On Query Answering in Description Logics with Number Restrictions on Transitive Roles*

Víctor Gutiérrez-Basulto[1], Yazmín Ibáñez-García[2], and Jean Christoph Jung[3]

[1] Cardiff University, UK (gutierrezbasultov@cardiff.ac.uk)
[2] TU Wien, Austria (ibanez@kr.tuwien.ac.at)
[3] Universität Bremen, Germany (jeanjung@informatik.uni-bremen.de)

Abstract. We study query answering in the description logic $\mathcal{SQ}$ supporting number restrictions on both transitive and non-transitive roles. Our main contributions are (i) a tree-like model property for $\mathcal{SQ}$ knowledge bases and, building upon this, (ii) an automata based decision procedure for answering two-way regular path queries, which gives a 3ExpTime upper bound.

1 Introduction

In the last years, several efforts have been put into the study of the query answering problem (QA) in description logics (DLs) featuring transitive roles (or generalisations thereof, such as regular expressions on roles) and number restrictions, see e.g., [10, 11, 9, 7, 8] and references therein. However, all these DLs heavily restrict the interaction between these two features, or altogether forbid number restrictions on transitive roles. Unfortunately, this comes as a shortcoming in crucial DL application areas like medicine and biology in which many terms, e.g., proteins, are defined and classified according to the number of components they contain or are part of (in a transitive sense) [27, 22, 24].

The lack of investigations of query answering in DLs of this kind is partly because (i) the interaction of these features often leads to undecidability of the standard reasoning tasks (e.g., satisfiability) - even in lightweight sub-Boolean DLs with unqualified number restrictions [17, 20, 15]; and (ii) for those DLs known to be decidable, such as $\mathcal{SQ}$ and $\mathcal{SOQ}$ [20, 18], only recently tight complexity bounds were obtained [15]. Moreover, even if these features (with restricted interaction) do not necessarily increase the complexity of QA, they do pose additional challenges for devising decision procedures [10, 11, 9] since they lead to the loss of properties, such as the tree model property, which make the design of algorithms for QA simpler. In fact, these difficulties are present already in DLs with transitivity, but without number restrictions [9]. Clearly, these issues are exacerbated if number restrictions are imposed on transitive roles.

The objective of this paper is to start the investigation of query answering in DLs supporting number restrictions on transitive roles. In particular, we look

* Gutiérrez-Basulto was funded by the EU's Horizon 2020 programme under the Marie Skłodowska-Curie grant No 663830

at the problem of answering regular path queries, which generalise standard query languages like positive existential queries, over $\mathcal{SQ}$ knowledge bases [16]. We first develop tree-like decompositions of $\mathcal{SQ}$-interpretations based on a novel unraveling that is specially tailored to handle the interaction of transitivity with number restrictions. Using these decompositions, we design an algorithm for the query answering problem using two-way alternating tree automata in the spirit of [10, 7, 8], resulting in a 3ExpTime upper bound (leaving an exponential gap).

Related Work. Schröder and Pattinson [23] investigate the DL $\mathcal{PHQ}$ supporting number restrictions on transitive parthood roles, which are, in contrast to $\mathcal{SQ}$, interpreted as trees: parthood-siblings cannot have a common part. They show that under this assumption decidability (for satisfiability) can be attained.

There has been some work on the extension of decidable first-order logic fragments, such as the guarded fragment, with transitivity and counting, see e.g., [25, 21]. Unfortunately, this case leads to undecidability unless the interaction is severely restricted [25]. Closer to DLs is the detailed investigation of modal logics with graded modalities carried out in [19]. Finally, in the context of existential rules, several efforts have been recently made to design languages with decidable QA supporting transitivity [12, 4, 1]. However, we are not aware of any attempts to additionally support number restrictions.

2 Preliminaries

Syntax. We introduce the DL $\mathcal{SQ}$, which extends the classical DL $\mathcal{ALC}$ with transitivity declarations on roles ($\mathcal{S}$) and qualified number restrictions ($\mathcal{Q}$). We consider a vocabulary consisting of countably infinite disjoint sets of *concept names* $\mathsf{N_C}$, *role names* $\mathsf{N_R}$, and *individual names* $\mathsf{N_I}$, and assume that $\mathsf{N_R}$ is partitioned into two countably infinite sets of *non-transitive role names* $\mathsf{N_R}^{nt}$ and *transitive role names* $\mathsf{N_R}^{t}$. The syntax of *$\mathcal{SQ}$-concepts* C, D is given by the grammar rule $C, D ::= A \mid \neg C \mid C \sqcap D \mid (\leq n\ r\ C)$ where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and n is a number given in binary. We use $(\geq n\ r\ C)$ as an abbreviation for $\neg(\leq (n-1)\ r\ C)$, and other standard abbreviations like $\bot$, $\top$, $C \sqcup D$, $\exists r.C$, $\forall r.C$. Concepts of the form $(\leq n\ r\ C)$ and $(\geq n\ r\ C)$ are called *at most-restrictions* and *at least-restrictions*, respectively.

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

Semantics. As usual, the semantics is defined in terms of interpretations. An *interpretation* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ consists 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 such that transitive role names are mapped to transitive relations. We define $C^\mathcal{I}$ for complex concepts C by interpreting $\neg$ and $\sqcap$ as usual and $(\leq n\ r\ D)^\mathcal{I}$ by taking

$$(\leq n\ r\ D)^\mathcal{I} = \{d \in \Delta^\mathcal{I} \mid \ |\{e \in C^\mathcal{I} \mid (d, e) \in r^\mathcal{I}\}| \leq n\}.$$

For ABoxes $\mathcal{A}$ we adopt the *standard name assumption (SNA)*, that is, $a^{\mathcal{I}} = a$, for all $a \in \mathsf{ind}(\mathcal{A})$, but we conjecture that our results hold without the unique name assumption. The satisfaction relation $\models$ is defined in the standard way:

$$\mathcal{I} \models C \sqsubseteq D \text{ iff } C^{\mathcal{I}} \subseteq D^{\mathcal{I}}; \quad \mathcal{I} \models A(a) \text{ iff } a \in A^{\mathcal{I}}; \quad \mathcal{I} \models r(a,b) \text{ iff } (a,b) \in r^{\mathcal{I}}.$$

An interpretation $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$, denoted $\mathcal{I} \models \mathcal{T}$, if $\mathcal{I} \models \alpha$ for all $\alpha \in \mathcal{T}$; it is a model of an ABox $\mathcal{A}$, written $\mathcal{I} \models \mathcal{A}$, if $\mathcal{I} \models \alpha$ for all $\alpha \in \mathcal{A}$; it is a model of a KB $\mathcal{K}$ if $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \models \mathcal{A}$. A KB is *satisfiable* if it has a model.

Query Language. As query language, we consider *regular path queries*, supporting regular expressions over roles. Recall that a *regular expression $\mathcal{E}$ over an alphabet Σ* is given by the grammar $\mathcal{E} ::= \varepsilon \mid \sigma \mid \mathcal{E}\cdot\mathcal{E} \mid \mathcal{E} \cup \mathcal{E} \mid \mathcal{E}^*$, where $\sigma \in \Sigma$ and ε denotes the *empty word*. We denote with $L(\mathcal{E})$ the language defined by $\mathcal{E}$.

We use $\mathsf{N}_{\mathsf{R}}^{\pm}$ to refer to $\mathsf{N}_{\mathsf{R}} \cup \{r^- \mid r \in \mathsf{N}_{\mathsf{R}}\}$ with $(r^-)^{\mathcal{I}}$ defined as $\{(d,e) \mid (e,d) \in r^{\mathcal{I}}\}$, and identify r^- with $s \in \mathsf{N}_{\mathsf{R}}$ if $r = s^-$. A *positive 2-way regular path query (P2RPQ)* is a formula of the form $q(\boldsymbol{x}) = \exists \boldsymbol{y}.\varphi(\boldsymbol{x}, \boldsymbol{y})$ where $\boldsymbol{x}$ and $\boldsymbol{y}$ are tuples of variables and φ is constructed using $\wedge$ and $\vee$ of atoms of the form $A(t)$ or $\mathcal{E}(t,t')$ where $A \in \mathsf{N}_{\mathsf{C}}$, $\mathcal{E}$ is a regular expression over $\mathcal{S} ::= \mathsf{N}_{\mathsf{R}}^{\pm} \cup \{A? \mid A \in \mathsf{N}_{\mathsf{C}}\}$, and t, t' are *terms*, i.e., individual names or variables from $\boldsymbol{x}, \boldsymbol{y}$. We define as usual when a possible answer tuple $\boldsymbol{a} \in \mathsf{ind}(\mathcal{A})$ is a *certain answer of q over $\mathcal{K}$* and write $\mathcal{K} \models q(\boldsymbol{a})$ in case it is [8, 6].

Reasoning Problem. We study the *certain answers problem:* Given a KB $\mathcal{K}$, a query $q(\boldsymbol{x})$ and a tuple of individuals $\boldsymbol{a}$, determine whether $\mathcal{K} \models q(\boldsymbol{a})$. Without loss of generality, we consider Boolean queries.

3 Decomposing $\mathcal{SQ}$-Interpretations

Existing algorithms for QA in expressive DLs, e.g., $\mathcal{SHIQ}$ (without number restrictions on transitive roles), exploit the fact that for answering queries it suffices to consider *canonical models* that are forest-like, roughly consisting of an interpretation of the ABox and a collection of tree-interpretations whose roots are elements of the ABox. However, for $\mathcal{SQ}$ this tree-model property is lost:

Example 1. Let $\mathcal{T} = \{A \sqsubseteq (\leq 1\ r\ C) \sqcap \exists r.B \sqcap \exists r.\neg B, \top \sqsubseteq \exists r.C\}$ with $r \in \mathsf{N}_{\mathsf{R}}^{t}$. The number restrictions in $\mathcal{T}$ force that every model of $\mathcal{T}$ satisfying A contains the structure in Fig. 1(a). Moreover, in $\mathcal{SQ}$ strongly connected components can be enforced. Let $\mathcal{T}' = \{A \sqsubseteq (= 3\ r\ B), B \sqsubseteq (= 3\ r\ B), A \sqsubseteq \neg B\}$ with $r \in \mathsf{N}_{\mathsf{R}}^{t}$. Then, in every model of $\mathcal{T}'$, an element satisfying A roots the structure depicted in Fig. 1(b), where the elements satisfying B form a strongly connected component.

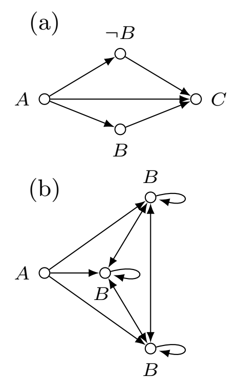

Fig. 1.

Nevertheless, we will define *tree-like* canonical models for $\mathcal{SQ}$ that suffice for query answering. We start with introducing a basic form of tree decompositions of $\mathcal{SQ}$-interpretations.

A *tree* is a connected, acyclic graph (T, E) with a distinguished root, which we usually denote with ε. We usually write only T instead of (T, E), thus leaving E implicit. Fix some interpretation $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$. A *bag* M is a set of assertions of the form $A(d), r(d, e)$ with $d, e \in \Delta^\mathcal{I}$. We denote with $\mathsf{dom}(M)$ the set of domain elements appearing in M. Given a set $\Lambda \subseteq \Delta^\mathcal{I}$, we denote with $\mathsf{bag}_\mathcal{I}(\Lambda)$ the set

$$\mathsf{bag}_\mathcal{I}(\Lambda) = \{A(d) \mid d \in \Lambda, d \in A^\mathcal{I}\} \cup \{r(d, e) \mid d, e \in \Lambda, (d, e) \in r^\mathcal{I}\}.$$

Definition 2. *A* tree decomposition of an interpretation $\mathcal{I}$ *is a tuple* (T, bg) *where* T *is a tree and* bg *assigns a bag to every node* w *in* T *such that:*

(i) $\mathsf{bg}(w) = \mathsf{bag}_\mathcal{I}(\mathsf{dom}(\mathsf{bg}(w)))$*;*
(ii) $\Delta^\mathcal{I} = \bigcup_{w \in T} \mathsf{dom}(\mathsf{bg}(w))$*;*
(iii) $r^\mathcal{I} = \chi_r$ *for non-transitive* r *and* $r^\mathcal{I} = \chi_r^+$ *for transitive* r*, where*

$$\chi_r = \{(d, e) \mid r(d, e) \in \mathsf{bg}(w), w \in T\};$$

(iv) for all $d \in \Delta^\mathcal{I}$*, the set* $\{w \in T \mid d \in \mathsf{dom}(\mathsf{bag}(w))\}$ *is connected in* T*.*

Definition 2 provides a variant of tree decompositions of interpretations with transitive relations. This formalisation does not yet enable tree automata to count over transitive roles (with a small number of states) since assertions $r(d, e)$ can appear far away from each other in the decomposition. To address this, we introduce *canonical tree decompositions* which extend tree decompositions with a third component rl which assigns a role name to every non-root node of T and $\perp$ to the root ε. Intuitively, a node $w \in T$ labeled with $r = \mathsf{rl}(w)$ will be responsible for capturing r-successors of some element(s) in the previous bag.

Fix a triple $(T, \mathsf{bg}, \mathsf{rl})$ such that (T, bg) is a tree decomposition of $\mathcal{I}$. By Item *(iv)* of Definition 2, for each $d \in \Delta^\mathcal{I}$, there is a unique node $w \in T$ such that all occurrences of d are in or below w in T. In this case, we say that d is *fresh* in w, and we denote with $F(w)$ the set of all fresh elements in w. We will also need a relativised version of freshness which takes into account the role associated to the predecessor bag. In particular, for each transitive role r and each $w \in T$ with $\mathsf{rl}(w) \in \{r, \perp\}$, we define a set $F_r(w)$ by taking:

- $F_r(w) = F(w)$ if the predecessor $\hat{w}$ of w satisfies $\mathsf{rl}(\hat{w}) \in \{r, \perp\}$;
- $F_r(w) = \mathsf{dom}(\mathsf{bg}(w))$ otherwise.

Intuitively, $F_r(w)$ contains all elements that are eligible as origins for r-successors.

For a transitive role r and a bag M, we call $\emptyset \subsetneq \mathbf{a} \subseteq \mathsf{dom}(M)$ an *r-cluster in M* if *(i)* $r(a, b) \in M$ for all $a \neq b \in \mathbf{a}$, and *(ii)* for all $a \in \mathbf{a}$, $b \in \mathsf{dom}(M)$ with $r(a, b), r(b, a) \in M$, we have $b \in \mathbf{a}$. An r-cluster $\mathbf{a}$ in M is an *r-root cluster in M* if $r(d, e) \in M$ for all $d \in \mathbf{a}$ and $e \in \mathsf{dom}(M) \setminus \mathbf{a}$.

