

## BIROn - Birkbeck Institutional Research Online

Kikot, Stanislav and Kontchakov, Roman and Podolskii, V. and Zakharyashev, Michael (2013) Query rewriting over shallow ontologies. In: Eiter, T. and Glimm, B. and Kazakov, Y. and Krötzsch, M. (eds.) Informal Proceedings of DL (Ulm, 23-26 July). CEUR Workshop Proceedings 1014. CEUR-WS, pp. 316-327.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/10306/>

*Usage Guidelines:*

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>

or alternatively

contact [lib-eprints@bbk.ac.uk](mailto:lib-eprints@bbk.ac.uk).

# Query Rewriting over Shallow Ontologies

S. Kikot<sup>1</sup>, R. Kontchakov<sup>1</sup>, V. Podolskii<sup>2</sup>, and M. Zakharyashev<sup>1</sup>

<sup>1</sup> Department of Computer Science and Information Systems  
Birkbeck, University of London, U.K.

{kikot,roman,michael}@dcs.bbk.ac.uk

<sup>2</sup> Steklov Mathematical Institute, Moscow, Russia  
podolskii@mi.ras.ru

**Abstract.** We investigate the size of rewritings of conjunctive queries over *OWL 2 QL* ontologies of depth 1 and 2 by means of a new hypergraph formalism for computing Boolean functions. Both positive and negative results are obtained. All conjunctive queries over ontologies of depth 1 have polynomial-size nonrecursive datalog rewritings; tree-shaped queries have polynomial-size positive existential rewritings; however, for some queries and ontologies of depth 1, positive existential rewritings can only be of superpolynomial size. Both positive existential and nonrecursive datalog rewritings of conjunctive queries and ontologies of depth 2 suffer an exponential blowup in the worst case, while first-order rewritings can grow superpolynomially unless  $\text{NP} \subseteq \text{P/poly}$ .

## 1 Introduction

This paper is a continuation of the series [11, 13, 14], where we investigated the following problems. Let  $q(\mathbf{x})$  be a conjunctive query (CQ) with answer variables  $\mathbf{x}$  and let  $\mathcal{T}$  be an *OWL 2 QL* ontology. It is known (see, e.g., [7, 3]) that there exists a first-order formula  $q'(\mathbf{x})$ , called an FO-rewriting of  $q$  and  $\mathcal{T}$ , such that  $(\mathcal{T}, \mathcal{A}) \models q(\mathbf{a})$  iff  $\mathcal{A} \models q'(\mathbf{a})$ , for any ABox  $\mathcal{A}$  and any vector  $\mathbf{a}$  of individuals in  $\mathcal{A}$  (of the same length as  $\mathbf{x}$ ). Thus, to find certain answers to  $q(\mathbf{x})$  over  $\mathcal{T}$  and  $\mathcal{A}$ , it suffices to find answers to  $q'(\mathbf{x})$  over the data  $\mathcal{A}$ , which can (hopefully) be done by conventional relational database management systems (RDBMSs). Various experiments showed, however, that rewritings  $q'$  can be too large for the RDBMSs to cope with. This put forward the followings problems:

- What is the overhead of answering CQs via ontologies compared to standard database query answering in the worst case?
- What is the size of FO-rewritings of CQs and *OWL 2 QL* ontologies in the worst case?
- Can rewritings of one type (say, nonrecursive datalog) be substantially shorter than rewritings of another type (say, positive existential)?
- Are there interesting and useful sufficient conditions on CQs and ontologies under which rewritings are short?

We showed [11, 13, 14] that, for a certain sequence of (tree-shaped) CQs  $\mathbf{q}_n$  and OWL 2 QL TBoxes  $\mathcal{T}_n$ , the problem ‘ $\mathcal{A} \models \mathbf{q}_n$ ?’ is in P for combined complexity, while the problem ‘ $(\mathcal{T}_n, \mathcal{A}) \models \mathbf{q}_n$ ?’ is NP-complete. Moreover, any positive existential (PE) or nonrecursive datalog (NDL) rewriting of  $\mathbf{q}_n$  and  $\mathcal{T}_n$  is of exponential size, while any FO-rewriting is of superpolynomial size unless  $\text{NP} \subseteq \text{P/poly}$ . We also showed that NDL-rewritings are in general exponentially more succinct than PE-rewritings, and FO-rewritings can be superpolynomially more succinct than PE-rewritings. On the other hand, Gottlob and Schwentick [8] demonstrated that one can always find a polynomial-size rewriting for the price of polynomially-many additional existential quantifiers over a domain with at least two constants (thus confirming once again that formalisms with nondeterminism are exponentially more succinct; cf. [4]). Finally, Kikot *et al.* [12] give a practically useful sufficient condition on CQs and ontologies under which PE-rewritings are of polynomial size.

