}$ with x mapped to d, then $d \in (A_{\mu,x})^{\mathcal{I}_Q}$, and otherwise $d \in \left(\bar{A}_{\mu,x}\right)^{\mathcal{I}_Q}$. By construction, $\mathcal{I}_Q$ is consistent, and so is each of its bags. Let T be the set of types realised in $\mathcal{I}_Q$. Because $\mathcal{I} \not\models Q$, no type from T contains both $A_{\text{TR},x}$ and $A_{\text{NT},x}$, which gives the first condition in Definition 13. The root bag of $\mathcal{I}_Q$ witnesses the second condition. As each element of $\mathcal{I}_Q$ belongs to a TR bag and a NT bag, each $\tau \in T$ is realised in some TR bag and in some NT bag. These bags witness the third condition.

Conversely, let $T \subseteq \mathsf{Tp}(\mathcal{K}_Q)$ be a counter-witness for Q. Let $\mathcal{I}_0$ be the interpretation guaranteed by the second condition, and let $\mathcal{I}_{\mu,\tau}$ be interpretations guaranteed by the third condition. From them we build a $\mathcal{SIF}$-tree counter-model for Q in a top-down fashion. The root bag is $\mathcal{I}_0$. Take an element d that so far only belongs to a μ-bag. By construction, the type τ of d belongs to T. Let $\nu \neq \mu$. We add to the $\mathcal{SIF}$-tree under construction a copy of $\mathcal{I}_{\mu,\tau}$, with one element of type τ replaced by d. Because $\mathcal{K}$ is normalised, the resulting $\mathcal{SIF}$-tree $\mathcal{I}$ is a model of $\mathcal{K}$. The tree partition of $\mathcal{I}$ has finite width because each bag is a copy of one of the finitely many finite interpretations $\mathcal{I}_0$ and $\mathcal{I}_{\mu,\tau}$.

It remains to see that $\mathcal{I} \not\models Q$. We first prove by induction on the size of $Q_{\mu,x}$ that for each homomorphism $f : Q_{\mu,x} \to \mathcal{I}$, it holds that $f(x) \in A^{\mathcal{I}}_{\mu,x}$. Let Z_x and $Z_{f(x)}$ be the μ-bags of x and $f(x)$, respectively. By the inductive assumption, $f(y) \in A^{\mathcal{I}}_{\nu,y}$ for all $y \in Z_x \setminus \{x\}$ and $\nu \neq \mu$. Because $Z_{f(x)}$ matches only consistent specialisations, there is a consistent specialisation $\widetilde{Z}_x$ of Z_x such that f induces a homomorphism from $\widetilde{Z}_x$ to $Z_{f(x)}$. From the consistency of $\widetilde{Z}_x$ it follows that $f(x) \in A^{\mathcal{I}}_{\mu,x}$. Now, if $\mathcal{I} \models Q$, then there is a homomorphism $f : Q_i \to \mathcal{I}$ for some i. Then, $f(x) \in A^{\mathcal{I}}_{\mathrm{TR},x} \cap A^{\mathcal{I}}_{\mathrm{NT},x}$ for all $x \in \mathit{var}(Q_i)$. Because all types realised in $\mathcal{I}$ occur in T, this contradicts Definition 13. $\square$

Theorem 7. *The finite query entailment problem for $\mathcal{SIF}$ is in* 2EXPTIME.

Proof. Let $\mathcal{K}$ be a $\mathcal{SIF}$ KB using only TR or only NT roles in the ABox and let Q be a union of connected CQs, each of size at most m. By Lemmas 6-7, testing $\mathcal{K} \models_{\mathsf{fin}} Q$ amounts to deciding if there exists a counter-witness for Q, which can be done using the following variant of type elimination (Pratt 1979; Rudolph, Krötzsch, and Hitzler 2012). Let T_0 be the set of types from $\mathsf{Tp}(\mathcal{K}_Q)$ that contain either $\bar{A}_{\mathrm{TR},x}$ or $\bar{A}_{\mathrm{NT},x}$ for all $x \in \bigcup_i \mathit{var}(Q_i)$. For $T \subseteq T_0$, let $F(T)$ be the set of types $\tau \in T_0$ such that for all $\mu \in \{\mathrm{TR}, \mathrm{NT}\}$ there exists a consistent finite model of the TBox of $\mathcal{K}_Q \upharpoonright \mu$ that realises type τ and realises only types from T. Then, a set T is a counter-witness if it is a fixed point of the operator F and satisfies the second condition of Definition 13. Notice that F is a monotone operator on subsets of T_0. Consequently, F has the greatest fixed point and it can be obtained by iterating F on T_0:

$$T_0 \supseteq F(T_0) \supseteq F^2(T_0) \supseteq \cdots \supseteq F^i(T_0) = F^{i+1}(T_0)$$

for some $i \leq |T_0|$. Thus, a counter-witness for Q exists iff $F^i(T_0)$ satisfies the second condition of Definition 13. It remains to see how to test this condition and how to compute $F(T)$ for a given T. Both these tasks reduce to finite query entailment modulo types for simpler logics.

A given T satisfies the second condition of Definition 13 iff it is not the case that $\mathcal{K}_Q \upharpoonright \mu_0 \models^T_{\mathsf{fin}} Q'$, where the UCQ Q' is the union of all inconsistent specialisations of the bags of queries $Q_1, Q_2, \ldots, Q_p$ (*). The size of $\mathcal{K}_Q \upharpoonright \mu_0$ is bounded by the size of $\mathcal{K}_Q$ which is $|\mathcal{K}| + \mathcal{O}(mp)$, and Q' is a union of at most $p \cdot 2^{2m}$ CQs of size $\mathcal{O}(m)$.

If $\mu_0 = \mathrm{NT}$, then $\mathcal{K}_Q \upharpoonright \mu_0$ is an $\mathcal{ALCIF}$ KB. By Theorem 6, we can decide if $\mathcal{K}_Q \upharpoonright \mu_0 \models^T_{\mathsf{fin}} Q'$ in time $2^{\mathcal{O}(|\mathcal{K}_Q \upharpoonright \mu_0| + |Q'| \cdot m^m)}$, which is $2^{\mathcal{O}(|\mathcal{K}| + mp \cdot 2^{\mathrm{poly}(m)})}$.

If $\mu_0 = \mathrm{TR}$, then $\mathcal{K}_Q \upharpoonright \mu_0$ is a $\mathcal{SOI}$ KB (with no nominals used). Using our previous results on $\mathcal{SOI}$, we can decide if $\mathcal{K}_Q \upharpoonright \mu_0 \models_{\mathsf{fin}} Q'$ in time $2^{|Q'| \cdot |\mathcal{K}_Q \upharpoonright \mu_0|^{\mathcal{O}(m)}}$, which is $2^{mp \cdot (|\mathcal{K}| + mp)^{\mathcal{O}(m)}}$. We can easily incorporate the set of types T without increasing the complexity: if the ABox contains some type not in T the algorithm immediately accepts; otherwise, the automaton is constructed like before, except that the set of all types is replaced everywhere with T.

To compute $F(T)$ for a given T we need to test for each $\tau \in T$ and $\mu \in \{\mathrm{TR}, \mathrm{NT}\}$ whether there is a consistent finite model of the TBox of $\mathcal{K}_Q \upharpoonright \mu$ that realises type τ and realises only types from T. For each τ and μ this test can be done just like above, except that in $\mathcal{K}_Q \upharpoonright \mu$ we replace the ABox with $\{A(b) \mid A \in \tau\}$ where b is a fresh individual name. The complexity bounds for a single test carry over. To compute the fixed point we need at most $2^{2mp+|\mathcal{K}|}$ iterations of F, each requiring at most $2^{2mp+|\mathcal{K}|}$ $\mathcal{SOI}$ tests and at most $2^{2mp+|\mathcal{K}|}$ $\mathcal{ALCIF}$ tests. These factors are absorbed by the asymptotic bounds on the cost of single tests. Substituting the bound $p \leq |Q| \cdot m^m$ we obtain the bound $2^{(|\mathcal{K}|+|Q|)^{\mathrm{poly}(m)}}$ for the total running time.