Definition 3. *A triple* $\mathfrak{T} = (T, \mathsf{bg}, \mathsf{rl})$ *is a* canonical tree decomposition of $\mathcal{I}$ *if* (T, bg) *is a tree decomposition of* $\mathcal{I}$ *and the following conditions are satisfied for every* $w \in T$ *with* $M = \mathsf{bg}(w)$ *and* $r = \mathsf{rl}(w)$ *and every successor* w' *of* w *with* $M' = \mathsf{bg}(w')$ *and* $r' = \mathsf{rl}(w')$*:*

(C1) if $r' \in \mathsf{N}_\mathsf{R}^{nt}$*, then* $\mathsf{dom}(M') = \{d, e\}$*, for some* $d \in F(w)$*,* $e \in F(w')$*, and* $r'(d, e)$ *is the only role assertion in* M'*;*

(C2) *if $r' \in \mathsf{N}_\mathsf{R}^t$ and $r \notin \{\bot, r'\}$, then there are $d \in F(w)$ and an r-root cluster $\mathbf{a}$ in M' such that $\mathsf{dom}(M) \cap \mathsf{dom}(M') = \{d\}$ and $d \in \mathbf{a}$; moreover, there is no successor v' of w different from w' satisfying this for d and $\mathsf{rl}(v') = r'$;*

(C3) *if $r' \in \mathsf{N}_\mathsf{R}^t$ and $r \in \{\bot, r'\}$, then there is an r'-cluster $\mathbf{a}$ in M with:*
- *(a) $\mathbf{a} \subseteq F_{r'}(w)$;*
- *(b) $\mathbf{a}$ is an r'-root cluster in M';*
- *(c) for all $d \in \mathbf{a}$ and $r'(d,e) \in M$, we have $e \in \mathsf{dom}(M')$; and*
- *(d) for all $r'(d,e) \in M'$, $d \in \mathbf{a} \cup F(w')$ or $e \notin F(w')$.*

Definition 3 imposes restrictions on the structural relation between neighbouring bags. Note that Condition (C1) is also satisfied by standard unravelings over non-transitive roles [2]. Condition (C2) reflects that neighbouring bags associated with different role names do only interact via single domain elements; this conforms with viewing $\mathcal{SQ}$ as a fusion logic [3]. Most interestingly, Condition (C3) plays the role of (C1), but for transitive roles. It is important to note that (C3) is based on r-clusters now since they can be enforced, see Example 1 above. Item *(a)* restricts for which elements $\mathbf{a}$ we can have successor bags; Item *(b)* requires that $\mathbf{a}$ is a root cluster in the successor bag M'; Item *(c)* states that everything which was reachable from $\mathbf{a}$ via r' in M should be also included in M'; finally, Item *(d)* requires that there are no connections $r(d,e) \in M'$ between elements d from M and fresh elements e from M'.

As a consequence of Definition 3, we can address r-successors of elements in a canonical tree decomposition $\mathfrak{T} = (T, \mathsf{bg}, \mathsf{rl})$ of $\mathcal{I}$. For a non-transitive role r, Condition (C1) ensures that r-successors e of d are contained only in successor nodes of the (unique) node where d is fresh. For a transitive role r, note first that $(d,e) \in r^\mathcal{I}$ iff there is a sequence $d_0, w_0, d_1, \ldots, w_{n-1}, d_n$ with $d_i \in \Delta^\mathcal{I}$ and $w_i \in T$ such that $d = d_0$, $e = d_n$, and $r(d_i, d_{i+1}) \in \mathsf{bg}(w_i)$, for all $0 \leq i < n$; we call such a sequence an *r-path from d to e in $\mathfrak{T}$*. We call an r-path $d_0, w_0, d_1, \ldots, w_{n-1}, d_n$ *downward* if w_i is a successor of w_{i-1} and d_i is contained in an r-root cluster of w_i, for all $0 < i < n$. An r-path in $\mathfrak{T}$ is *canonical* if **P1**: it is downward; or **P2**: $d_0 \in F_r(w_0)$, $d_1 \notin F_r(w_0)$, and, if $n > 1$, then $d_1, w_1, \ldots, d_n$ is a downward path in $\mathfrak{T}$ and the predecessor $\hat{w}$ of w_1 is an ancestor of w_0 and satisfies $d_1 \in F_r(\hat{w})$. Two r-paths $d_0, w_0, d_1, \ldots, w_{n-1}, d_n$ and $e_0, w'_0, e_1, \ldots, w'_{m-1}, e_m$ from d to e are *equivalent* if $n = m$, $w_i = w'_i$, for $0 \leq i < n$, and d_i and e_i are in the same r-cluster in $\mathsf{bg}(w_i)$, for every $1 \leq i < n$.

Lemma 4 establishes the basis for uniquely identifying transitive r-successors in a canonical tree decomposition which is essential for the design of automata.

Lemma 4. *Let $\mathfrak{T}$ be a canonical decomposition of $\mathcal{I}$, r transitive, and $(d,e) \in r^\mathcal{I}$. Then there is a unique canonical r-path up to equivalence from d to e in $\mathfrak{T}$.*

3.1 Unraveling into Canonical Decompositions

We give now the main technical contribution of our paper: an unraveling operation into canonical decompositions of small width, and consequently a tree-like model property for $\mathcal{SQ}$-interpretations. A canonical tree decomposition $(T, \mathsf{bg}, \mathsf{rl})$ has *width* $k-1$ if k is the maximum size of $\mathsf{dom}(\mathsf{bg}(w))$, where w ranges over T; its *outdegree* is the outdegree of the underlying tree T.

Theorem 5. *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{SQ}$ KB and $\mathcal{I} \models \mathcal{K}$. Then, there is an interpretation $\mathcal{J}$ and a canonical tree decomposition $(T, \mathsf{bg}, \mathsf{rl})$ of $\mathcal{J}$ such that:*

(1) $\mathcal{A} \subseteq \mathsf{bg}(\varepsilon)$;
(2) $\mathcal{J} \models \mathcal{K}$;
(3) there is a homomorphism from $\mathcal{J}$ to $\mathcal{I}$;
(4) width and outdegree of $(T, \mathsf{bg}, \mathsf{rl})$ are bounded by $O(|\mathcal{A}| \cdot 2^{\mathsf{poly}(|\mathcal{T}|)})$.

We outline the proof of Theorem 5. As a first step, we show that wlog. we can assume that $\mathcal{I}$ has a restricted outdegree and width, as defined below. This will be used later on to ensure the satisfaction of Condition (4) above. Given $d \in \Delta^{\mathcal{I}}$ and a transitive role r, the *r-cluster of d in $\mathcal{I}$,* denoted by $Q_{\mathcal{I},r}(d)$, is the set of all elements $e \in \Delta^{\mathcal{I}}$ such that both $(d, e) \in r^{\mathcal{I}}$ and $(e, d) \in r^{\mathcal{I}}$. The *width of* $\mathcal{I}$ is the minimum k such that $|Q_{\mathcal{I},r}(d)| \leq k$ for all $d \in \Delta^{\mathcal{I}}$, $r \in \mathsf{N}_{\mathsf{R}}^t$. Moreover, for a transitive role r, we say that e is a *direct r-successor of d* if $(d, e) \in r^{\mathcal{I}}$ but $e \notin Q_{\mathcal{I},r}(d)$, and for each f with $(d, f), (f, e) \in r^{\mathcal{I}}$, we have $f \in Q_{\mathcal{I},r}(d)$ or $f \in Q_{\mathcal{I},r}(e)$; if r is non-transitive, then e is a *direct r-successor of* d if $(d, e) \in r^{\mathcal{I}}$. The *breadth of* $\mathcal{I}$ is the maximum k such that there are $d, d_1, \ldots, d_k$ and a role name r, all d_i are direct r-successors of d, and

– if r is non-transitive, then $d_i \neq d_j$ for all $i \neq j$;
– if r is transitive, then $Q_{\mathcal{I},r}(d_i) \neq Q_{\mathcal{I},r}(d_j)$, for all $i \neq j$.

We can assume that width and breath of $\mathcal{I}$ are within the following boundaries.

Lemma 6 (adapting [19, 15]). *For each $\mathcal{I} \models \mathcal{K}$, there is a sub-interpretation $\mathcal{I}'$ of $\mathcal{I}$ with $\mathcal{I}' \models \mathcal{K}$ and width and breadth of $\mathcal{I}'$ are bounded by $O(|\mathcal{A}| + 2^{\mathsf{poly}(|\mathcal{T}|)})$.*

We need to introduce one more notion for dealing with at-most restrictions over transitive roles. Let $\mathsf{cl}(\mathcal{T})$ be the set of all subconcepts occurring in $\mathcal{T}$, closed under single negation. For each transitive role r, define a binary relation $\rightsquigarrow_{\mathcal{I},r}$ on $\Delta^{\mathcal{I}}$, by taking $d \rightsquigarrow_{\mathcal{I},r} e$ if there is some $(\leq n\ r\ C) \in \mathsf{cl}(\mathcal{T})$ such that $d \in (\leq n\ r\ C)^{\mathcal{I}}$, $e \in C^{\mathcal{I}}$, and $(d, e) \in r^{\mathcal{I}}$. Based on the transitive, reflexive closure $\rightsquigarrow^*_{\mathcal{I},r}$ of $\rightsquigarrow_{\mathcal{I},r}$, we define, for every $d \in \Delta^{\mathcal{I}}$, the set $\mathsf{Wit}_{\mathcal{I},r}(d)$ of *witnesses for* d by taking

$$\mathsf{Wit}_{\mathcal{I},r}(d) = \textstyle\bigcup_{e \mid d \rightsquigarrow^*_{\mathcal{I},r} e} Q_{\mathcal{I},r}(e).$$

Intuitively, $\mathsf{Wit}_{\mathcal{I},r}(d)$ contains all witnesses of at-most restrictions of some element d, and due to using $\rightsquigarrow^*_{\mathcal{I},r}$, also the witnesses of at-most restrictions of those witnesses and so on. It is important to note that the size of $\mathsf{Wit}_{\mathcal{I},r}(\mathcal{T})$ is bounded exponentially in $\mathcal{T}$ (and linearly in $\mathcal{A}$), see appendix.

We describe now the construction of the interpretation $\mathcal{J}$ and its tree decomposition via a possibly infinite unraveling process. Elements of $\Delta^{\mathcal{J}}$ will be either of the form a with $a \in \mathsf{ind}(\mathcal{A})$ or of the form d_x with $d \in \Delta^{\mathcal{I}}$ and some index x. We usually use δ to refer to domain elements in $\mathcal{J}$ (in either form), and define a function $\tau : \Delta^{\mathcal{J}} \to \Delta^{\mathcal{I}}$ by setting $\tau(\delta) = \delta$, for all $\delta \in \mathsf{ind}(\mathcal{A})$, and $\tau(\delta) = d$, for all δ of the form d_x in $\Delta^{\mathcal{J}}$.

To start the construction of $\mathcal{J}$ and $(T, \mathsf{bg}, \mathsf{rl})$, we set $\mathcal{J} = \mathcal{I}|_{\mathsf{ind}(\mathcal{A})}$ and, for every transitive role r, define two sets Δ_r, Δ'_r by taking

$$\Delta_r = \{d_r \mid d \in \bigcup_{a \in \mathsf{ind}(\mathcal{A})} \mathsf{Wit}_{\mathcal{I},r}(a) \setminus \mathsf{ind}(\mathcal{A})\} \quad \text{and} \quad \Delta'_r = \Delta_r \cup \mathsf{ind}(\mathcal{A}).$$

Then extend $\mathcal{J}$ by adding, for each transitive r, Δ_r to the domain and extending the interpretation of concept and role names such that, for all $\delta, \delta' \in \Delta'_r$, we have

$$\delta \in A^{\mathcal{J}} \Leftrightarrow \tau(\delta) \in A^{\mathcal{I}}, \text{ for all } A \in \mathsf{N_C}, \text{ and } (\delta, \delta') \in r^{\mathcal{J}} \Leftrightarrow (\tau(\delta), \tau(\delta')) \in r^{\mathcal{I}}. \quad (\dagger)$$

Now, initialise $(T, \mathsf{bg}, \mathsf{rl})$ with $T = \{\varepsilon\}$, $\mathsf{bg}(\varepsilon) = \mathsf{bag}_{\mathcal{J}}(\Delta^{\mathcal{J}})$, and $\mathsf{rl}(\varepsilon) = \bot$. Intuitively, this first step ensures that all witnesses of ABox individuals appear in the first bag. This finishes the initialisation phase.

Next, extend $\mathcal{J}$ and $(T, \mathsf{bg}, \mathsf{rl})$ by applying the following rules exhaustively and in a fair way:

$\mathbf{R_1}$ Let r be non-transitive, $w \in T$, $\delta \in F(w)$, and d a direct r-successor of $\tau(\delta)$ in $\mathcal{I}$ with $\{\delta, d\} \not\subseteq \mathsf{ind}(\mathcal{A})$. Then, add a fresh successor v of w to T, add a fresh element d_v to $\Delta^{\mathcal{J}}$, extend $\mathcal{J}$ by adding $(\delta, d_v) \in r^{\mathcal{J}}$ and $d_v \in A^{\mathcal{J}}$ iff $d \in A^{\mathcal{I}}$, for all $A \in \mathsf{N_C}$, and set $\mathsf{bg}(v) = \mathsf{bag}_{\mathcal{J}}(\{\delta, d_v\})$ and $\mathsf{rl}(v) = r$.

$\mathbf{R_2}$ Let r be transitive, $w \in T$, and $\delta \in F(w)$ such that:
(a) $w = \varepsilon$ and $\delta \in \Delta_s$, $s \neq r$ (Δ_s defined in the initialisation phase), or
(b) $w \neq \varepsilon$ and $\mathsf{rl}(w) \neq r$.
Then add a fresh successor v of w to T, and define

$$\Delta = \{e_v \mid e \in \mathsf{Wit}_{\mathcal{I},r}(\tau(\delta)) \setminus \{\tau(\delta)\}\} \quad \text{and} \quad \Delta' = \Delta \cup \{\delta\}.$$

Extend the domain of $\mathcal{J}$ with Δ and the interpretation of concept and role names such that ($\dagger$) is satisfied for all $\delta, \delta' \in \Delta'$. Finally, set $\mathsf{bg}(v) = \mathsf{bag}_{\mathcal{J}}(\Delta')$ and $\mathsf{rl}(v) = r$.

$\mathbf{R_3}$ Let r be transitive, $w \in T$, $\mathbf{a} \subseteq F_r(w)$ an r-cluster in $\mathsf{bg}(w)$ such that:
(a) $w = \varepsilon$ and $\mathbf{a} \subseteq \Delta'_r$, or
(b) $w \neq \varepsilon$ and $\mathsf{rl}(w) = r$.
If there is a direct r-successor e of $\tau(\delta)$ in $\mathcal{I}$ for some $\delta \in \mathbf{a}$ such that $(\delta, \delta') \notin r^{\mathcal{J}}$ for any δ' with $\tau(\delta') = e$, then add a fresh successor v of w to T, and define

$$\Delta = \{f_v \mid f \in \mathsf{Wit}_{\mathcal{I},r}(e) \setminus \mathsf{Wit}_{\mathcal{I},r}(\tau(\delta))\} \quad \text{and}$$
$$\Delta' = \Delta \cup \mathbf{a} \cup \{\delta'' \mid r(\delta', \delta'') \in \mathsf{bg}(w) \text{ for some } \delta' \in \mathbf{a}\}.$$

Then extend the domain of $\mathcal{J}$ with Δ and the interpretation of concept names such that ($\dagger$) is satisfied for all pairs δ, δ' with $\delta \in \mathbf{a} \cup \Delta$ and $\delta' \in \Delta'$. Finally, set $\mathsf{bg}(v) = \mathsf{bag}_{\mathcal{J}}(\Delta')$ and $\mathsf{rl}(v) = r$.

Rules $\mathbf{R_1}$–$\mathbf{R_3}$ are, respectively, in one-to-one correspondence with Conditions (C1)-(C3) in Definition 3. In particular, $\mathbf{R_1}$ implements the well-known unraveling procedure for non-transitive roles. $\mathbf{R_2}$ is used to change the 'role component' for transitive roles by creating a bag which contains all witnesses of the chosen element δ. Finally, $\mathbf{R_3}$ describes how to unravel direct r-successors in case of transitive roles r. In the definition of Δ it is taken care that witnesses which are 'inherited' from predecessors are not introduced again, in order to preserve at-most restrictions.

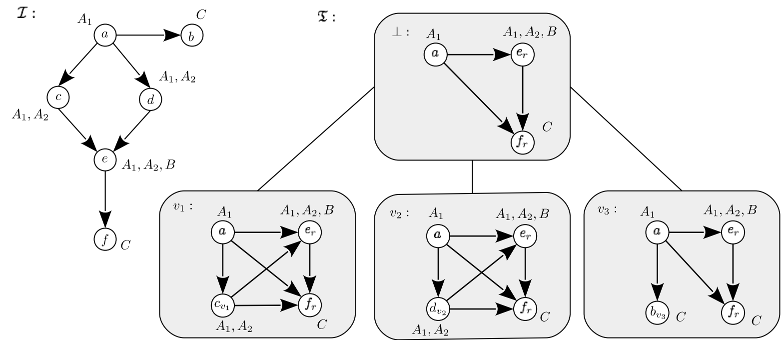

Fig. 2.