The problem we address in this paper is whether the depth of TBoxes (that is, the maximal depth of the canonical models with single-individual ABoxes) has any impact on the size of rewritings. (The TBoxes  $\mathcal{T}_n$  mentioned above are of depth  $n$ .) In particular, what happens if we restrict the depth of TBoxes to 1 or 2? (PE-rewritings over TBoxes of depth 0 are trivially polynomial.) The obtained results are summarised below:

- (1) For any CQ and TBox of depth 1, there is a polynomial-size NDL-rewriting.
- (2) PE-rewritings of some CQs and TBoxes of depth 1 are of superpolynomial size.
- (3) All tree-shaped CQs and TBoxes of depth 1 have polynomial-size PE-rewritings.
- (4) For TBoxes of depth 2, both NDL- and PE-rewritings can suffer an exponential blowup, while FO-rewritings can suffer a superpolynomial blowup (unless  $\text{NP} \subseteq \text{P/poly}$ ).

We begin by observing that the tree-witness PE-rewritings, representing possible homomorphisms of subqueries of a given CQ to the canonical models with one ABox individual, give rise to a class of monotone Boolean functions associated with hypergraphs and called *hypergraph functions*. In particular, hypergraphs  $H$  of degree  $\leq 2$  (every vertex in which belongs to at most 2 hyperedges) correspond to Boolean CQs  $\mathbf{q}_H$  and TBoxes  $\mathcal{T}_H$  of depth 1 such that answering  $\mathbf{q}_H$  over  $\mathcal{T}_H$  and single-individual ABoxes amounts to computing the hypergraph function for  $H$ . We show then that representing Boolean functions as hypergraphs of degree  $\leq 2$  is polynomially equivalent to representing them by nondeterministic branching programs (NBP) [10]. This correspondence and known results about NBPs [17, 9] give (1) and (2) above. We show (3) using the tree form of CQs and the fact that, over TBoxes of depth 1, CQs  $\mathbf{q}$  can only have  $\leq |\mathbf{q}|$  tree witnesses. To obtain (4), we observe that hypergraphs of degree  $> 2$  are computationally as powerful as nondeterministic polynomial Boolean circuits (NP/poly) and encode computing the function  $\text{CLIQUE}_{n,k}(\mathbf{e})$  (a graph with  $n$  vertices has a  $k$ -clique) as answering some CQs over TBoxes of depth 2. Although hypergraph representations of Boolean functions are introduced as a technical means to investigate the

size of rewritings, they may be of independent interest to the complexity theory of Boolean functions. All the omitted proofs can be found in the full version of the paper available at [www.dcs.bbk.ac.uk/~roman](http://www.dcs.bbk.ac.uk/~roman).

## 2 OWL 2 QL and the Tree-Witness Rewriting

In this paper, we use the following (simplified) syntax of *OWL 2 QL*. It contains *individual names*  $a_i$ , *concept names*  $A_i$ , and *role names*  $P_i$  ( $i \geq 1$ ). *Roles*  $R$  and *basic concepts*  $B$  are defined by the grammar:

$$R ::= P_i \mid P_i^-, \quad B ::= \perp \mid A_i \mid \exists R.$$

A *TBox*,  $\mathcal{T}$ , is a finite set of *inclusions* of the form

$$B_1 \sqsubseteq B_2, \quad B_1 \sqcap B_2 \sqsubseteq \perp, \quad R_1 \sqsubseteq R_2, \quad R_1 \sqcap R_2 \sqsubseteq \perp.$$

An *ABox*,  $\mathcal{A}$ , is a finite set of atoms of the form  $A_k(a_i)$  or  $P_k(a_i, a_j)$ . The set of individual names in  $\mathcal{A}$  is denoted by  $\text{ind}(\mathcal{A})$ .  $\mathcal{T}$  and  $\mathcal{A}$  together form the *knowledge base* (KB)  $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ . The semantics for *OWL 2 QL* is defined in the usual way based on interpretations  $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$  [5]. We write  $B_1 \equiv B_2$  as a shortcut for  $B_1 \sqsubseteq B_2$  and  $B_2 \sqsubseteq B_1$ .