Let us now lift the provisos. Take an arbitrary $\mathcal{SIF}$ KB $\mathcal{K}$ and arbitrary UCQ Q. Like for $\mathcal{SOI}$, we can assume that each individual has its unary type fully specified in the ABox. Consider two KBs $\mathcal{K}_1$ and $\mathcal{K}_2$ obtained from $\mathcal{K}$ by removing from the ABox of $\mathcal{K}$ all transitive and all non-transitive roles, respectively. It is not hard to prove (see the full version) that $\mathcal{K} \not\models_{\mathsf{fin}} Q$ iff there exist finite interpretations $\mathcal{F}_1 \models \mathcal{K}_1$ and $\mathcal{F}_2 \models \mathcal{K}_2$ such that for each disjunct P of Q, for each $V \subseteq \mathit{var}(P)$, for each function $h : V \to \mathsf{Ind}(\mathcal{K})$, for each partition of the atoms of P into P_1 and P_2 with $\mathit{var}(P_1) \cap \mathit{var}(P_2) \subseteq V$, for some i it holds that $\mathcal{F}_i \not\models h(P_i)$, where $h(P_i)$ is a CQ with constants obtained from P_i by applying h to variables in V. For each P, V, h and each partition P_1, P_2 of P, guess whether it is $h(P_1)$ or $h(P_2)$ that will not hold. Let Q_i be the union of all chosen $h(P_i)$; note that this is a union of exponentially many CQs of size bounded by the maximal size of Q's CQs. (The number of possible Q_i is doubly exponential, so eliminating this nondeterminism adds a doubly exponential factor to the running time.) It holds that $\mathcal{K} \not\models_{\mathsf{fin}} Q$ iff $\mathcal{K}_i \not\models_{\mathsf{fin}} Q_i$ for all i, and each $\mathcal{K}_i$ respects the proviso. For the second proviso, consider $R = R_1 \cup \cdots \cup R_p$ with $R_j = R^1_j \wedge \cdots \wedge R^{q_j}_j$, where R^k_j are connected CQs over disjoint sets of variables and constants. Then for any KB $\mathcal{L}$, $\mathcal{L} \not\models_{\mathsf{fin}} R$ iff $\mathcal{L} \not\models_{\mathsf{fin}} R^{k_1}_1 \cup \cdots \cup R^{k_p}_p$ for some $k_1, \ldots, k_p$. The number of sequences $k_1, \ldots, k_p$ to check is singly exponential in p. Applying this construction to $\mathcal{K}_i$ and Q_i, we arrive at the case where both provisos are satisfied. Because Q_i is an exponential union of CQs, this step introduces a doubly exponential factor to the running time, but the size bounds for the involved KBs and UCQs are not affected. After eliminating constants from Q_i in the usual way, we can use the algorithm described above. $\square$

Conclusions and Discussion

We have established decidability of finite query entailment of $\mathcal{SOI}$, $\mathcal{SOF}$ and $\mathcal{SIF}$, and proved that the combined complexity coincides with that of unrestricted query entailment (2EXPTIME-complete in all cases). Decidability of finite query entailment for $\mathcal{SOIF}$ remains open.

Since existing 2EXPTIME-hardness proofs hold for finite query answering for both $\mathcal{ALCI}$ and $\mathcal{ALCO}$, our upper bound is tight for all logics containing either of these. For $\mathcal{SF}$ and its fragments, the best known lower bound is co-NEXPTIME of query answering in $\mathcal{S}$ (Eiter et al. 2009).

One crucial aspect in our techniques is the ability to define a suitable notion of decomposition of counter-models. This appears to be more challenging for logics with role inclusions, and we conjecture that for fragments of $\mathcal{SOIF}$ extended with role inclusions a different approach is needed. A promising direction for future work is to push our techniques to establish tight bounds for Horn fragments of $\mathcal{SOIF}$.

Acknowledgements. This work was done in the course of several meetings in Vienna and Warsaw, made possible by a grant from the Austrian Agency for International Cooperation in Education and Research (OeAD-GmbH) awarded for the project *Logic-based Methods in Data Management and Knowledge Representation* within the call *WTZ Poland 2017-19*. The whole project and this work in particular is a result of three great brainstorming sessions with Claire David, Magdalena Ortiz, and Mantas Simkus back in 2016. The first author was supported by Poland's National Science Centre grant 2016/21/D/ST6/01485. Last but not least, we salute the reviewers of KR 2018 for their hard work.