Example 7. Let $\mathcal{K} = (\{A_1 \sqsubseteq (\leq 1\ r\ B), A_2 \sqsubseteq (\leq 1\ r\ C)\}, \{A_1(a)\})$ with $r \in \mathsf{N}_\mathsf{R}^t$. Fig. 2 shows a model $\mathcal{I}$ of $\mathcal{T}$ and a canonical decomposition $\mathfrak{T}$ of its unraveling (transitivity connections are omitted). In the initialisation phase, the $\perp$-bag is constructed starting from individual a. Since $a \rightsquigarrow_{\mathcal{I},r} e$ and $e \rightsquigarrow_{\mathcal{I},r} f$, we have $\mathsf{Wit}_{\mathcal{I},r}(a) = \{e, f\}$, thus e_r and f_r are added in this phase.

It is verified in the appendix that $(T, \mathsf{bg}, \mathsf{rl})$ and $\mathcal{J}$ satisfy the conditions from Theorem 5. Theorem 5 yields a tree-like model property for $\mathcal{SQ}$-knowledge bases, which is interesting on its own since existing decidability results (for satisfiability) [20, 15] are based on the finite model property.

4 Automata-Based Approach to Query Answering

In this section, we devise an automata-based decision procedure for query answering in $\mathcal{SQ}$. By Theorem 5, if $\mathcal{K} \not\models q$, there is an interpretation of small width and outdegree witnessing this. The idea is now to design two automata $\mathfrak{A}_\mathcal{K}$ and $\mathfrak{A}_q$ working over tree decompositions which accept precisely the models (of a fixed width) of the KB $\mathcal{K}$ and the query q, respectively. Query answering is then reduced to the question if some tree is accepted by $\mathfrak{A}_\mathcal{K}$, but not by $\mathfrak{A}_q$ [8].

Trees are represented as prefix-closed subsets $T \subseteq (\mathbb{N} \setminus \{0\})^*$ such that additionally, $wc \in T$ implies $w(c-1) \in T$ for all $c > 1$. A tree is *k-ary* if each node has *exactly* k successors. As a convention, we set $w \cdot 0 = w$ and $wc \cdot (-1) = w$, leave $\varepsilon \cdot (-1)$ undefined, and for any $k \in \mathbb{N}$, set $[k] = \{-1, 0, \dots, k\}$. Let Σ be a finite alphabet. A *Σ-labeled tree* is a pair (T, τ) with T a tree and $\tau : T \to \Sigma$ assigns a letter from Σ to each node. An *alternating 2-way tree automaton (2ATA)* over Σ-labeled k-ary trees is a tuple $\mathfrak{A} = (Q, \Sigma, q_0, \delta, F)$ where Q is a finite set of *states*, $q_0 \in Q$ is an *initial state*, δ is the *transition function*, and F is the *(parity) acceptance condition* [26]. The transition function maps a state q and an input letter $a \in \Sigma$ to a positive Boolean formula over the constants true and false, and variables from $[k] \times Q$. The semantics is given in terms of *runs*, see the appendix. As usual, $L(\mathfrak{A})$ denotes the set of trees accepted by $\mathfrak{A}$. Emptiness of $L(\mathfrak{A})$ can be checked in exponential time in the number of states of $\mathfrak{A}$ [26].

In principle, tree decompositions $\mathfrak{T}$ can be represented as labeled trees, where each node label consists of a bag and a role name (or $\perp$). However, 2ATAs cannot run over such labeled trees because the domain underlying the bags is potentially infinite. Exploiting the bounded width, we encode the infinite domain with finitely many elements in the following well-known way [14, 5]. Let $\mathcal{K}$ be an $\mathcal{SQ}$ KB, let K be the bound on the width obtained in Theorem 5, and choose a set of elements Δ of size $2K$. We define the input alphabet Σ as the set of all pairs $\langle M, x\rangle$ such that M is a bag that uses only constants from Δ, $|\mathsf{dom}(M)| \leq K$, and x is a role appearing in $\mathcal{K}$ or $\perp$. A Σ-labeled tree (T, τ) represents a tree decomposition (and thus an interpretation) as follows. Each domain element $d \in \Delta$ induces an equivalence relation $\sim_d$ on T by taking $v \sim_d w$ iff d appears in all bags on the path from v to w. Domain elements in the represented interpretation are then all equivalence classes obtained in this way. Moreover, for all $w \in T$, $\tau(w)$ represents the following bag:

$$\mathsf{bg}(w) = \{A([w]_{\sim_d}) \mid A(d) \in \tau(w))\} \cup \{r([w]_{\sim_d}, [w]_{\sim_e}) \mid r(d, e) \in \tau(w)\}.$$

We denote the interpretation associated with a Σ-labeled tree (T, τ) with $\mathcal{I}_{T,\tau}$. Moreover, we consider only k-ary trees where k is the bound on the outdegree given by Theorem 5. Since tree decompositions are not necessarily uniformly branching, we include an auxiliary symbol $\bullet$ to refer to non-existing branches.

Lemma 8. *There are 2ATAs $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3$ of size $O(|\mathcal{A}| \cdot 2^{\mathsf{poly}(|\mathcal{T}|)})$ such that: $(T, \tau) \in L(\mathfrak{A}_1)$ iff (T, τ) is the canonical decomposition of some interpretation; $(T, \tau) \in L(\mathfrak{A}_2)$ iff $\mathcal{I}_{T,\tau} \models \mathcal{A}$, and $(T, \tau) \in L(\mathfrak{A}_3)$ iff $\mathcal{I}_{T,\tau} \models \mathcal{T}$.*

The mentioned automaton $\mathfrak{A}_\mathcal{K}$ is obtained as the conjunction of $\mathfrak{A}_1$, $\mathfrak{A}_2$, and $\mathfrak{A}_3$. Note that $\mathfrak{A}_\mathcal{K}$ can be used to decide KB satisfiability in double exponential time, thus not optimal [15]. We concentrate here on the most interesting 2ATA $\mathfrak{A}_3$. Denote with $\mathsf{nnf}(C)$ the *negation normal form* of a concept C, and define $\mathsf{nnf}(\mathcal{T}) = \{\mathsf{nnf}(C) \mid C \in \mathsf{cl}(\mathcal{T})\}$. Moreover, let $\mathsf{Rol}(\mathcal{K})$ be the set of role names appearing in $\mathcal{K}$. Then, define $\mathfrak{A}_3 = (Q_3, \Sigma, q_0, \delta_3, F_3)$; start by including in Q_3

$$\{q_0\} \cup Q^{nt} \cup Q^t \cup \{F_{x,d}, F'_{x,d}, \overline{F}_{x,d}, \overline{F}'_{x,d} \mid d \in \Delta, x \in \{\perp\} \cup \mathsf{Rol}(\mathcal{K})\} \cup$$
$$\{q_d, q_{C,d} \mid C \in \mathsf{nnf}(\mathcal{T}), d \in \Delta\} \cup \{q^*_{C,d}, q'_{C,d} \mid C = (\sim\ n\, r\, D) \in \mathsf{nnf}(\mathcal{T}), d \in \Delta\}$$

where Q^t and Q^{nt} are the states that are used after entering states $q^*_{(\sim n\, r\, D),d}$ for transitive and non-transitive roles, respectively. Then, we define the transition function for all states except states of the form $q^*_{(\sim n\, r\, D),d}$:

$$\begin{aligned}
\delta_3(q_0, \langle M, x\rangle) &= \bigwedge_{i \in [k]} (i, q_0) \wedge \bigwedge_{d \in \mathsf{dom}(M)} \bigwedge_{C \sqsubseteq D \in \mathcal{T}} \big((0, q_{\mathsf{nnf}(\neg C),d}) \vee (0, q_{D,d})\big)\\
\delta_3(q_0, \bullet) &= \mathsf{true}\\
\delta_3(q_{A,d}, \langle M, x\rangle) &= \text{if } A(d) \in M, \text{ then } \mathsf{true} \text{ else } \mathsf{false}\\
\delta_3(q_{\neg A,d}, \langle M, x\rangle) &= \text{if } A(d) \notin M, \text{ then } \mathsf{true} \text{ else } \mathsf{false}\\
\delta_3(q_{C_1 \sqcup C_2,d}, \langle M, x\rangle) &= (0, q_{C_1,d}) \vee (0, q_{C_2,d})
\end{aligned}$$

$$\delta_3(q_{C_1 \sqcap C_2,d}, \langle M, x\rangle) = (0, q_{C_1,d}) \wedge (0, q_{C_2,d})$$

$$\delta_3(q_{(\sim n\, r\, D),d}, \langle M, x\rangle) = \big((0, F_{x,d}) \wedge (0, q^*_{(\sim n\, r\, D),d})\big) \vee \bigvee_{i \in [k]} (i, q_{(\sim n\, r\, D),d}) \wedge (i, q_d)$$

$$\delta_3(q_d, \langle M, x\rangle) = \text{if } d \in \mathsf{dom}(M)\text{, then } \mathsf{true} \text{ else } \mathsf{false}$$

$$\delta_3(F_{\perp,d}, \langle M, x\rangle) = \mathsf{true}$$

$$\delta_3(F_{r,d}, \langle M, x\rangle) = \begin{cases} (-1, F'_{r,d}) & \text{if } r \in \mathsf{N}_\mathsf{R}^{nt} \text{ or } (r \in \mathsf{N}_\mathsf{R}^{t} \text{ and } x = r) \\ \mathsf{false} & \text{otherwise} \end{cases}$$

$$\delta_3(F'_{r,d}, \langle M, x\rangle) = \begin{cases} \mathsf{true} & d \notin \mathsf{dom}(M) \text{ or } (r \in \mathsf{N}_\mathsf{R}^{t} \text{ and } x \notin \{\perp, r\}) \\ \mathsf{false} & \text{otherwise} \end{cases}$$

Intuitively, q_0 is used to verify that the TBox is *globally* satisfied. A state $q_{C,d}$ assigned to a node w is used as an obligation to verify that the element d satisfies the concept C. This can be done locally for Boolean concept constructors $\sqcap, \sqcup, \neg$, as implemented in the transitions above. For concepts of the form $(\sim n\ r\ D)$, we have to be more careful: the automaton has to move to the unique node w where $d \in F_r(w)$, identified using states $F_{r,d}$ and q_d (and the acceptance condition).

The transitions for number restrictions on non-transitive roles are defered to the appendix. For transitive roles, Lemma 4 provides the following observation: For counting the r-successors satisfying D of some $d \in \mathsf{dom}(\mathsf{bg}(w))$, it suffices to look at three "locations" in the tree decomposition: in the bag at w itself, along canonical paths satisfying **P1**, and along canonical paths satisfying **P2**. We next implement this strategy for at-least restrictions. In the following transitions, we assume that $\mathbf{a}_1, \ldots, \mathbf{a}_\ell$ are all r-clusters in M, and that $a_1, \ldots, a_\ell$ are representatives of each cluster. A partition $n_1 + \ldots + n_\ell = n$ *respects* M *relative to* d if $n_i = 0$ whenever $r(d, a_i) \notin M$; it *d-respects* M *relative to* d if $n_i = 0$ whenever $r(d, a_i) \notin M$ or $d \in \mathbf{a}_i$. Moreover, we define $M_r(d) = \{e \mid r(d, e), r(e, d) \in M\}$, and define transitions for (the complement of $F_{x,d}$) $\overline{F}_{x,d}$ similar to $F_{x,d}$.

$$\delta_3(q^*_{(\geq n\, r\, D),d}, \langle M, x\rangle) = \bigvee_{\substack{n_1+\ldots+n_\ell=n \\ \text{respects } M \text{ rel. to } d}} \ \bigwedge_{n_i \neq 0} (0, q^0_{(\geq n_i\, r\, D),a_i}) \vee (0, q^1_{(\geq n_i\, r\, D),a_i})$$

$$\delta_3(q^0_{(\geq n\, r\, D),d}, \langle M, x\rangle) = (0, F_{r,d}) \wedge (0, q^{\downarrow}_{(\geq n\, r\, D),d})$$

$$\delta_3(q^1_{(\geq n\, r\, D),d}, \langle M, x\rangle) = (0, \overline{F}_{r,d}) \wedge (-1, q^{\uparrow}_{(\geq n\, r\, D),d})$$

$$\delta_3(q^{\downarrow}_{(\geq n\, r\, D),d}, \langle M, x\rangle) = \bigvee_{n_0+n_1+\ldots+n_k=n} (0, p_{n_0,r,D,d}) \wedge \bigwedge_{n_i \neq 0} (i, p_{(\geq n_i\, r\, D),d})$$

$$\delta_3(p_{n,r,D,d}, \langle M, x\rangle) = \bigvee_{Y \subseteq M_r(d), |Y|=n} \Big(\bigwedge_{e \in Y} q_{D,e} \wedge \bigwedge_{y \in M_r(d) \setminus Y} q_{\sim D,e} \Big)$$

$$\delta_3(p_{(\geq n\, r\, D),d}, \bullet) = \text{if } n = 0\text{, then } \mathsf{true} \text{ else } \mathsf{false}$$

$$\delta_3(p_{(\geq n\, r\, D),d}, \langle M, x\rangle) = \begin{cases} \mathsf{false} & \text{if } x \neq r \text{ or } d \text{ not in root cluster} \\ \displaystyle\bigvee_{\substack{n_1+\ldots+n_\ell=n \\ \text{d-respects } M \text{ rel. to } d}} \ \bigwedge_{n_i \neq 0} (0, q^0_{(\geq n_i\, r\, D),a_i}) & \text{otherwise} \end{cases}$$

$$\delta_3(q^{\uparrow}_{(\geq n\, r\, D),d}, \langle M, x\rangle) = (0, q_d) \wedge \big((0, q^0_{(\geq n\, r\, D),d}) \vee (0, q^1_{(\geq n\, r\, D),d})\big)$$

Intuitively, the automaton non-deterministically guesses a partition $n_1 + \ldots + n_k$ of n and verifies that, starting from $\mathbf{a}_i$ at least n_i elements are reachable via r and satisfy D. For each such r-cluster, it proceeds either downwards (in states of the form $q^0_\cdot$ and $q^{\downarrow}_\cdot$) or looks for the world where the cluster $\mathbf{a}_i$ was a root (in states $q^1_\cdot$ and $q^{\uparrow}_\cdot$) and proceeds downwards from there on. In states $q^{\downarrow}_{(\geq n\, r\, D),d}$, the automaton again partitions n this time into $n_0, \ldots, n_k$; it then verifies that there are n_0 elements in the r-cluster of d satisfying D and, recursively, that via the i-th successor of the current node, there are n_i elements that are reachable via r and satisfy D. Using the parity condition, we make sure that states $q^{\downarrow}_{(\geq n\, r\, D),d}$ with $n \geq 1$ are not suspended forever, that is, eventualities are finally satisfied.

For the at-most restrictions, recall that $(\leq\ n\ r\ D)$ is equivalent to $\neg(\geq n{+}1\ r\ D)$; we can thus obtain the transitions for $q_{(\leq n\, r\, D),d}$ by "complementing" the transitions for $q_{(\geq n+1\, r\, D),d}$; details are given in the appendix.

In order to construct, for a given query q, an automaton $\mathfrak{A}_q$ which accepts a tree $(T, \tau) \in L(\mathfrak{A}_1)$ iff $\mathcal{I}_{T,\tau} \models q$, we adapt and extend ideas from [8] to canonical tree decompositions. The result is a nondeterministic parity tree automaton [13] of size exponential in q, and doubly exponential in $\mathcal{K}$. In contrast to [8], the query automaton depends on the KB. Indeed, for checking whether a fact $r(x, y)$ from the query is true (given some match candidate), it has to recall domain elements in the states; their number, however, is bounded by the width only. It remains to remark that the question of whether $L(\mathfrak{A}_\mathcal{K}) \setminus L(\mathfrak{A}_q)$ is empty can be decided in 3ExpTime, given the mentioned bounds on the sizes of the involved automata.