For every role name  $R$  in  $\mathcal{T}$ , we take two fresh concept names  $A_R, A_{R^-}$  and add to  $\mathcal{T}$  the axioms  $A_R \equiv \exists R$  and  $A_{R^-} \equiv \exists R^-$ . We say that the resulting TBox is in *normal form* and assume, without loss of generality, that every TBox in this paper is in normal form. We denote by  $\sqsubseteq_{\mathcal{T}}$  the subsumption relation induced by  $\mathcal{T}$  and write  $S_1 \sqsubseteq_{\mathcal{T}} S_2$  if  $\mathcal{T} \models S_1 \sqsubseteq S_2$ , where  $S_1, S_2$  are both concepts or roles. We say that an ABox  $\mathcal{A}$  is *H-complete with respect to  $\mathcal{T}$*  in case

$$\begin{aligned} R_2(a, b) \in \mathcal{A} & \quad \text{if} \quad R_1(a, b) \in \mathcal{A} \text{ and } R_1 \sqsubseteq_{\mathcal{T}} R_2, \\ A_2(a) \in \mathcal{A} & \quad \text{if} \quad A_1(a) \in \mathcal{A} \text{ and } A_1 \sqsubseteq_{\mathcal{T}} A_2, \end{aligned}$$

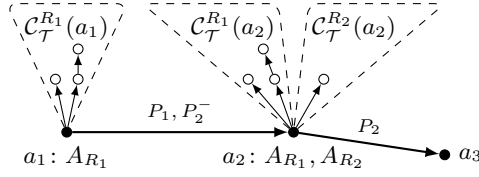
for all concept names  $A_i$  (including the  $A_R$ ) and roles  $R_i$ . We write  $R(a, b) \in \mathcal{A}$  for  $P(a, b) \in \mathcal{A}$  if  $R = P$  and for  $P(b, a)$  if  $R = P^-$ ; also, we write  $A_R(a) \in \mathcal{A}$  if  $R(a, b) \in \mathcal{A}$ , for some  $b$ .

A *conjunctive query* (CQ)  $\mathbf{q}(\mathbf{x})$  is a formula  $\exists \mathbf{y} \varphi(\mathbf{x}, \mathbf{y})$ , where  $\varphi$  is a conjunction of atoms of the form  $A_k(z_1)$  or  $P_k(z_1, z_2)$  with  $z_i \in \mathbf{x} \cup \mathbf{y}$  (without loss of generality, we assume that CQs do not contain constants). A tuple  $\mathbf{a} \subseteq \text{ind}(\mathcal{A})$  is a *certain answer* to  $\mathbf{q}(\mathbf{x})$  over  $\mathcal{K} = (\mathcal{T}, \mathcal{A})$  if  $\mathcal{I} \models \mathbf{q}(\mathbf{a})$  for all  $\mathcal{I} \models \mathcal{K}$ ; in this case we write  $\mathcal{K} \models \mathbf{q}(\mathbf{a})$ . If  $\mathbf{x} = \emptyset$ , the CQ  $\mathbf{q}$  is called *Boolean*; a certain answer to such a  $\mathbf{q}$  over  $\mathcal{K}$  is ‘yes’ if  $\mathcal{K} \models \mathbf{q}$  and ‘no’ otherwise. Where convenient, we regard a CQ  $\mathbf{q}$  as the set of its atoms.