References

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

Amendola, G.; Leone, N.; and Manna, M. 2017. Finite model reasoning over existential rules. *TPLP* 17(5-6):726–743.

Baader, F.; Calvanese, D.; McGuinness, D. L.; Nardi, D.; and Patel-Schneider, P. F. 2010. *The Description Logic Handbook: Theory, Implementation and Applications*. New York, NY, USA: Cambridge University Press, 2nd edition.

Baget, J.; Leclère, M.; Mugnier, M.; and Salvat, E. 2011. On rules with existential variables: Walking the decidability line. *Artif. Intell.* 175(9-10):1620–1654.

Bárány, V.; Gottlob, G.; and Otto, M. 2014. Querying the guarded fragment. *Log. Meth. Comput. Sci.* 10(2).

Bienvenu, M., and Ortiz, M. 2015. Ontology-mediated query answering with data-tractable description logics. In *Reasoning Web*, volume 9203 of *LNCS*, 218–307. Springer.

Calì, A.; Lembo, D.; and Rosati, R. 2003. On the decidability and complexity of query answering over inconsistent and incomplete databases. In *PODS*, 260–271. ACM.

Calvanese, D. 1996. Finite model reasoning in description logics. In *KR*, 292–303. Morgan Kaufmann.

Civili, C., and Rosati, R. 2012. A broad class of first-order rewritable tuple-generating dependencies. In *Datalog 2.0*, 68–80.

Cosmadakis, S. S.; Kanellakis, P. C.; and Vardi, M. Y. 1990. Polynomial-time implication problems for unary inclusion dependencies. *J. ACM* 37(1):15–46.

Eiter, T.; Lutz, C.; Ortiz, M.; and Simkus, M. 2009. Query answering in description logics with transitive roles. In *IJCAI*, 759–764.

Gogacz, T., and Marcinkowski, J. 2013. On the BDD/FC conjecture. In *PODS*, 127–138.

Gogacz, T., and Marcinkowski, J. 2017. Converging to the chase – A tool for finite controllability. *J. Comput. Syst. Sci.* 83(1):180–206.

Ibáñez-García, Y. A.; Lutz, C.; and Schneider, T. 2014. Finite model reasoning in horn description logics. In *KR*. AAAI Press.

Johnson, D. S., and Klug, A. C. 1984. Testing containment of conjunctive queries under functional and inclusion dependencies. *J. Comput. Syst. Sci.* 28(1):167–189.

Kazakov, Y. 2008. RIQ and SROIQ are harder than SHOIQ. In *KR*, 274–284. AAAI Press.

Lutz, C.; Sattler, U.; and Tendera, L. 2005. The complexity of finite model reasoning in description logics. *Inf. Comput.* 199(1-2):132–171.

Lutz, C. 2008. The complexity of conjunctive query answering in expressive description logics. In *IJCAR*, volume 5195 of *LNCS*, 179–193. Springer.

Ngo, N.; Ortiz, M.; and Simkus, M. 2016. Closed predicates in description logics: Results on combined complexity. In *AMW*, volume 1644 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Pratt-Hartmann, I. 2007. Complexity of the guarded two-variable fragment with counting quantifiers. *J. Log. Comput.* 17(1):133–155.

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

Pratt, V. R. 1979. Models of program logics. In *FOCS*, 115–122. IEEE Computer Society.

Rosati, R. 2008. Finite model reasoning in dl-lite. In *ESWC*, volume 5021 of *LNCS*, 215–229. Springer.

Rosati, R. 2011. On the finite controllability of conjunctive query answering in databases under open-world assumption. *J. Comput. Syst. Sci.* 77(3):572–594.

Rudolph, S., and Glimm, B. 2010. Nominals, inverses, counting, and conjunctive queries or: Why infinity is your friend! *J. Artif. Intell. Res.* 39:429–481.

Rudolph, S.; Krötzsch, M.; and Hitzler, P. 2012. Type-elimination-based reasoning for the description logic SHIQbs using decision diagrams and disjunctive datalog. *Log. Meth. Comput. Sci.* 8(1).

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