Theorem 9. *The certain answers problem for P2RPQs over $\mathcal{SQ}$-KBs is decidable in* 3ExpTime.

5 Discussion and Future Work

We have developed a tree-like decomposition for $\mathcal{SQ}$ which handles the interaction of number restrictions over transitive roles and enables the use of automata-based techniques for query answering. Our techniques yield a 3ExpTime upper bound, leaving an exponential gap to the known 2ExpTime lower bound, for answering positive existential queries over $\mathcal{ALC}$ KBs [8].

As immediate future work, we plan to close this gap, taking into account also other techniques for query answering such as rewriting [11]. Another relevant question is the precise *data complexity* – the present techniques give only exponential bounds, but we expect coNP-completeness. Moreover, we plan to extend our approach to nominals and inverses. We will also look at the problem of answering *conjunctive queries (CQs)* in $\mathcal{SQ}$; in general, the proposed automata-based approach yields the same upper bound for the problem of answering P2RPQs or CQs, but we expect it to be easier for CQs. Finally, we plan to see whether our techniques extend to the query containment problem, and develop techniques for *finite* query answering in (extensions of) $\mathcal{SQ}$.

References

1. Amarilli, A., Benedikt, M., Bourhis, P., Vanden Boom, M.: Query answering with transitive and linear-ordered data. In: Proc. of IJCAI-16. pp. 893–899 (2016)
2. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2003)
3. Baader, F., Lutz, C., Sturm, H., Wolter, F.: Fusions of description logics and abstract description systems. J. Artif. Intell. Res. (JAIR) 16, 1–58 (2002)
4. Baget, J., Bienvenu, M., Mugnier, M., Rocher, S.: Combining existential rules and transitivity: Next steps. In: Proc. of IJCAI-15. pp. 2720–2726 (2015)
5. Benedikt, M., Bourhis, P., Vanden Boom, M.: A step up in expressiveness of decidable fixpoint logics. In: Proc. of LICS-16. pp. 817–826 (2016)
6. Bienvenu, M., Ortiz, M., Simkus, M.: Regular path queries in lightweight description logics: Complexity and algorithms. J. Artif. Intell. Res. (JAIR) 53, 315–374 (2015)
7. Calvanese, D., Eiter, T., Ortiz, M.: Regular path queries in expressive description logics with nominals. In: Proc. of IJCAI-09. pp. 714–720 (2009)
8. Calvanese, D., Eiter, T., Ortiz, M.: Answering regular path queries in expressive description logics via alternating tree-automata. Inf. Comput. 237, 12–55 (2014), http://dx.doi.org/10.1016/j.ic.2014.04.002
9. Eiter, T., Lutz, C., Ortiz, M., Simkus, M.: Query answering in description logics with transitive roles. In: Proc. of IJCAI-09. pp. 759–764 (2009)
10. Glimm, B., Horrocks, I., Sattler, U.: Unions of conjunctive queries in SHOQ. In: Proc. of KR-08. pp. 252–262 (2008)
11. Glimm, B., Lutz, C., Horrocks, I., Sattler, U.: Conjunctive query answering for the description logic SHIQ. J. Artif. Intell. Res. (JAIR) 31, 157–204 (2008)
12. Gottlob, G., Pieris, A., Tendera, L.: Querying the guarded fragment with transitivity. In: Proc. of ICALP-13. pp. 287–298 (2013)
13. Grädel, E., Thomas, W., Wilke, T. (eds.): Automata, Logics, and Infinite Games: A Guide to Current Research, LNCS, vol. 2500. Springer
14. Grädel, E., Walukiewicz, I.: Guarded fixed point logic. In: Proc. of LICS-99. pp. 45–54 (1999)
15. Gutiérrez-Basulto, V., Ibáñez-García, Y., Jung, J.C.: Number restrictions on transitive roles in description logics with nominals. In: Proc. of AAAI-17 (2017)
16. Hollunder, B., Baader, F.: Qualifying number restrictions in concept languages. In: Proc. of KR-91. pp. 335–346 (1991)
17. Horrocks, I., Sattler, U., Tobies, S.: Practical reasoning for very expressive description logics. Logic Journal of the IGPL 8(3), 239–263 (2000)
18. Kaminski, M., Smolka, G.: Terminating tableaux for $\mathcal{SOQ}$ with number restrictions on transitive roles. In: Proc. of the 6th IFIP TC. pp. 213–228 (2010)
19. Kazakov, Y., Pratt-Hartmann, I.: A note on the complexity of the satisfiability problem for graded modal logics. In: Proc. of LICS-09. pp. 407–416 (2009)
20. Kazakov, Y., Sattler, U., Zolin, E.: How many legs do I have? non-simple roles in number restrictions revisited. In: Proc. of LPAR-07. pp. 303–317 (2007)
21. Pratt-Hartmann, I.: The two-variable fragment with counting and equivalence. Math. Log. Q. 61(6), 474–515 (2015)
22. Rector, A.L., Rogers, J.: Ontological and practical issues in using a description logic to represent medical concept systems: Experience from GALEN. In: In Proc. of RW-06. pp. 197–231 (2006)

23. Schröder, L., Pattinson, D.: How many toes do I have? parthood and number restrictions in description logics. In: Proc. of KR-08. pp. 307–317 (2008)
24. Stevens, R., Aranguren, M.E., Wolstencroft, K., Sattler, U., Drummond, N., Horridge, M., Rector, A.L.: Using OWL to model biological knowledge. International Journal of Man-Machine Studies 65(7), 583–594 (2007)
25. Tendera, L.: Counting in the two variable guarded logic with transitivity. In: Proc. of STACS-05. pp. 83–96 (2005)
26. Vardi, M.Y.: Reasoning about the past with two-way automata. In: Proc. ICALP-98. pp. 628–641 (1998)
27. Wolstencroft, K., Brass, A., Horrocks, I., Lord, P., Sattler, U., Turi, D., Stevens, R.: A little semantic web goes a long way in biology. In: Proc. of ISWC-05 (2005)

APPENDIX

Additional Preliminaries

The semantics of P2RPQs is defined as follows. Let $\mathcal{I}$ be an interpretation and $\mathcal{E}$ a regular expression over $\mathcal{S}$. We write $(d,e) \in \mathcal{E}^\mathcal{I}$, if there is a word $w_1 \ldots w_n \in L(\mathcal{E})$ and a sequence $d_0, \ldots, d_n \in \Delta^\mathcal{I}$ such that $d_0 = d, d_n = e$, and for all $i \in [1,n]$ we have that *(i)* if $w_i = A?$, then $d_{i-1} = d_i \in A^\mathcal{I}$, and *(ii)* if $w_i = r \in \mathsf{N}_\mathsf{R}^\pm$, then $(d_{i-1}, d_i) \in r^\mathcal{I}$.

Let $q(\boldsymbol{x}) = \exists \boldsymbol{y}.\varphi$ be a P2RPQ with $\boldsymbol{x} = (x_1, \ldots, x_k)$, $\mathcal{I}$ an interpretation. A map π from the set of terms in q to $\Delta^\mathcal{I}$ is a *match for q in $\mathcal{I}$* if *(i)* $\pi(c) = c$ for all individual names c in q, *(ii)* $\pi(t) \in A^\mathcal{I}$ for every $A(t) \in q$, and *(iii)* $(\pi(t), \pi(t')) \in \mathcal{E}^\mathcal{I}$, for every $\mathcal{E}(t,t') \in q$. A tuple $\boldsymbol{c} = (c_1, \ldots, c_k) \in \mathsf{ind}(\mathcal{A})^k$ is an *answer to q in $\mathcal{I}$*, written $\mathcal{I} \models q(\boldsymbol{c})$, if there is a match π for q in $\mathcal{I}$ with $\pi(x_i) = c_i$, for all $i \in [1,k]$. Moreover, $\boldsymbol{c}$ is a *certain answer to q over a KB $\mathcal{K}$*, written $\mathcal{K} \models q(\boldsymbol{c})$, if $\mathcal{I} \models q(\boldsymbol{c})$ for every model $\mathcal{I}$ of $\mathcal{K}$. The set of all certain answers to q over $\mathcal{K}$ is denoted $\mathsf{cert}(q, \mathcal{K})$.

Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be two interpretations. A *homomorphism from $\mathcal{I}_1$ to $\mathcal{I}_2$* is a function $h : \Delta^{\mathcal{I}_1} \to \Delta^{\mathcal{I}_2}$ such that *(i)* $h(a) = a$ for all $a \in \mathsf{N}_\mathsf{I}$, *(ii)* if $d \in A^{\mathcal{I}_1}$, then $h(d) \in A^{\mathcal{I}_2}$, for all $A \in \mathsf{N}_\mathsf{C}$, and *(iii)* if $(d,e) \in r^{\mathcal{I}_1}$, then $(h(d), h(e)) \in r^{\mathcal{I}_2}$, for all $r \in \mathsf{N}_\mathsf{R}$. It is folklore that P2RQPs are *preserved under homomorphisms*, that is, if $\mathcal{I}_1 \models q(\boldsymbol{c})$ and there is a homomorphism from $\mathcal{I}_1$ to $\mathcal{I}_2$, then $\mathcal{I}_2 \models q(\boldsymbol{c})$.

Proofs for Section 3

Before we can prove Lemma 4, we prove an auxiliary lemma. Let us fix an interpretation $\mathcal{J} = (\Delta^\mathcal{J}, \cdot^\mathcal{J})$ and let $\mathfrak{T} = (T, \mathsf{bg}, \mathsf{rl})$ be a canonical decomposition of $\mathcal{I}$. Observe that as a consequence of Definition 3, particularly, Condition (C2), for every $d \in \Delta^\mathcal{J}$, $r \in \mathsf{N}_\mathsf{R}^t$, there is a unique node $w \in T$ with $\mathsf{rl}(w) = r$ and $d \in F_r(w)$. We denote this node with $w_{d,r}$.

Lemma 10. *Let $r \in \mathsf{N}_\mathsf{R}^t$. For every $u \in T$ with $\mathsf{rl}(u) = r$ and $r(d,e) \in \mathsf{bg}(u)$, exactly one of the following holds:*

- *$w_{d,r} = w_{e,r}$ and $r(d,e) \in \mathsf{bg}(w_{d,r})$;*
- *$w_{e,r}$ is a successor of $w_{d,r}$, $r(d,e) \in \mathsf{bg}(w_{e,r})$ and d belongs to an r-root cluster in $w_{e,r}$;*
- *$w_{e,r}$ is an ancestor of $w_{d,r}$ and $r(d,e) \in \mathsf{bg}(w_{d,r})$.*

Proof. Since $d, e \in \mathsf{dom}(\mathsf{bg}(u))$, we know that $w_{d,r}$ and $w_{e,r}$ are either equal to u or ancestors of u. We distinguish three cases:

- If $w_{d,r} = w_{e,r}$, then $d, e \in \mathsf{dom}(\mathsf{bg}(v'))$ for every v' on the path from $w_{d,r}$ to u, by Definition 2 *(iv)*. By Definition 2 *(i)*, we obtain that $r(d,e) \in \mathsf{bg}(v')$ for every such v', hence $r(d,e) \in \mathsf{bg}(w_{d,r})$.

– If $w_{d,r}$ is an ancestor of $w_{e,r}$, then we know by the same reasoning as in the previous point that $r(d,e) \in \mathsf{bg}(w_{e,r})$. Let w' be the predecessor of $w_{e,r}$. Since $\mathfrak{T}$ is a canonical decomposition, either (C2) or (C3) applies to w' and $w_{e,r}$. Assume first that $w' = w_{d,r}$.
  - In case of (C2), since $d \in \mathsf{dom}(\mathsf{bg}(w_{d,r})) \cap \mathsf{dom}(\mathsf{bg}(w_{e,r}))$ we know that there is a r-root cluster $\mathbf{a} \subseteq \mathsf{dom}(\mathsf{bg}(w_{e,r}))$ such that $d \in \mathbf{a}$.
  - In case of (C3), let $\mathbf{a} \subseteq F_r(w')$ be the cluster witnessing this. By Item *(b)*, $\mathbf{a}$ is an r-root cluster in $\mathsf{bg}(w_{e,r})$. By definition, we know $e \in F_r(w_{e,r})$ and thus $e \in F(w_{e,r})$. From this and Item *(c)*, we obtain that $d \in \mathbf{a} \cup F(w_{e,r})$, and since $d \notin F(w_{e,r})$ we know $d \in \mathbf{a}$.

  Thus, in both cases, we are in the second case of the lemma. Assume now that $w' \neq w_{d,r}$. We show that it leads to a contradiction in both cases:
  - In case of (C2), since $d \in \mathsf{dom}(\mathsf{bg}(w_{e,r})) \cap \mathsf{dom}(\mathsf{bg}(w'))$ we know $d \in F(w')$. On the other hand, $d \in F_r(w_{d,r})$ implies that either $d \in F(w_{d,r})$ or $w_{d,r}$ has a predecessor w'' such that $d \in F(w'')$. This is a contradiction since $w' \neq w''$ since $w_{d,r}$ is an ancestor of $w_{e,r}$.
  - In case of (C3), let $\mathbf{a}$ be the r-cluster witnessing this. By (C3) and the Definition of F_r, $e \in F_r(w_{e,r})$, implies $e \in F(w_{e,r})$. Since $w_{d,r} \neq w'$ but $w_{d,r}$ is an ancestor of $w_{e,r}$, we know that $w' \neq \varepsilon$ and $\mathsf{rl}(w') = r$. From Item *(d)* we obtain that $d \in \mathbf{a} \cup F(w_{e,r})$, and since $d \notin F(w_{e,r})$, we know $d \in \mathbf{a}$. By Item *(a)*, we know that $\mathbf{a} \subseteq F_r(w')$, but then $w' = w_{d,r}$, contradiction.

– If $w_{e,r}$ is an ancestor of $w_{d,r}$, then we know by the reasoning in the first point that $r(d,e) \in \mathsf{bg}(w_{d,r})$; thus, we are in the last case of the lemma.

□

In the proof of the following lemma, we denote with $\mathsf{pre}(w)$ the predecessor of w in the underlying tree.

Lemma 4. *Let $\mathfrak{T}$ be a canonical decomposition of $\mathcal{I}$, r a transitive role and $d, e \in \Delta^{\mathcal{I}}$ with $(d,e) \in r^{\mathcal{I}}$. Then there is a unique canonical r-path (up to equivalence) from d to e in $\mathfrak{T}$.*

Proof. If $(d,e) \in r^{\mathcal{J}}$, then by Item *(iii)* in the definition of a tree decomposition, there is an r-path $d_0, w_0, \ldots, w_{n-1}, d_n$ from d to e in $(T, \mathsf{bg}, \mathsf{rl})$. We show first that it is without loss of generality to assume that for all $0 \leq j < n-1$, we have:

(a) $w_{j+1} = w_j$,
(b) w_{j+1} is a successor of w_j, $d_{j+1} \in F_r(w_j)$ and d_{j+1} is in an r-root cluster of $\mathsf{bg}(w_{j+1})$
(c) w_{j+1} is an ancestor of w_j and $d_{j+1} \in F_r(w_{j+1})$, or
(d) $\mathsf{pre}(w_{j+1})$ is an ancestor of w_j, $d_j \in F_r(w_j)$, and $d_{j+1} \in F_r(\mathsf{pre}(w_{j+1}))$.

Observe that, by Lemma 10, we can assume that $w_i \in \{w_{d_i,r}, w_{d_{i+1},r}\}$. Moreover, if $w_i = w_{d_i,r}$, then either $w_{d_i,r} = w_{d_{i+1},r}$ or $w_{d_{i+1},r}$ is an ancestor of $w_{d_i,r}$ and $d_{i+1} \in F_r(w_{j+1})$; if $w_i = w_{d_{i+1},r}$, then $w_{d_{i+1},r}$ is a successor of $w_{d_i,r}$ and $d_{i+1} \in F_r(w_i)$. Let now be $0 \leq j < n-1$. We distinguish four cases:

– If $w_j = w_{d_j,r}$ and $w_{j+1} = w_{d_{j+1},r}$, then Case (a) or Case (c) applies.
– If $w_j = w_{d_{j+1},r}$ and $w_{j+1} = w_{d_{j+1},r}$, then Case (a) applies.
– If $w_j = w_{d_j,r}$ and $w_{j+1} = w_{d_{j+2},r}$, then Case (b) or (d) applies.
– If $w_j = w_{d_{j+1},r}$ and $w_{j+1} = w_{d_{j+2},r}$, then Case (b) applies.