Suppose  $\mathcal{T}$  is a TBox and  $\mathbf{q}(\mathbf{x})$  a CQ. An FO-formula  $\mathbf{q}'(\mathbf{x})$  with free variables  $\mathbf{x}$  and without constants is an *FO-rewriting of  $\mathbf{q}$  and  $\mathcal{T}$  over H-complete ABoxes* if, for any H-complete (with respect to  $\mathcal{T}$ ) ABox  $\mathcal{A}$  and any  $\mathbf{a} \subseteq \text{ind}(\mathcal{A})$ , we have  $(\mathcal{T}, \mathcal{A}) \models \mathbf{q}(\mathbf{a})$  iff  $\mathcal{A} \models \mathbf{q}'(\mathbf{a})$ . If an FO-rewriting  $\mathbf{q}'$  is a positive existential formula (with only  $\exists, \wedge, \vee$ ), we call it a *PE-rewriting of  $\mathbf{q}$  and  $\mathcal{T}$* . We also consider

rewritings in the form of nonrecursive Datalog queries. We remind the reader that a *Datalog program*,  $\Pi$ , is a finite set of Horn clauses  $\forall \mathbf{x} (\gamma_1 \wedge \dots \wedge \gamma_m \rightarrow \gamma_0)$ , where each  $\gamma_i$  is an atom of the form  $P(x_1, \dots, x_l)$  with  $x_i \in \mathbf{x}$ . The atom  $\gamma_0$  is called the *head* of the clause, and  $\gamma_1, \dots, \gamma_m$  its *body*. All variables in the head must also occur in the body. A predicate  $P$  *depends* on a predicate  $Q$  in  $\Pi$  if  $\Pi$  contains a clause whose head is  $P$  and whose body contains  $Q$ .  $\Pi$  is called *nonrecursive* if this dependence relation for  $\Pi$  is acyclic. For a nonrecursive Datalog program  $\Pi$  and an atom  $\mathbf{q}'(\mathbf{x})$ , we say that  $(\Pi, \mathbf{q}')$  is an *NDL-rewriting of  $\mathbf{q}(\mathbf{x})$  and  $\mathcal{T}$  over H-complete ABoxes* in case  $(\mathcal{T}, \mathcal{A}) \models \mathbf{q}(\mathbf{a})$  iff  $\Pi, \mathcal{A} \models \mathbf{q}'(\mathbf{a})$ , for any H-complete ABox  $\mathcal{A}$  and any  $\mathbf{a} \subseteq \text{ind}(\mathcal{A})$ . Rewritings *over arbitrary ABoxes* are defined by dropping the condition that the ABoxes are H-complete. They can be obtained from rewritings over H-complete ABoxes at the price of a polynomial blowup.

Recall [7, 12] that, for any TBox  $\mathcal{T}$  and ABox  $\mathcal{A}$ , there is a *canonical model*  $\mathcal{C}_{\mathcal{T}, \mathcal{A}}$  of  $(\mathcal{T}, \mathcal{A})$  such that  $(\mathcal{T}, \mathcal{A}) \models \mathbf{q}(\mathbf{a})$  iff  $\mathcal{C}_{\mathcal{T}, \mathcal{A}} \models \mathbf{q}(\mathbf{a})$ , for all CQs  $\mathbf{q}(\mathbf{x})$  and  $\mathbf{a} \subseteq \text{ind}(\mathcal{A})$ . The domain of  $\mathcal{C}_{\mathcal{T}, \mathcal{A}}$  consists of the individuals in  $\text{ind}(\mathcal{A})$  and the witnesses introduced by the existential quantifiers in  $\mathcal{T}$ . Every individual  $a \in \text{ind}(\mathcal{A})$  with  $(\mathcal{T}, \mathcal{A}) \models A_R(a)$  is the root of a (possibly infinite) subtree  $\mathcal{C}_{\mathcal{T}}^R(a)$  of  $\mathcal{C}_{\mathcal{T}, \mathcal{A}}$ , which may intersect another such tree only on their common root  $a$ . Every  $\mathcal{C}_{\mathcal{T}}^R(a)$  is isomorphic to the canonical model of  $(\mathcal{T}, \{A_R(a)\})$ .



We say that  $\mathcal{T}$  is of *depth*  $\omega$  if at least one of  $\mathcal{C}_{\mathcal{T}}^R(a)$  is infinite;  $\mathcal{T}$  is of *depth*  $d$ ,  $0 \leq d < \omega$ , if there is a chain of the form  $aR_0w_1 \dots w_{d-1}R_{d-1}w_d$  (for distinct  $w_i$ ) in the trees  $\mathcal{C}_{\mathcal{T}}^R(a)$ ,  $R$  a role in  $\mathcal{T}$ , but there is no such chain of greater length.