Note then, that in case (a) is satisfied for some j, we can safely drop d_{j+1} and w_j and the remaining sequence is still an r-path, due to Item (iv) of the definition of tree decomposition. So from now on, we assume that for all $0 \leq j < n-1$, one of (b)–(d) is the case.

If Condition (b) applies for all j, then the r-path is downward, thus by the third item in Lemma 10, canonical. Otherwise, we modify the sequence by performing the following operation exhaustively. Let $0 \leq k < n-1$ be some index satisfying (c), that is, w_{k+1} is an ancestor of w_k, and let k' be minimal such that all i with $k' \leq i < k$ satisfy (b). If $k' = k$, then do nothing, otherwise we distinguish the following cases:

Case 1: $w_{k+1} = w_j$ for some $k' \leq j < k$. We show inductively that then $r(d_i, d_{k+1}) \in \mathsf{bg}(w_i)$, for all $j \leq i \leq k$. For $i = k$ it is clear by assumption. For the inductive step, assume $j \leq i < k$. Clearly, we have $r(d_i, d_{i+1}) \in \mathsf{bg}(w_i)$ and, by the choice of k and the assumption $w_j = w_k$, also $r(d_{k+1}, d_{k+2}) \in \mathsf{bg}(w_j)$. Moreover, by induction, we can assume that $r(d_{i+1}, d_{k+1}) \in \mathsf{bg}(w_{i+1})$. By Item (v) of the definition of tree decomposition, we know that $d_{k+1} \in \mathsf{dom}(\mathsf{bg}(w_i))$. By (C1), we know that $r(d_{i+1}, d_{k+1}) \in \mathsf{bg}(w_i)$. Item (iv) of the definition of tree decomposition yields also $r(d_i, d_{k+1}) \in \mathsf{bg}(w_i)$, thus finishing the inductive step.
This implies $r(d_j, d_{k+1}) \in \mathsf{bg}(w_j)$. Since also $r(d_{k+1}, d_{k+2}) \in \mathsf{bg}(w_j)$, we know $r(d_j, d_{k+2}) \in \mathsf{bg}(w_j)$. Thus, dropping the subsequence

$$d_{j+1}, w_{j+1}, \ldots, w_{k+1}$$

yields an r-path satisfying (b)–(d) for all j.

Case 2: w_{k+1} is an ancestor of $w_{k'}$. We can argue as in Case 1 that $r(d_{k'}, d_{k+1}) \in \mathsf{bg}(w_{k'})$. Thus, we can drop the subsequence $d_{k'+1}, \ldots, d_k, w_k$ obtaining an r-path which satisfies (b)–(d), for all j.

We can deal similarly with an index satisfying (d). After performing this step exhaustively, we obtain an r-path $e_0, v_0, \ldots, v_{m-1}, e_m$ from d to e which is downward, and thus already canonical, or

($*$) there is some $0 \leq j < m$ such that (c) holds for all $0 \leq i < j$, and if $j < m-1$, then (d) holds for j, and (b) holds for all $j < i < m$.

In case of ($*$), we show how to obtain an r-path satisfying ($*$) with $j = 0$.

Claim. If $j \geq 1$, then $r(e_0, e_2) \in \mathsf{bg}(v_0)$.

Proof of the Claim. We show inductively that $r(e_1, e_2) \in \mathsf{bg}(u)$ for all u on the path between v_1 and v_0. It is obviously true for $u = v_1$.

Let now u be the successor of some u_0 on the path from v_1 to v_0, and assume by induction that u_0 satisfies $r(e_1, e_2) \in \mathsf{dom}(\mathsf{bg}(u_0))$. Suppose that (C3) holds for u. Then $\mathsf{rl}(u) \neq r$. But since $\mathsf{rl}(w_0) = r$, we know that (C3) holds again for some node between u and v_0. The only possible witness for this is $e = e_2$. However, this leads to a contradiction as well, because $e_2 \notin F(w)$ for any w on the path between u and v_0. Hence, we know that (C4) holds for u. Let $\mathbf{a}$ be the r-cluster in $\mathsf{bg}(u)$ witnessing this.

- If $e_2 \in \mathbf{a}$, (C4) (b) implies that $e_2 \in \mathsf{dom}(\mathsf{bg}(u))$, since $r(e_1, e_2) \in \mathsf{bg}(u_0)$.
- If $e_2 \notin \mathbf{a}$, then we know by (C4) that $r(e, e_2) \in \mathsf{bg}(u)$, for some $e \in \mathbf{a}$. Thus, $r(e, e_2) \in \mathsf{bg}(u_0)$. Again, (C4) (b) implies that $e_2 \in \mathsf{dom}(\mathsf{bg}(u))$.

By (C1), we obtain in both cases $r(e_1, e_2) \in \mathsf{bg}(u)$, thus finishing the induction. Since also $r(e_0, e_1) \in \mathsf{bg}(v_0)$, we obtain $r(e_0, e_2) \in \mathsf{bg}(w_0)$. This finishes the proof of the Claim.

It is now easy to verify that dropping e_1, v_1 from the sequence preserves $(*)$, but with j and m decreased by one. By the Claim, we can perform this operation repeatedly until $j = 0$.

We argue that the remaining r-path is canonical.

- If $m = 1$, we distinguish cases according to Lemma 10:
  - if $w_{e_0,r} = w_{e_1,r}$, then $e_0, w_{e_1,r}, e_1$ is a downward path from d to e;
  - if $w_{e_1,r}$ is a successor of $w_{e_0,r}$, then $e_0, w_{e_1,r}, e_1$ is a downward path from d to e;
  - if $w_{e_1,r}$ is an ancestor of $w_{e_0,r}$, then $e_0, w_{e_0,r}, e_1$ is an r-path from d to e satisfying **P2**.

Thus, in all cases, we obtain a canonical r-path from d to e. In case $m > 2$, the resulting path satisfies **P2** because of $(*)$, in particular, (d) holds for 0 and (b) holds for all $0 < j < m$.

It remains to show that this path is unique up to equivalence. Let

$$\begin{aligned}\pi_1 =& d_0, w_0, \ldots, d_{n-1}, w_{n-1}, d_n \text{ and} \\ \pi_2 =& e_0, v_0, \ldots, e_{m-1}, v_{m-1}, e_m\end{aligned}$$

be two canonical paths from d to e in $\mathfrak{T}$. We distinguish the following cases:

- **P1** holds for both π_1 and π_2. Since $d_n \in F_r(w_{n-1})$, $e_n \in F_r(v_{m-1})$, $r(d_{n-1}, d_n) \in \mathsf{bg}(w_{n-1})$, $r(e_{m-1}, e_m) \in \mathsf{bg}(v_{m-1})$ and $d_n = e_m$ we know $w_{n-1} = w_{d_n,r} = w_{e_m,r} = v_{m-1}$. And because both π_1 and π_2 are downward, we can conclude $n = m$ since T is a tree, $w_i = v_i$ for $0 \leq i < n-1$. If $n = 1$, clearly π_1 and π_2 are equivalent. Otherwise, since d_{i+1} and e_{i+1} are both in an r-root cluster in w_{i+1} we have that either $d_{i+1} = e_{i+1}$ or $r(d_{i+1}, e_{i+1}), r(e_{i+1}, d_{i+1}) \in \mathsf{bg}(w_{i+1})$. That is, they are in the same r-cluster in $\mathsf{bg}(w_{i+1})$, and thus π_1 and π_2 are equivalent.
- **P2** holds for both π_1 and π_2. Since $d_0 \in F_r(w_0)$, $e_0 \in F_r(v_0)$, $r(d_0, d_1) \in \mathsf{bg}(w_0)$, $r(e_0, e_1) \in \mathsf{bg}(v_0)$ and $d_0 = e_0$, we have that $w_0 = w_{d_0,r} = w_{e_0,r} = v_0$. Further, using the same arguments as in the previous case we can conclude that $w_i = v_i$, and d_i and e_i are in the same r-cluster in $\mathsf{bg}(w_i)$, for $1 \leq i \leq n_1$; and thus, $n = m$. Therefore, π_1 and π_2 are equivalent.

– **P1** holds for π_1 and **P2** holds for π_2. In this case, we derive a contradiction. Since $d_n = e_m$, and $d_n \in F_r(w_{n-1})$ and $e_n \in F_r(v_{m-1})$, we can conclude as in the previous cases that $w_{n-1} = w_{d_{n-1},r} = w_{e_m,r} = v_{m-1}$. Since π_1 and $e_1, v_1, \ldots, e_{m-1}, v_{m-1}, e_m$ are downward paths, we have that $m = n+1$ and form this we obtain that $w_i = v_{i+1}$ for all $0 \leq i \leq n-1$. Moreover, d_i and e_{i+1} are in the same r-cluster in $\mathsf{bg}(w_i)$, for all $1 \leq i \leq n-1$. In particular $w_0 = v_1$, and thus $d_0 \in \mathsf{dom}(\mathsf{bg}(v_1))$. Since $d_0 = e_0 \in \mathsf{dom}(v_0)$, and $\mathsf{pre}(v_1)$ is an ancestor of v_0 we know that $d_0 \notin F(v_0)$. Further, by condition *(v)* of the definition of tree decomposition, we have that $d_0, e_1 \in \mathsf{dom}(\mathsf{bg}(w))$, for every node w in the path from $\mathsf{pre}(v_1)$ to v_0 and by (C1) $r(d_0, e_1) \in \mathsf{bg}(w)$. From this, we obtain that that (C4) applies to every such w, and therefore $\mathsf{rl}(w) = r$. On the other hand, by **P2** we have that $d_0 \in F_r(v_0)$. Since $d_0 \notin F(v_0)$, we know $\mathsf{pre}(v_0) \neq \varepsilon$ and $\mathsf{rl}(\mathsf{pre}(v_0)) \neq r$, a contradiction.

□

Lemma 6. *For each $\mathcal{I} \models \mathcal{K}$, there is a sub-interpretation $\mathcal{I}'$ of $\mathcal{I}$ with $\mathcal{I}' \models \mathcal{K}$ and width and breadth of $\mathcal{I}'$ are bounded by $O(|\mathcal{A}| + 2^{\mathsf{poly}(|\mathcal{T}|)})$.*

Proof. We show the lemma in two stages, adapting a technique from [19, 15]. Let $\hat{m}$ be the maximal number appearing in $\mathcal{T}$.

Stage 1 (Bounded breadth). As it is standard to achieve bounded breadth for non-transitive roles [11], we only deal with transitive roles here.

An element e is an *strict r-successor of d* if $(d, e) \in r^{\mathcal{I}}$, but $e \notin Q_{\mathcal{I},r}(d)$. Let $W_r(d)$ be the set of strict r-successors of d and $W_r(d, C) \subseteq W_r(d)$ be the set of all strict r-successors of d satisfying C. Then, fix a subset $W'_r(d) \subseteq W_r(d)$ by adding, for each $C \in \mathsf{cl}(\mathcal{T})$, $\min(\hat{m}, |W_r(d, C)|)$ elements from $W_r(d, C)$.

Assume without loss of generality that $W'_r(d_1) = W'_r(d_2)$ if $d_1 \in Q_{\mathcal{I},r}(d_2)$, and define relations S^1_r, S^2_r, and S^3_r, for each $r \in \mathsf{Rol}_t(\mathcal{K})$, as follows:

$$S^1_r = \{(d, d') \in r^{\mathcal{I}} \mid d' \in Q_{\mathcal{I},r}(d)\};$$
$$S^2_r = \{(d, d') \in r^{\mathcal{I}} \mid r(d, d') \in \mathcal{A}\};$$
$$S^3_r = \{(d, d') \in r^{\mathcal{I}} \mid d' \in W'_r(d)\}.$$

Intuitively, S^1_r is the restriction of $r^{\mathcal{I}}$ to the clusters, S^2_r takes care of the ABox, and and S^3_r keeps a sufficient set of successors to witness all at-least restrictions.

Finally, obtain $\mathcal{I}'$ from $\mathcal{I}$ by taking $\Delta^{\mathcal{I}'} = \Delta^{\mathcal{I}}$, $A^{\mathcal{I}'} = A^{\mathcal{I}}$ for all concept names A, $r^{\mathcal{I}'} = r^{\mathcal{I}}$, for all non-transitive roles r, and, for all transitive roles r,

$$r^{\mathcal{I}'} = (S^1_r \cup S^2_r \cup S^3_r)^+.$$

Claim 1. $C^{\mathcal{I}} = C^{\mathcal{I}'}$, for all $C \in \mathsf{cl}(\mathcal{T})$.

Proof of Claim 1. This is shown by induction on the structure of concepts. The only non-trivial case are concepts $C = (\leq n\ r\ D)$, r transitive. Clearly, $d \in C^{\mathcal{I}}$ implies $d \in C^{\mathcal{I}'}$ since $r^{\mathcal{I}'} \subseteq r^{\mathcal{I}}$. The converse is a direct consequence of the definition of $W'_r(d)$ and S^3_r. In particular, only r-successors that "cannot be seen" by at-most restrictions (due to the choice of $\hat{m}$) are removed.

From Claim 1, we conclude that $\mathcal{I}' \models \mathcal{T}$; by Claim 1 and the definition of $r^{\mathcal{I}'}$, particularly S_r^2, we also have $\mathcal{I}' \models \mathcal{A}$, thus $\mathcal{I} \models \mathcal{K}$. Since $r^{\mathcal{I}'} \subseteq r^{\mathcal{I}}$ and $A^{\mathcal{I}} = A^{\mathcal{I}'}$, for all $A \in \mathsf{N_C}$, the identity is an homomorphism from $\mathcal{I}'$ to $\mathcal{I}$. Finally note that, by construction, the breadth of $\mathcal{I}'$ is at most $|\mathcal{A}| + |\mathsf{cl}(\mathcal{T})| \cdot \hat{m}$ and thus $O(|\mathcal{A}| + 2^{\mathsf{poly}(|\mathcal{T}|)})$.

Stage 2 (Bounded Width). For every transitive role r, and every $d \in \Delta^{\mathcal{I}}$, fix a set $W_r(d) \subseteq Q_{\mathcal{I},r}(d)$ as follows. For each $C \in \mathsf{cl}(\mathcal{T})$, $W_r(d)$ contains the set $Q_{\mathcal{I},r}(d) \cap C^{\mathcal{I}}$ if this set has size at most $\hat{m}$, and otherwise a subset thereof having size $\hat{m}$. Without loss of generality, we assume that $W_r(d) = W_r(e)$ for all $e \in Q_{\mathcal{I},r}(d)$. Now, define a set Δ_r, for each transitive r, by taking

$$\Delta_r = \mathsf{ind}(\mathcal{A}) \cup \bigcup_{d \in \Delta^{\mathcal{I}}} W_r(d),$$

and define an interpretation $\mathcal{I}' = (\Delta^{\mathcal{I}'}, \cdot^{\mathcal{I}'})$ by setting $\Delta^{\mathcal{I}'} = \Delta^{\mathcal{I}}$, $A^{\mathcal{I}'} = A^{\mathcal{I}}$, for all $A \in \mathsf{N_C}$, $r^{\mathcal{I}'} = r^{\mathcal{I}}$, for all non-transitive roles r, and

$$r^{\mathcal{I}'} = r^{\mathcal{I}} \cap (\Delta^{\mathcal{I}} \times \Delta_r), \text{ for all transitive roles } r.$$

It is not hard to verify that $r^{\mathcal{I}'}$ is indeed transitive.

Claim 2. $C^{\mathcal{I}} = C^{\mathcal{I}'}$, for all $C \in \mathsf{cl}(\mathcal{T})$.