By definition,  $\mathcal{C}_{\mathcal{T}, \mathcal{A}} \models \mathbf{q}(\mathbf{a})$  iff there is a homomorphism  $h: \mathbf{q}(\mathbf{a}) \rightarrow \mathcal{C}_{\mathcal{T}, \mathcal{A}}$ . Such a homomorphism  $h$  splits  $\mathbf{q}$  into the subquery mapped by  $h$  to  $\text{ind}(\mathcal{A})$  and the subquery mapped to the trees  $\mathcal{C}_{\mathcal{T}}^R(a)$ . We can think of a rewriting of  $\mathbf{q}$  and  $\mathcal{T}$  as listing possible splits of  $\mathbf{q}$  into such subqueries.

Suppose  $\mathbf{q}'$  is a subset of the atoms of  $\mathbf{q}$  and there is a homomorphism  $h: \mathbf{q}' \rightarrow \mathcal{C}_{\mathcal{T}}^R(a)$ , for some  $a$ , such that  $h$  maps all answer variables in  $\mathbf{q}'$  to  $a$ . Let  $\mathfrak{t}_r = h^{-1}(a)$  and let  $\mathfrak{t}_i$  be the remaining set of (existentially quantified) variables in  $\mathbf{q}'$ . We call the pair  $\mathfrak{t} = (\mathfrak{t}_r, \mathfrak{t}_i)$  a *tree witness for  $\mathbf{q}$  and  $\mathcal{T}$  generated by  $R$*  if  $\mathfrak{t}_i \neq \emptyset$  and  $\mathbf{q}'$  is a *minimal* subset of  $\mathbf{q}$  such that, for any  $y \in \mathfrak{t}_i$ , every atom in  $\mathbf{q}$  containing  $y$  belongs to  $\mathbf{q}'$ . In this case, we denote  $\mathbf{q}'$  by  $\mathbf{q}_{\mathfrak{t}}$ . By definition,

$$\mathbf{q}_{\mathfrak{t}} = \{S(\mathbf{z}) \in \mathbf{q} \mid \mathbf{z} \subseteq \mathfrak{t}_r \cup \mathfrak{t}_i \text{ and } \mathbf{z} \not\subseteq \mathfrak{t}_r\}.$$

Note that the same tree witness  $\mathfrak{t} = (\mathfrak{t}_r, \mathfrak{t}_i)$  can be generated by different roles  $R$ . We denote the set of all such roles by  $\Omega_{\mathfrak{t}}$  and define the formula

$$\text{tw}_{\mathfrak{t}} = \bigvee_{R \in \Omega_{\mathfrak{t}}} \exists z (A_R(z) \wedge \bigwedge_{x \in \mathfrak{t}_r} (x = z)). \quad (1)$$

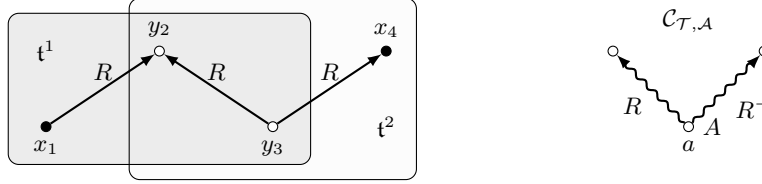
Tree witnesses  $\mathfrak{t}$  and  $\mathfrak{t}'$  are *consistent* if  $\mathfrak{q}_{\mathfrak{t}} \cap \mathfrak{q}_{\mathfrak{t}'} = \emptyset$ . Each *consistent set*  $\Theta$  of tree witnesses (in which any pair of distinct tree witnesses is consistent) determines a subquery  $\mathfrak{q}_{\Theta}$  of  $\mathfrak{q}$  that comprises all atoms of  $\mathfrak{q}_{\mathfrak{t}}$ , for  $\mathfrak{t} \in \Theta$ . The subquery  $\mathfrak{q}_{\Theta}$  is to be mapped to the  $\mathcal{C}_{\mathcal{T}}^R(a)$ , whereas the remainder,  $\mathfrak{q} \setminus \mathfrak{q}_{\Theta}$ , obtained by removing the atoms of  $\mathfrak{q}_{\Theta}$  from  $\mathfrak{q}$ , is mapped to  $\text{ind}(\mathcal{A})$ . The following PE-formula  $\mathfrak{q}_{\text{tw}}$  is called the *tree-witness rewriting* of  $\mathfrak{q}$  and  $\mathcal{T}$  over H-complete ABoxes:

$$\mathfrak{q}_{\text{tw}}(\mathbf{x}) = \bigvee_{\Theta \text{ consistent}} \exists \mathbf{y} \left( (\mathfrak{q} \setminus \mathfrak{q}_{\Theta}) \wedge \bigwedge_{\mathfrak{t} \in \Theta} \text{tw}_{\mathfrak{t}} \right). \quad (2)$$

*Example 1.* Consider the KB  $\mathcal{K} = (\mathcal{T}, \{A(a)\})$ , where

$$\mathcal{T} = \{A \sqsubseteq \exists R, A \sqsubseteq \exists R^-, A_R \equiv \exists R, A_{R^-} \equiv \exists R^-\},$$

and the CQ  $\mathfrak{q}(x_1, x_4) = \{R(x_1, y_2), R(y_3, y_2), R(y_3, x_4)\}$  shown in the picture below alongside the canonical model  $\mathcal{C}_{\mathcal{T}, \mathcal{A}}$  (with  $A_R$  and  $A_{R^-}$  omitted).



There are two tree witnesses for  $\mathfrak{q}$  and  $\mathcal{T}$ :  $\mathfrak{t}^1 = (\mathfrak{t}_r^1, \mathfrak{t}_i^1)$  generated by  $R$  and  $\mathfrak{t}^2 = (\mathfrak{t}_r^2, \mathfrak{t}_i^2)$  generated by  $R^-$ , with

$$\begin{aligned} \mathfrak{t}_r^1 &= \{x_1, y_3\}, & \mathfrak{t}_i^1 &= \{y_2\}, & \text{tw}_{\mathfrak{t}^1} &= \exists z (A_R(z) \wedge (x_1 = z) \wedge (y_3 = z)), \\ \mathfrak{t}_r^2 &= \{y_2, x_4\}, & \mathfrak{t}_i^2 &= \{y_3\}, & \text{tw}_{\mathfrak{t}^2} &= \exists z (A_{R^-}(z) \wedge (x_4 = z) \wedge (y_2 = z)). \end{aligned}$$

We have  $\mathfrak{q}_{\mathfrak{t}^1} = \{R(x_1, y_2), R(y_3, y_2)\}$  and  $\mathfrak{q}_{\mathfrak{t}^2} = \{R(y_3, y_2), R(y_3, x_4)\}$ , so  $\mathfrak{t}^1$  and  $\mathfrak{t}^2$  are inconsistent. Thus, we obtain the following tree-witness rewriting:

$$\mathfrak{q}_{\text{tw}}(x_1, x_4) = \exists y_2, y_3 [(R(x_1, y_2) \wedge R(y_3, y_2) \wedge R(y_3, x_4)) \vee (R(y_3, x_4) \wedge \text{tw}_{\mathfrak{t}^1}) \vee (R(x_1, y_2) \wedge \text{tw}_{\mathfrak{t}^2})].$$

**Theorem 1.** *For any ABox  $\mathcal{A}$  that is H-complete with respect to  $\mathcal{T}$  and any  $\mathbf{a} \subseteq \text{ind}(\mathcal{A})$ , we have  $\mathcal{C}_{\mathcal{T}, \mathcal{A}} \models \mathfrak{q}(\mathbf{a})$  iff  $\mathcal{A} \models \mathfrak{q}_{\text{tw}}(\mathbf{a})$ .*

### 3 Hypergraph Functions

The tree-witness rewriting  $\mathfrak{q}_{\text{tw}}$  above gives rise to monotone Boolean functions we call hypergraph functions. For the complexity theory of monotone Boolean functions the reader is referred to [2, 10].