Proof of Claim 2. This is again shown by induction on the structure of concepts. The only non-trivial case are concepts $C = (\leq n\ r\ D)$, r transitive. Clearly, $d \in C^{\mathcal{I}}$ implies $d \in C^{\mathcal{I}'}$ since $r^{\mathcal{I}'} \subseteq r^{\mathcal{I}}$. The converse is a direct consequence of the definition of $W_r(d)$, in particular the choice of $\hat{m}$, and the definition of $r^{\mathcal{I}'}$. In particular, we remove only r-successors that cannot contribute to at-least restrictions.

Based on Claim 2, it is easy to see that $\mathcal{I}' \models \mathcal{T}$ and $\mathcal{I}' \models \mathcal{A}$. Moreover, the identity is a homomorphism from $\mathcal{I}'$ to $\mathcal{I}$. Finally, by definition of $W_r(d)$, particularly the choice of $\hat{m}$, it should be clear that the width of $\mathcal{I}'$ is bounded by $|\mathcal{A}| + 2^{\mathsf{poly}(|\mathcal{T}|)}$. ◻

We next verify two properties of the witness set $\mathsf{Wit}_{\mathcal{I},r}(d)$, which are needed later on.

Lemma 11. *For every interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$, we have:*

1. $\mathsf{Wit}_{\mathcal{I},r}(d) \subseteq \mathsf{ind}(\mathcal{A}) \cup X$ *for some set X of size bounded by* $2^{\mathsf{poly}(|\mathcal{T}|)}$*;*
2. *for all $e \in \mathsf{Wit}_{\mathcal{I},r}(d)$, we have that* $\mathsf{Wit}_{\mathcal{I},r}(e) \subseteq \mathsf{Wit}_{\mathcal{I},r}(d)$*.*

Proof. Throughout the proof, we denote with $W^{\leadsto}_{\mathcal{I},r}(d)$ the set $\{e \mid d \leadsto^*_{\mathcal{I},r} e\}$.

Item 1. Construct a tree (T, E) by starting with $T = \{d\}$ and exhaustively performing the following operation:

(∗) Choose a leaf $e \in T$ and add, for all $f \in \Delta^{\mathcal{I}} \setminus T$ with $e \leadsto_{\mathcal{I},r} f$, f to T and (e, f) to E.

By definition of $\rightsquigarrow_{\mathcal{I},r}$ and $(*)$, the obtained graph is indeed a tree which additionally satisfies $W^{\rightsquigarrow}_{\mathcal{I},r}(d) \subseteq T$. Now, consider the labelling $\ell : T \to 2^{\mathsf{cl}(\mathcal{T})}$ given by

$$\ell(e) = \{C \mid e \in (\leq n\ r\ C)^{\mathcal{I}}, (\leq n\ r\ C) \in \mathsf{cl}(\mathcal{T})\}.$$

By construction of (T, E), we know that $(e, f) \in E$ implies

- $\ell(e) \subseteq \ell(f)$ for all $e, f \in T$ such that f is a leaf in (T, E);
- $\ell(e) \subsetneq \ell(f)$ for all $e, f \in T$ such that f is not a leaf in (T, E).

Thus, the depth of (T, E) is bounded by $|\mathcal{T}|$. Since, for any e, there are at most exponentially (in $\mathcal{T}$) many f such that $e \rightsquigarrow_{\mathcal{I},r} f$, we know that the outdegree of (T, E) is bounded exponentially in $\mathcal{T}$. Overall, we get that the size of T, and thus of the set $W^{\rightsquigarrow}_{\mathcal{I},r}(d)$ is bounded by an exponential in $\mathcal{T}$. Note next that, by Lemma 6, $Q_{\mathcal{I},r}(d) \subseteq \mathsf{ind}(\mathcal{A}) \cup X_d$, for some set X_d of size bounded by $2^{\mathsf{poly}(|\mathcal{T}|)}$. Overall, this implies the statement in the lemma.

Item 2. Let $e \in \mathsf{Wit}_{\mathcal{I},r}(d)$. By definition of $\mathsf{Wit}_{\mathcal{I},r}$, it suffices to show that $W^{\rightsquigarrow}_{\mathcal{I},r}(e) \subseteq \mathsf{Wit}_{\mathcal{I},r}(d)$. To this end, suppose $f \in W^{\rightsquigarrow}_{\mathcal{I},r}(e)$. By definition of $W^{\rightsquigarrow}_{\mathcal{I},r}$, there is a sequence $e_1 \rightsquigarrow_{\mathcal{I},r} \cdots \rightsquigarrow_{\mathcal{I},r} e_n$ with $e = e_1$ and $f = e_n$ (possibly $n = 1$). As $e \in \mathsf{Wit}_{\mathcal{I},r}(d)$, we have either *(i)* $e \in W^{\rightsquigarrow}_{\mathcal{I},r}(d)$ or *(ii)* there is some $e' \in W^{\rightsquigarrow}_{\mathcal{I},r}(d)$ such that $e \in Q_{\mathcal{I},r}(e')$. We distinguish cases.

(i) $e \in W^{\rightsquigarrow}_{\mathcal{I},r}(d)$ implies that there is a sequence $d_1 \rightsquigarrow_{\mathcal{I},r} \cdots \rightsquigarrow_{\mathcal{I},r} d_m$ with $d_1 = d$ and $d_m = e$. Thus, there is a sequence $d_1 \rightsquigarrow_{\mathcal{I},r} \cdots \rightsquigarrow_{\mathcal{I},r} d_m = e = e_1 \rightsquigarrow_{\mathcal{I},r} \cdots \rightsquigarrow_{\mathcal{I},r} e_n = f$. Hence, $f \in W^{\rightsquigarrow}_{\mathcal{I},r}(d) \subseteq \mathsf{Wit}_{\mathcal{I},r}(d)$.

(ii) Similar to Case *(i)*, there is a sequence $d_1 \rightsquigarrow_{\mathcal{I},r} \cdots \rightsquigarrow_{\mathcal{I},r} d_m$ with $d_1 = d$ and $d_m = e'$. If $e = f$, that is $n = 1$ in the sequence above, we know that $f \in Q_{\mathcal{I},r}(e')$ and thus $f \in \mathsf{Wit}_{\mathcal{I},r}(d)$. Otherwise, observe that we can assume that $|Q_{\mathcal{I},r}(d)| \geq 2$ (otherwise $e' = e$ and we are in Case *(i)*). Thus, we have $e' \in (\sim \ell\ r\ C)^{\mathcal{I}}$ iff $e \in (\sim \ell\ r\ C)^{\mathcal{I}}$, for all $\sim$, ℓ, and C, and hence also $e' \rightsquigarrow_{\mathcal{I},r} e_2$ implying that $d_1 \rightsquigarrow_{\mathcal{I},r} \cdots \rightsquigarrow_{\mathcal{I},r} d_m = e' \rightsquigarrow_{\mathcal{I},r} e_2 \rightsquigarrow_{\mathcal{I},r} \cdots \rightsquigarrow_{\mathcal{I},r} e_n = f$. Hence, $f \in W^{\rightsquigarrow}_{\mathcal{I},r}(d) \subseteq \mathsf{Wit}_{\mathcal{I},r}(d)$.

□

We will establish next Theorem 5; the proof is separated in two lemmata.

Lemma 12. *$(T, E, \mathsf{bg}, \mathsf{rl})$ is a canonical tree decomposition of $\mathcal{J}$.*

Proof. Items (i) and (ii) of $(T, E, \mathsf{bg}, \mathsf{rl})$ being a tree decomposition of $\mathcal{J}$ are an immediate consequence of the definition of $\mathcal{J}$, bg, and rl, particularly due to (†). Items (iii) and (iv) are also clear from the construction. For Item (v), we denote, for every $\delta \in \Delta^{\mathcal{J}}$, with w_δ the node where δ was introduced in the domain, either in the initialization or in rules $\mathbf{R_1}$–$\mathbf{R_3}$. It suffices to show the following claim.

Claim. For every $w \in T$ and every $\delta \in \mathsf{dom}(\mathsf{bg}(w))$, we have $\delta \in \mathsf{dom}(\mathsf{bg}(v))$ for every v on the path from w_δ to w.

Proof of the Claim. The proof is by induction. Obviously, $\delta \in \mathsf{dom}(\mathsf{bg}(w))$. Now take any v on the path from w_δ to w, and let v' be its successor on this path. By induction, we can assume that $\delta \in \mathsf{dom}(\mathsf{bg}(v'))$. We distinguish cases which rule has been applied in order to obtain v' from v:

- For $\mathbf{R_1}$, it suffices to note that the only element that is in $\mathsf{dom}(\mathsf{bg}(v'))$, but not in $\mathsf{dom}(\mathsf{bg}(v))$, is the freshly introduced element $d_{v'}$. By choice of v', we have $v' \neq w_\delta$. Hence, also $\delta \neq d_{v'}$ and thus $\delta \in \mathsf{dom}(\mathsf{bg}(v))$.
- For $\mathbf{R_2}$, we can argue analogously.
- For $\mathbf{R_3}$, assume that it was applied due to some r-cluster $\mathbf{a}$ in $\mathsf{bg}(v)$. By assumption, we have that $\delta \in \mathit{\Delta}' = \mathsf{dom}(\mathsf{bg}(v'))$. Observe first that $\delta \notin \mathit{\Delta}$ by the same reasoning as for $\mathbf{R_1}$. Thus, $\delta \in \mathbf{a}$ or $\delta \in \{\delta'' \mid r(\delta', \delta'') \in \mathsf{bg}(v)$ for some $\delta' \in \mathbf{a}\}$. However, in both cases, we obtain $\delta \in \mathsf{dom}(\mathsf{bg}(v))$.

This finishes the proof of the Claim, and thus of Item (v). Hence, $(T, E, \mathsf{bg}, \mathsf{rl})$ is a tree decomposition of $\mathcal{J}$.

We argue next that $(T, E, \mathsf{bg}, \mathsf{rl})$ is canonical. Let $w \in T$ with $M = \mathsf{bg}(w)$ and $r = \mathsf{rl}(w)$, and w' a successor of w with $M' = \mathsf{bg}(w')$ and $r' = \mathsf{rl}(w')$ For (C1) we distinguish cases which rule has been applied to obtain w' from w.

- For $\mathbf{R_1}$ and $\mathbf{R_2}$, observe that the overlap $\mathsf{dom}(M) \cap \mathsf{dom}(M')$ is a singleton set $\{d\}$. It suffices to note that by the definition of $\mathbf{R_1}$ and $\mathbf{R_2}$ (in case of $\mathbf{R_2}$ particularly because of (†)), we have that $A(d) \in M$ iff $A(d) \in M'$ for all $A \in \mathsf{N_C}$.
- For $\mathbf{R_3}$, it follows from the fact that (†) is only applied to all $\delta \in \mathbf{a} \cup \mathit{\Delta}$ and $\delta' \in \mathit{\Delta}'$. Thus, the interpretation of elements already present is not changed.

For verifying that (C2)–(C4) are satisfied, we also distinguish cases which rule has been applied.

- In case of $\mathbf{R_1}$, it is clear from the definition of $\mathbf{R_1}$, that (C2) is satisfied.
- If $\mathbf{R_2}$ has been applied, it is clear from the definition of $\mathbf{R_2}$ that the overlap is a singleton $\{d\}$ and that $r \neq r'$. By the premise of the rule, we know that $d \in F(w)$. By definition of $\mathsf{Wit}_{\mathcal{I},r}$, we know that there is an r-root cluster $\mathbf{a}$ in M'. Moreover, observe that both premises (a) and (b) imply that there is no $r(d, d')$ or $r(d', d)$ in M. Finally, observe that $\mathbf{R_2}$ is applied only once to every $d \in F(w)$, and $r'' \neq r$. Thus, $\mathbf{R_2}$ satisfies (C3).
- Suppose $\mathbf{R_3}$ has been applied to some r-cluster $\mathbf{a}$ in M with $\mathbf{a} \subseteq F_r(w)$ and a direct r-successor $\hat{e}$ of $\tau(\delta)$ in $\mathcal{I}$, for some $\delta \in \mathbf{a}$ such that $(\delta, \delta') \notin r^{\mathcal{J}}$, for any δ' with $\tau(\delta') = \hat{e}$. We show that $\mathbf{a}$ witnesses (C4).
  By assumption, $\mathbf{a}$ satisfies the precondition of (C4). By definition of $\mathit{\Delta}'$ and (†), $\mathbf{a}$ is an r-root cluster in M', thus (C4) (a) is satisfied. Condition (C4) (b) is clear from the definition of $\mathit{\Delta}'$. Finally, for (C4) (c) assume $r(d, e) \in M$ and suppose that $e \in F(w)$. By definition of $\mathbf{R_3}$, we know that $(d, e) \in r^{\mathcal{J}}$. Because $e \in F(w)$, the tuple (d, e) has been added to $r^{\mathcal{J}}$ in this step via the application of (†). Thus, we obtain $d \in \mathbf{a} \cup \mathit{\Delta} = \mathbf{a} \cup F(w)$, as required.

Lemma 13. $(T, E, \mathsf{bg}, \mathsf{rl})$ *and* $\mathcal{J}$ *satisfy Conditions (1)–(4) of Theorem 5.*

Proof. We start with Condition (1). Obviously, the initialization phase ensures that $\mathcal{A} \subseteq \mathsf{bg}(\varepsilon)$. Moreover, by construction of $\mathsf{bg}(\varepsilon)$ and Lemma 11, we get the desired bound. Condition (3) is witnessed by the fact that τ is in fact the required homomorphism.

For Condition (2), we verify the following claim by induction on the structure of concept names.

Claim. For all $\delta \in \Delta^{\mathcal{J}}$ and $C \in \mathsf{cl}(\mathcal{T})$

$$\delta \in C^{\mathcal{J}} \text{ iff } \tau(\delta) \in C^{\mathcal{I}}$$

Proof of the Claim. The proof is by induction on C. The case $C = A$ for $A \in \mathsf{N_C}$ follows from τ being a homomorphism and rules $\mathbf{R_1}$- $\mathbf{R_3}$. The Boolean cases $C = \neg D$ and $C = C_1 \sqcap C_2$ are consequences of the induction hypothesis. It thus remains to consider concepts of the form $C = (\sim n\ r\ D)$. If $r \in \mathsf{N}_\mathsf{R}^{nt}$, the claim is a straightforward consequence of the induction hypothesis and construction rule $\mathbf{R_1}$. Now, assume that $r \in \mathsf{N}_\mathsf{R}^{t}$. It suffices to show that:

(a) If $\tau(\delta) \in (\geq n\ r\ D)^{\mathcal{I}}$ then $\delta \in (\geq n\ r\ D)^{\mathcal{J}}$ and
(b) if $\tau(\delta) \in (\leq n\ r\ D)^{\mathcal{I}}$ then $\delta \in (\leq n\ r\ D)^{\mathcal{J}}$.

For Point (a), let $d_0, \dots, d_n$ be a sequence such that d_{i+1} is a direct r-successor of d_i in $\mathcal{I}$, for all $0 \leq i < n$. We start with showing that for every $\delta \in \Delta^{\mathcal{J}}$ with $\tau(\delta) = d_0$ and every $0 \leq i \leq n$ there is an r-path from δ to some δ_i with $\tau(\delta_i) = d_i$. This is clear for $i = 0$. So suppose $i > 0$ and let $\pi = \delta_0, w_0, \dots, \delta_{i-1}$ be an r-path from δ_0 to δ_{i-1} with $\tau(\delta_{i-1}) = d_{i-1}$. We fix $w \in T$ and an r-cluster $\mathbf{a} \subseteq F_r(w)$ with $\delta_{i-1} \in \mathbf{a}$ as follows. If $\delta_{i-1} \in \Delta'_r$, then set $w = \varepsilon$ and $\mathbf{a}$ the r-cluster containing δ_{i-1}. Otherwise, by construction, there is some $w \neq \varepsilon$ such that $\mathsf{rl}(w) = r$ and δ_{i-1} belongs to some r-cluster $\mathbf{a} \subseteq F_r(w)$. We distinguish two cases:

- If $(\delta, \delta') \in r^{\mathcal{J}}$, for some δ' with $\tau(\delta') = d_i$. Then, there is some r-path $\delta, v_0, \dots, \delta'$ from δ to δ' in $(T, E, \mathsf{bg}, \mathsf{rl})$. Then $\pi, v_0, \dots, \delta'$ is the required r-path.
- If $(\delta, \delta') \notin r^{\mathcal{J}}$, for all δ' with $\tau(\delta') = d_i$, then $\mathbf{R_3}$ applies to w, $\mathbf{a}$, and d_i. In particular, it adds a successor v of w to T and adds a domain element $\delta' = (d_i)_v \in \Delta$ to $\mathcal{J}$ such that $(\delta_i, \delta') \in r^{\mathcal{J}}$. By construction, $\mathsf{bg}(v)$ contains $r(\delta, \delta')$ and π, v, δ' is the required r-path.