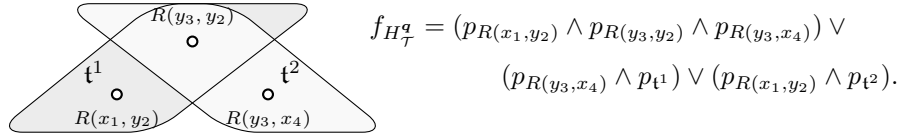
Let  $H = (V, E)$  be a hypergraph with *vertices*  $v \in V$  and *hyperedges*  $e \in E$ ,  $E \subseteq 2^V$ . We call a subset  $X \subseteq E$  *independent* if  $e \cap e' = \emptyset$ , for any distinct

$e, e' \in X$ . The set of vertices that occur in the hyperedges of  $X$  is denoted by  $V_X$ . With each vertex  $v \in V$  and each hyperedge  $e \in E$  we associate propositional variables  $p_v$  and  $p_e$ , respectively. The *hypergraph function*  $f_H$  for  $H$  is given by the Boolean formula

$$f_H = \bigvee_{X \text{ independent}} \left( \bigwedge_{v \in V \setminus V_X} p_v \wedge \bigwedge_{e \in X} p_e \right). \quad (3)$$

The rewriting  $\mathbf{q}_{\text{tw}}$  of  $\mathbf{q}$  and  $\mathcal{T}$  defines a hypergraph whose vertices are the atoms of  $\mathbf{q}$  and hyperedges are the sets  $\mathbf{q}_{\mathbf{t}}$ , for tree witnesses  $\mathbf{t}$  for  $\mathbf{q}$  and  $\mathcal{T}$ . We denote this hypergraph by  $H_{\mathcal{T}}^{\mathbf{q}}$ . The formula (3) defining  $f_{H_{\mathcal{T}}^{\mathbf{q}}}$  is basically the same as the rewriting (2) with the atoms  $S(\mathbf{z})$  in  $\mathbf{q}$  and tree-witness formulas  $\text{tw}_{\mathbf{t}}$  treated as propositional variables. We denote these variables by  $p_{S(\mathbf{z})}$  and  $p_{\mathbf{t}}$  (rather than  $p_v$  and  $p_e$ ), respectively.

*Example 2.* Consider again the CQ  $\mathbf{q}$  and TBox  $\mathcal{T}$  from Example 1. The hypergraph  $H_{\mathcal{T}}^{\mathbf{q}}$  and its hypergraph function  $f_{H_{\mathcal{T}}^{\mathbf{q}}}$  are shown below:



Suppose the function  $f_{H_{\mathcal{T}}^{\mathbf{q}}}$  is computed by some Boolean formula  $\chi_{H_{\mathcal{T}}^{\mathbf{q}}}$ . Consider the FO-formula  $\widehat{\chi}_{H_{\mathcal{T}}^{\mathbf{q}}}$  obtained by replacing each  $p_{S(\mathbf{z})}$  in  $\chi_{H_{\mathcal{T}}^{\mathbf{q}}}$  with  $S(\mathbf{z})$ , each  $p_{\mathbf{t}}$  with  $\text{tw}_{\mathbf{t}}$ , and adding the prefix  $\exists \mathbf{y}$ . By comparing (3) and (2), we see that the resulting FO-formula is a rewriting of  $\mathbf{q}$  and  $\mathcal{T}$  over H-complete ABoxes.

**Theorem 2.** (i) If the function  $f_{H_{\mathcal{T}}^{\mathbf{q}}}$  is computed by a Boolean formula  $\chi_{H_{\mathcal{T}}^{\mathbf{q}}}$ , then  $\widehat{\chi}_{H_{\mathcal{T}}^{\mathbf{q}}}$  is an FO-rewriting of  $\mathbf{q}$  and  $\mathcal{T}$  over H-complete ABoxes.

(ii) If  $f_{H_{\mathcal{T}}^{\mathbf{q}}}$  is computed by a monotone Boolean circuit  $\mathbf{C}$ , then there is an NDL-rewriting of  $\mathbf{q}$  and  $\mathcal{T}$  over H-complete ABoxes of size  $O(|\mathbf{C}| \cdot (|\mathbf{q}| + |\mathcal{T}|))$ .

Thus, the problem of constructing short rewritings is reducible to the problem of finding short Boolean formulas computing the hypergraph functions. Hypergraphs of *degree*  $\leq 2$ , in which every vertex belongs to at most two hyperedges, are of particular interest to us because (i) TBoxes of depth 1 have hypergraphs of degree  $\leq 2$ , and (ii) their hypergraph functions are the functions computed by nondeterministic branching programs (NBPs), also known as switching-and-rectifier networks [10] of polynomial size. Recall [17] that, in terms of the expressive power, NBPs sit between Boolean formulas and Boolean circuits. On the other hand, hypergraphs of unbounded degree are substantially more powerful than hypergraphs of degree 2; more precisely, polynomial-size hypergraphs of unbounded degree can define NP-hard Boolean functions. Here is an example.