Let now be $\tau(\delta) \in (\geq n\ r\ D)^{\mathcal{I}}$ and $\tau(\delta) = d$. Let $d' \in \Delta^{\mathcal{I}}$ be any domain element such that $(d, d') \in r^{\mathcal{I}}$ and $d' \in D^{\mathcal{I}}$. Thus, either $d' \in Q_{\mathcal{I},r}(d)$ or there is a sequence $d_0, \dots, d_n$ with $d_0 = d$, $d_n = d'$, and d_{i+1} is a direct r-successor of d_i in $\mathcal{I}$, for all $0 \leq i < n$. In the first case, by construction, there is $\delta' \in Q_{\mathcal{J},r}(\delta)$ with $\tau(\delta') = d'$. By induction, $\delta' \in D^{\mathcal{J}}$. In the second case, by the statement above, there is an r-path from δ to some δ' with $\tau(\delta') = d'$. By induction, we know that $\delta' \in D^{\mathcal{J}}$. Since distinct d' with $(d, d') \in r^{\mathcal{I}}$ and $d' \in D^{\mathcal{I}}$ yield distinct δ', this finishes the proof of (a).

For Point (b), assume $\tau(\delta) \in (\leq n\ r\ D)^{\mathcal{I}}$ with $\tau(\delta) = d$ and set $w = w_{\delta,r}$. It clearly suffices to show that (i) for every $e \in \mathsf{Wit}_{\mathcal{I},r}(d)$, there is precisely one $\delta' \in \mathsf{dom}(\mathsf{bg}(w))$ with $r(\delta,\delta') \in \mathsf{bg}(w)$ and $\tau(\delta') = e$, and (ii) whenever $(\delta,\delta') \in r^{\mathcal{J}}$ and $\delta' \in D^{\mathcal{J}}$, then $r(\delta,\delta') \in \mathsf{bg}(w)$.

For (i), observe that either $w = \varepsilon$ or it was added either by an application of $\mathbf{R_2}$ or $\mathbf{R_3}$. Suppose first $w = \varepsilon$. If $\delta = a \in \mathsf{ind}(\mathcal{A})$ then (i) is clear due to the initialization of $\mathcal{J}$ and $(T, E, \mathsf{bg}, \mathsf{rl})$. If $\delta = d_r \in \Delta_r$, we have that $d \in \mathsf{Wit}_{\mathcal{I},r}(a)$ for some individual a, and by Lemma 11, Point 2, we have that $\mathsf{Wit}_{\mathcal{I},r}(d) \subseteq \mathsf{Wit}_{\mathcal{I},r}(a)$. Together with the initialization of $\mathcal{J}$ and of $(T, E, \mathsf{bg}, \mathsf{rl})$, this yields that $\mathsf{dom}(\varepsilon)$ contains exactly one element δ_e with $r(\delta,\delta_e) \in \mathsf{bg}(\varepsilon)$ and $\tau(\delta_e) = e$, for each $e \in \mathsf{Wit}_{\mathcal{I},r}(d)$. Assume that (i) holds for $w' = \mathsf{pre}(w)$ for $w \neq \varepsilon$.

- If w was created by $\mathbf{R_2}$, then there is some $\hat{\delta} \in \mathsf{dom}(\mathsf{bg}(w)) \cap \mathsf{dom}(\mathsf{bg}(w'))$ such that for every $\delta' \in \mathsf{dom}(\mathsf{bg}(w))$, with $\hat{\delta} \neq \delta'$, $\tau(\delta') \in \mathsf{Wit}_{\mathcal{I},r}(\tau(\hat{\delta}))$. From this, we have that if $\delta' = \hat{\delta}$, then the claim follows; if $\delta \neq \hat{\delta}$, then the claim follows from point (2) in Lemma 11.
- If w was created by $\mathbf{R_3}$, let $\mathbf{a} \subseteq F_r(w')$ be the r-cluster in $\mathsf{dom}(\mathsf{bg}(w))$ that witnesses this. Then there is some $\hat{\delta} \in \mathbf{a}$ and $\delta' \in \mathsf{dom}(\mathsf{bg}(w))$ such that $\tau(\delta')$ is a direct successor of $\tau(\hat{\delta})$. Further, since $w = w_{\delta,r}$ we have $\delta \in F(w)$. Then, by the definition of $\mathbf{R_3}$, $\tau(\delta) \in \mathsf{Wit}_{\mathcal{I},r}(\tau(\delta')) \setminus \mathsf{Wit}_{\mathcal{I},r}(\hat{\delta})$. By point 2 of Lemma 11, the definition of $\mathbf{R_3}$ ensures that there is exactly one fresh element in $\mathsf{dom}(\mathsf{bg}(w))$ for every $e \in \mathsf{Wit}_{\mathcal{I},r}(\delta) \setminus \mathsf{Wit}_{\mathcal{I},r}(\hat{\delta})$. It remains to show that there is exactly one element in $\mathsf{dom}(\mathsf{bg}(w))$ for every $e \in \mathsf{Wit}_{\mathcal{I},r}(\hat{\delta}) \cap \mathsf{Wit}_{\mathcal{I},r}(\delta')$. Indeed, by assumption, for every such e, there is some $d'' \in \mathsf{dom}(w')$ such that $\tau(\delta'') = e$. Moreover $r(\hat{\delta}, \delta'') \in \mathsf{bg}(w')$, and by the definition of $\mathbf{R_3}$, we have $\delta'' \in \mathsf{dom}(\mathsf{bg}(w))$.

For (ii), take any δ' with $(\delta,\delta') \in r^{\mathcal{J}}$ and $\delta' \in D^{\mathcal{J}}$. By Lemma 4, there is a canonical r-path $\pi = \delta_0, w_0, \ldots, \delta_k$ from δ to δ'. We have to show that $r(\delta,\delta') \in \mathsf{bg}(w)$ with $w = w_{\delta,r}$. We distinguish cases according to the shape of π.

- Suppose **P1** applies to π.
  - If $k = 1$ and $w_0 = w_{\delta,r}$, we are done.
  - If $k = 1$ and $w_0 \neq w_{\delta,r}$, (C4) (c) implies that w_0 was introduced via an application of $\mathbf{R_3}$ to $w_{\delta,r}$ and the r-cluster $\mathbf{a}$ containing δ. Indeed, since $\delta \notin F(w_0)$ and $\delta' \in F(w_0)$, we have that $\delta \in \mathbf{a}$. Let e be the direct r-successor from this application of $\mathbf{R_3}$. Then, δ' is of the shape f_{w_0} for some $f \in \mathsf{Wit}_{\mathcal{I},r}(e) \setminus \mathsf{Wit}_{\mathcal{I},r}(\tau(\delta))$. In particular, we have $f \notin \mathsf{Wit}_{\mathcal{I},r}(\tau(\delta))$, hence $f \notin D^{\mathcal{I}}$, and thus $\delta' \notin D^{\mathcal{J}}$.
  - If $k \geq 2$, consider δ_{k-1}, w_{k-1}, $\delta' = \delta_k$. By canonicity, we know that $\delta_k \in F_r(w_{k-1})$, thus $\delta_k \in F(w_{k-1})$, and $\delta_{k-1} \notin F(w_{k-1})$. Thus, we can argue as in the previous item that w_{k-1} was $\mathbf{R_3}$ was applied to some r-cluster $\mathbf{a}$ in w_{k-2} with $\delta_{k-1} \in \mathbf{a}$. By the r-path from δ to δ_{k-1}, we know that $(\delta, \delta_{k-1}) \in r^{\mathcal{J}}$ and hence $(\tau(\delta), \tau(\delta_{k-1})) \in r^{\mathcal{I}}$. Thus, also $\tau(\delta_{k-1}) \in (\leq n\ r\ D)^{\mathcal{I}}$. This leads to a contradiction as in the previous case.

– Suppose **P2** applies to π. By canonicity, we know $w_0 = w_{\delta,r}$ and $\delta_1 \notin F_r(w_0)$. If $k = 1$, then we are done.
If $k > 1$, then $\mathsf{pre}(w_1) = w_{\delta_1,r}$ and $\delta_1, w_1, \ldots, w_{k-1}, \delta_k$ is downward. From $(\delta, \delta_1) \in r^{\mathcal{J}}$, we have that $(\tau(\delta), \tau(\delta_1)) \in r^{\mathcal{I}}$ and $\tau(\delta_1) \in (\leq n\ r\ D)^{\mathcal{I}}$; and since $\tau(\delta') \in D^{\mathcal{I}}$, $\tau(\delta') \in \mathsf{Wit}_{\mathcal{I},r}(\delta_1)$. Then, using similar arguments as in the previous case, we can conclude that $k = 2$, $r(\delta_1, \delta') \in \mathsf{bg}(w_1)$ and $\delta' \in F_r(w_1)$. Moreover, we know that w_1 was added either by $\mathbf{R_2}$ or $\mathbf{R_3}$. The former cannot be the case because $\mathsf{pre}(w_1) = w_{\delta_1,r}$ and $\mathsf{dom}(\mathsf{pre}(w_1)) \cap \mathsf{dom}(w_1)$. Thus, w_1 was created by an application of $\mathbf{R_3}$. Since $r(\delta_1, \delta') \in \mathsf{bg}(w_1)$, by the definition of $\mathbf{R_3}$, we know that $\delta' \notin F(w_1)$ because $\tau(\delta') \in \mathsf{Wit}_{\mathcal{I},r}(\tau(\delta'))$ and $d_1 \notin F(w_1)$. Then, $\delta' \in \mathsf{dom}(\mathsf{pre}(w_1))$ and $r(\delta_1, \delta') \in \mathsf{bg}(\mathsf{pre}(w_1))$, but then $w_1 \neq w_{\delta_1,r}$, a contradiction.

For Condition (4), we distinguish cases according to which rule was applied. We start with the width.

– By Condition (1), we know that $|\mathsf{dom}(\mathsf{bg}(\varepsilon))| \leq O(|\mathcal{A}| \cdot 2^{\mathsf{poly}(|\mathcal{T}|)})$.
– If w was created by $\mathbf{R_1}$, then $|\mathsf{dom}(\mathsf{bg}(w))| = 2$.
– If w was created by $\mathbf{R_2}$, then $\mathsf{dom}(\mathsf{bg}(w)) \subseteq \mathcal{A} \cup X$, for some set X of size $2^{\mathsf{poly}(|\mathcal{T}|)}$), by Lemma 11.
– If w was created by $\mathbf{R_3}$, we make the following observations, where $\mathcal{C}$ denotes the set of all concepts $(\leq m\ r\ D)$ appearing in $\mathcal{T}$.
  (i) For all $(d, e) \in r^{\mathcal{I}}$, $C \in \mathcal{C}$: if $d \in C^{\mathcal{I}}$, then $e \in C^{\mathcal{I}}$;
  (ii) Let $d_1, \ldots, d_n$ be such that $(d_i, d_{i+1}) \in r^{\mathcal{I}}$, for all $1 \leq i < n$ and $n > |\mathcal{T}|$. Then $\mathsf{Wit}_{\mathcal{I},r}(d_i) = \mathsf{Wit}_{\mathcal{I},r}(d_j)$ for some $i \neq j$.
  Point (i) follows from the semantic of $(\leq m\ r\ D)$ and transitivity. For Point (ii) observe that $|\mathcal{C}| < n$, thus there are $i \neq j$ such that $d_i \in C^{\mathcal{I}}$ iff $d_j \in C^{\mathcal{I}}$, for all $C \in \mathcal{C}$. By definition of $\rightsquigarrow_{\mathcal{I},r}$ and $\mathsf{Wit}_{\mathcal{I},r}$, we also have $\mathsf{Wit}_{\mathcal{I},r}(d_i) = \mathsf{Wit}_{\mathcal{I},r}(d_j)$.
  Now consider some branch of applications of $\mathbf{R_3}$. Each application adds (copies of) elements which are new witnesses, that is, they are in $\mathsf{Wit}_{\mathcal{I},r}(e) \setminus \mathsf{Wit}_{\mathcal{I},r}(d)$, for some $(d, e) \in r^{\mathcal{I}}$. By Point (ii), along such a branch, elements are added at most $|\mathcal{T}|$ times. Each time, at most $|\mathsf{Wit}_{\mathcal{I},r}(d)|$ elements are added. Overall, the size is bounded by $|\mathcal{T}| \cdot (|\mathcal{A}| \cdot 2^{\mathsf{poly}(|\mathcal{T}|)}) = O(|\mathcal{A}| \cdot 2^{\mathsf{poly}(|\mathcal{T}|)})$.

Finally, the outdegree is bounded by $k_1 \cdot k_2 \cdot k_3$, where k_1 is the number of elements in a bag, k_2 is the number of role names, and k_3 is the maximal outdegree in $\mathcal{I}$. We have seen bounds for k_1 and k_2. So it remains to note that the outdegree in $\mathcal{I}$ is bounded by $|\mathcal{A}| + 2^{\mathsf{poly}(|\mathcal{T}|)}$, by Lemma 6.

□

Proofs of Section 4

Semantics of 2ATAs. A *run* of $\mathfrak{A}$ on a labelled tree (T, τ) is a $T_r \times Q$-labelled tree (T_r, r) such that $r(\varepsilon) = (\varepsilon, q_0)$ and whenever $x \in T_r, r(x) = (w, q)$, and $\delta(q, \tau(w)) = \theta$, then there is a set $\mathcal{S} = \{(m_1, q_1), \ldots, (m_n, q_n)\} \subseteq [k] \times Q$ such that $\mathcal{S}$ satisfies θ and for $1 \leq i \leq n$, we have $x \cdot i \in T_r$, $w \cdot m_i$ is defined, and $\tau_r(x \cdot i) = (w \cdot m_i, q_i)$. A run is *accepting* if every infinite path π satisfies the *parity condition*. A *parity condition* F *over* Q is a finite sequence $G_1, \dots, G_m$ with $G_1 \subseteq G_2 \subseteq \dots \subseteq G_m = Q$. An infinite path π satisfies F if there is an even i such that $\mathsf{inf}(\pi) \cap G_i \neq \emptyset$ and $\mathsf{inf}(\pi) \cap G_{i-1} = \emptyset$, where $\mathsf{inf}(\pi) \subseteq Q$ denotes the set of states that occur infinitely often in π. The automaton accepts an input tree if there is an accepting run for it. We use $L(\mathfrak{A})$ to denote the set of trees accepted by $\mathfrak{A}$. The *nonemptiness problem* is to decide, given a 2ATA $\mathfrak{A}$, whether $L(\mathfrak{A})$ is nonempty. The nonemptiness problem for 2ATAs is ExpTime-complete in the number of states [26].

Lemma 8. *There are 2ATAs $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3$ of size $O(|\mathcal{A}| \cdot 2^{\mathsf{poly}(|\mathcal{T}|)})$ such that: $(T, \tau) \in L(\mathfrak{A}_1)$ iff (T, τ') is a canonical decomposition (of some interpretation); $(T, \tau) \in L(\mathfrak{A}_2)$ iff $\mathcal{I}_{T,\tau} \models \mathcal{A}$, and $(T, \tau) \in L(\mathfrak{A}_3)$ iff $\mathcal{I}_{T,\tau} \models \mathcal{T}$.*

Proof. For devising $\mathfrak{A}_1$, observe that the properties of canonical tree compositions can easily be checked using a 2ATA; we leave the details to the reader.

For devising $\mathfrak{A}_2$, we assume without loss of generality that $\mathsf{ind}(\mathcal{A}) \subseteq \Delta$ and identify, in the translation to (T, τ'), $[\varepsilon]_a$ with a, for each $a \in \mathsf{ind}(\mathcal{A})$, to reflect the SNA. Checking whether $\mathcal{I}_{T,\tau} \models \mathcal{A}$ then amounts to checking that $\alpha \in M$ for each $\alpha \in \mathcal{A}$ and for $\tau(\varepsilon) = \langle M, x \rangle$, which can easily be done using a 2ATA.

For $\mathfrak{A}_3$, we give the missing transitions and states. We start with the transitions for number restrictions on non-transitive roles. The automaton uses the following additional states for this:

$$\begin{aligned} Q^{nt} = \{ & q_{r,d,e}, \overline{q}_{r,d,e} \mid r \in \mathsf{N}_\mathsf{R}^{nt}, d, e \in \Delta\} \cup \\ \{ & q_{d,r,D,b} \mid d \in \Delta, r \in \mathsf{N}_\mathsf{R}^{nt}, D \in \mathsf{nnf}(\mathcal{T}), b \in \{0,1\}\} \cup \\ \{ & q_{n,r,D,d} \mid n \leq N_0, r \in \mathsf{N}_\mathsf{R}^{nt}, D \in \mathsf{nnf}(\mathcal{T}), d \in \Delta\} \end{aligned}$$

The transitions for these states are based on (C2):

$$\delta_3(q^*_{(\sim n\, r\, D),d}, \langle M, x \rangle) = \bigvee_{\substack{b_1,\dots,b_k \in \{0,1\}^k \\ \sum_i b_i \sim n}} \bigwedge_{i=1}^{k} (i, q_{d,r,D,b_i}) \qquad \text{if } x \neq \varepsilon$$

$$\delta_3(q_{(\sim n\, r\, D),d}, \langle M, \varepsilon \rangle) = \bigvee_{\substack{b_1 \dots, b_k \in \{0,1\}^k, m \\ \sum_i b_i \sim n - m}} (0, q_{m,r,D,d}) \wedge \bigwedge_{i=1}^{k} (i, q_{d,r,D,b_i})$$

$$\delta_3(q_{n,r,D,d}, \langle M, \varepsilon \rangle) = \bigvee_{\substack{Y \subseteq \mathsf{dom}(M), \\ |Y| = n}} \Big(\bigwedge_{e \in Y} (q_{r,d,e} \wedge q_{D,e}) \wedge \bigwedge_{e \in \mathsf{dom}(M) \setminus Y} (\overline{q}_{r,d,e} \vee q_{\sim D,e}) \Big)$$

$$\delta_3(q_{r,d,e}, \langle M, x \rangle) = \begin{cases} \mathsf{true} & \text{if } r(d,e) \in M \\ \mathsf{false} & \text{otherwise} \end{cases}$$

$$\delta_3(\overline{q}_{r,d,e}, \langle M, x \rangle) = \begin{cases} \mathsf{true} & \text{if } r(d,e) \notin M \\ \mathsf{false} & \text{otherwise} \end{cases}$$

$$\delta_3(q_{d,r,D,1}, \bullet) = \mathsf{false}$$

$$\delta_3(q_{d,r,D,1}, \langle M, x\rangle) = \begin{cases} q_{D,e} & \text{if } x = r \text{ and } r(d,e) \in M \text{ (e well-defined by (C2))} \\ \mathsf{false} & \text{if } x \neq r \end{cases}$$

$$\delta_3(q_{d,r,D,0}, \bullet) = \mathsf{true}$$

$$\delta_3(q_{d,r,D,0}, \langle M, x\rangle) = \begin{cases} q_{\sim D,e} & \text{if } x = r \text{ and } r(d,e) \in M \text{ (e well-defined by (C2))} \\ \mathsf{true} & \text{if } x \neq r \end{cases}$$

Let us consider the first transition, which is used for inner nodes of the input tree. There, the automaton chooses a bit-vector $b_1, \ldots, b_k$ over all successor nodes such that the number of 1's in the b_i satisfies the number restriction. It then verifies that the choice was correct in the sense that b_i is 1 iff successor i contains an r-successor satisfying D. For the root node, we have to take into account witnesses for the number restriction that are contained in the bag; this is done by guessing an additional number m and verifying, in state $q_{m,r,D,d}$, that this bag contains *precisely* m witnesses.

Next, we give the missing transitions for states $\overline{F}_{x,d}$ and $\overline{F}'_{x,d}$ which intuitively behave complementary to $F_{x,d}$ and $F'_{x,d}$:

$$\delta_3(\overline{F}_{\varepsilon,d}, \langle M, x\rangle) = \mathsf{false}$$

$$\delta_3(\overline{F}_{r,d}, \langle M, x\rangle) = \begin{cases} (-1, \overline{F}'_{r,d}) & \text{if } r \notin \mathsf{N}_\mathsf{R}^{nt} \text{ and } (r \notin \mathsf{N}_\mathsf{R}^{t} \text{ or } x = r) \\ \mathsf{true} & \text{otherwise} \end{cases}$$

$$\delta_3(\overline{F}'_{r,d}, \langle M, x\rangle) = \begin{cases} \mathsf{false} & d \notin \mathsf{dom}(M) \text{ or } (r \in \mathsf{N}_\mathsf{R}^{t} \text{ and } x \notin \{\varepsilon, r\}) \\ \mathsf{true} & \text{otherwise} \end{cases}$$

Finally, the following are the transitions for the at-most restrictions (for transitive roles). The strategy there is to try to find $n + 1$ r-successors satisfying D and accept if this fails; thus "complementing" the strategy for the at-least restrictions.

$$\delta_3(q^*_{(\leq n\, r\, D),d}, \langle M, x\rangle) = \bigwedge_{\substack{n_1+\ldots+n_\ell=n+1 \\ \text{respects } M \text{ rel. to } d}} \bigvee_{n_i \neq 0} (0, q^0_{(\leq n_i\, r\, D),a_i}) \wedge (0, q^1_{(\leq n_i\, r\, D),a_i})$$

$$\delta_3(q^0_{(\leq n\, r\, D),d}, \langle M, x\rangle) = (0, \overline{F}_{r,d}) \vee (0, q^{\downarrow}_{(\geq n\, r\, D),d})$$

$$\delta_3(q^1_{(\leq n\, r\, D),d}, \langle M, x\rangle) = (0, F_{r,d}) \vee (-1, q^{\uparrow}_{(\geq n\, r\, D),d})$$

$$\delta_3(q^{\downarrow}_{(\leq n\, r\, D),d}, \langle M, x\rangle) = \bigwedge_{n_0+n_1+\ldots+n_k=n} (0, p'_{n_0,r,D,d}) \vee \bigvee_{i=1}^{k} (i, p_{(\leq n_i\, r\, D),d})$$

$$\delta_3(p'_{n,r,D,d}, \langle M, x\rangle) = \bigwedge_{Y \subseteq M_r(d), |Y|=n} \Big(\bigvee_{e \in Y} q_{\sim D,e} \vee \bigvee_{y \in M_r(d)\setminus Y} q_{D,e} \Big)$$

$$\delta_3(p_{(\leq n\, r\, D),d}, \bullet) = \begin{cases} \mathsf{true} & \text{if } n > 0 \\ \mathsf{false} & \text{otherwise} \end{cases}$$

$$\delta_3(p_{(\leq n\, r\, D),d}, \langle M, x\rangle) = \begin{cases} \mathsf{true} & \text{if } x \neq r \text{ or } d \text{ not in root cluster} \\ \displaystyle\bigwedge_{\substack{n_1+\ldots+n_\ell=n \\ \text{d-respects } M \text{ rel. to } d}} \bigvee_{n_i\neq 0} (0, q^0_{(\leq n_i\, r\, D),a_i}) & \text{otherwise} \end{cases}$$

$$\delta_3(q^{\uparrow}_{(\leq n\, r\, D),d}, \langle M, x\rangle) = q_d \wedge (0, q^0_{(\leq n\, r\, D),d}) \wedge (0, q^1_{(\leq n\, r\, D),d})$$

It remains to define the acceptance condition F. We set $F = G_1, G_2, G_3$ where $G_1 = \emptyset$, G_2 contains all states of the form $q^*_{(\sim n\, r\, D),d}$ and $p_{n,r,D,d}$ with $n \geq 1$, and $G_3 = Q_3$. □

Lemma 14. *There is a non-deterministic tree automaton $\mathfrak{A}_q$ of size $O(2^{\mathsf{poly}(|q|)} \cdot 2^{2^{\mathsf{poly}|\mathcal{K}|}})$ such that for all $(T, \tau) \in L(\mathfrak{A}_1)$, we have $(T, \tau) \in L(\mathfrak{A}_q)$ iff $\mathcal{I}_{T,\tau} \models q$.*

Proof. We start with automata which work over an extended alphabet to also encode a possible query match. Formally, let $\mathsf{term}(q)$ be the set of all terms, that is, constants and variables, in q and Γ_q be the set of all assignments $t \to d$ with $t \in \mathsf{term}(q)$ and $d \in \Delta$. The extended alphabet is $\Sigma_q = \Sigma \times 2^{\Gamma_q}$, that is, each node is additionally labeled with a set of assignments from Γ_q. We call a Σ_q-labelled tree (T, τ) *proper*, if the following conditions are satisfied:

- For all $w \in T$ with $\tau(w) = \langle M, x, N\rangle$:
  - $t \to d \in N$ implies $d \in \mathsf{dom}(M)$;
  - $a \to d \in N$, a a constant, implies $d = a$ and $w = \varepsilon$;
- for each term $t \in \mathsf{term}(q)$, there is precisely one pair d, w with $\tau(w) = \langle M, x, N\rangle$ such that $t \to d \in N$, that is, every term is mapped to precisely one element.

Intuitively, a proper (T, τ) corresponds to an interpretation and a possible match for the query. It is straightforward to give a 2ATA $\mathfrak{A}^0_q$ of size polynomial in q that checks properness of its input.

As the next step, we develop a 2ATA $\mathfrak{A}^1_q$ such that for all $(T, \tau) \in L(\mathfrak{A}^0_q)$, we have $(T, \tau) \in L(\mathfrak{A}^1_q)$ iff the mapping π_τ encoded by a proper (T, τ) is a match for q in $\mathcal{I}_{T,\tau}$. Let $q = \exists \boldsymbol{y}.\varphi(\boldsymbol{y})$, and denote with $\mathsf{sub}(\varphi)$ the set of sub-formulas of φ. Moreover, for every regular expression $\mathcal{E}$ appearing in φ, fix a (polynomially-sized) NFA $\mathcal{B}_\mathcal{E} = (Q_\mathcal{E}, \mathcal{S}, s^0_\mathcal{E}, \Delta_\mathcal{E}, F_\mathcal{E})$ with $L(\mathcal{E}) = L(\mathcal{B}_\mathcal{E})$ over the alphabet $\mathcal{S}$ defined in the preliminaries.

For defining the automaton $\mathfrak{A}^1_q = (Q, \Sigma_q, q_\varphi, \delta, F)$, we take as the set of states

$$\begin{aligned} Q = {} & \{q_\psi \mid \psi \in \mathsf{sub}(\varphi)\} \cup \{q_{t\to d}, q^*_{t\to d}, q_d \mid t \in \mathsf{term}(q), d \in \Delta\} \cup \\ & \{q^\mathcal{E}_{s,d,t}, \mathsf{acc}^\mathcal{E}_{s,d,t}, q^\mathcal{E}_{s,u,s',d,t} \mid \mathcal{E}(t', t) \in \mathsf{sub}(\varphi), s, s' \in Q_\mathcal{E}, u \in \mathcal{S}, d \in \Delta\} \end{aligned}$$

and use the following transitions:

$$\delta(q_{\varphi_1\wedge\varphi_2}, \langle M, x, N\rangle) = (0, q_{\varphi_1}) \wedge (0, q_{\varphi_2})$$
$$\delta(q_{\varphi_1\vee\varphi_2}, \langle M, x, N\rangle) = (0, q_{\varphi_1}) \vee (0, q_{\varphi_2})$$

$$\delta(q_{A(t)}, \langle M, x, N\rangle) = \begin{cases} \mathsf{true} & \text{if } t \to d \in N \text{ and } A(d) \in M \\ \mathsf{false} & \text{if } t \to d \in N \text{ and } A(d) \notin M \\ \bigvee_{i\in[k]}(i, q_{A(t)}) & \text{if } t \to d \notin N \end{cases}$$

$$\delta(q_{\mathcal{E}(t,t')}, \langle M, x, N\rangle) = \Big(\bigvee_{d\in\Delta} (0, q_{t\to d}) \wedge (0, q^{\mathcal{E}}_{s^0_{\mathcal{E}},d,t'})\Big) \vee \bigvee_{i\in[k]} (i, q_{\mathcal{E}(t,t')})$$

$$\delta(q_{t\to d}, \langle M, x, N\rangle) = \begin{cases} \mathsf{true} & \text{if } t \to d \in N \\ \mathsf{false} & \text{otherwise} \end{cases}$$

$$\delta(q^*_{t\to d}, \langle M, x, N\rangle) = (0, q_{t\to d}) \vee \bigvee_{i\in[k]} \big((i, q_d) \wedge (i, q^*_{t\to d})\big)$$

$$\delta(q_d, \langle M, x, N\rangle) = \begin{cases} \mathsf{true} & \text{if } d \in \mathsf{dom}(M) \\ \mathsf{false} & \text{otherwise} \end{cases}$$

$$\delta(q^{\mathcal{E}}_{s,d,t}, \langle M, x, N\rangle) = (0, \mathsf{acc}^{\mathcal{E}}_{s,d,t}) \vee \bigvee_{(s,u,s')\in\Delta_{\mathcal{E}}} (0, q^{\mathcal{E}}_{s,u,s',d,t})$$

$$\delta(\mathsf{acc}^{\mathcal{E}}_{s,d,t}) = \begin{cases} (0, q^*_{t\to d}) & \text{if } s \in F_{\mathcal{E}} \\ \mathsf{false} & \text{if } s \notin F_{\mathcal{E}} \end{cases}$$

$$\delta(q^{\mathcal{E}}_{s,A?,s',d,t}, \langle M, x, N\rangle) = \begin{cases} (0, q^{\mathcal{E}}_{s',d,t}) & \text{if } A(d) \in M \\ \mathsf{false} & \text{otherwise} \end{cases}$$

$$\delta(q^{\mathcal{E}}_{s,r,s',d,t}, \langle M, x, N\rangle) = \bigvee_{d'\in\Delta} \big((0, q_{r,d,d'}) \wedge (0, q^{\mathcal{E}}_{s',d',t})\big) \vee \bigvee_{i\in[k]} \big((i, q_d) \wedge (i, q^{\mathcal{E}}_{s,r,s',d,t})\big)$$

We use the parity condition to enforce that *no* state is allowed to occur infinitely often by setting $F = Q$.

Note that $\mathfrak{A}^1_q$ has size polynomial in q and exponential in $\mathcal{K}$ since its states depend on Δ. Now, $\mathfrak{A}_q$ is obtained from $\mathfrak{A}^0_q$ and $\mathfrak{A}^1_q$ as follows. First take the conjunction $\mathfrak{A}'_q$ of $\mathfrak{A}^0_q$ and $\mathfrak{A}^1_q$ (without blowup in the size). Then transform $\mathfrak{A}'_q$ into an equivalent nondeterministic tree automaton $\mathfrak{A}''_q$; thus, the size of $\mathfrak{A}''_q$ is exponentially bigger, that is, exponential in q and double exponential in $\mathcal{K}$. Finally, obtain $\mathfrak{A}_q$ from $\mathfrak{A}''_q$ by projecting away the extended alphabet, without increasing the size of the automaton. $\square$