We remind the reader (see, e.g., [2] for details) that the monotone Boolean function  $\text{CLIQUE}_{n,k}(\mathbf{e})$  of  $n(n-1)/2$  variables  $e_{jj'}$ ,  $1 \leq j < j' \leq n$ , returns 1 iff the graph with vertices  $\{1, \dots, n\}$  and edges  $\{\{i, j\} \mid e_{jj'} = 1\}$  contains a  $k$ -clique. Clearly,  $\text{CLIQUE}_{n,k}$  is NP-complete. A series of papers, started by

Razborov's [16], gave an exponential lower bound for the size of monotone circuits computing  $\text{CLIQUE}_{n,k}$ :  $2^{\Omega(\sqrt{k})}$  for  $k \leq \frac{1}{4}(n/\log n)^{2/3}$  [1]. For monotone formulas, an even better lower bound is known:  $2^{\Omega(k)}$  for  $k = 2n/3$  [15].

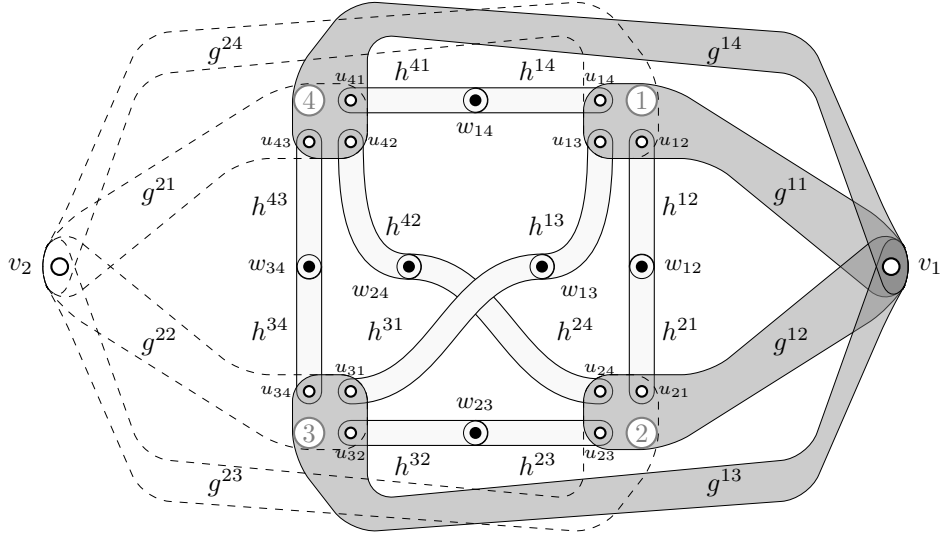
Given  $n$  and  $k$  as above, let  $H_{n,k}$  be a hypergraph with the vertices

$$\{v_i \mid 1 \leq i \leq k\} \cup \{u_{jj'}, w_{jj'}, u_{j'j} \mid 1 \leq j < j' \leq n\}$$

and the hyperedges

- $g^{ij} = \{v_i\} \cup \{u_{jj'} \mid 1 \leq j' \leq n, j' \neq j\}$ , for  $1 \leq i \leq k, 1 \leq j \leq n$ ,
- $h^{jj'} = \{w_{jj'}, u_{jj'}\}$  and  $h^{j'j} = \{w_{jj'}, u_{j'j}\}$ , for  $1 \leq j < j' \leq n$ .

Informally, the vertices  $v_i$  of  $H_{n,k}$  represent a  $k$ -clique in a given graph with  $n$  vertices. The vertices  $w_{jj'}$  represent the edges of the complete graph with  $n$  vertices; they can be turned 'on' or 'off' by means of the Boolean variables  $e_{jj'}$ . The vertex  $u_{jj'}$  together with the hyperedge  $h^{jj'}$  represent the 'half' of the edge connecting  $j$  and  $j'$  that is adjacent to  $j$ . The hyperedges  $g^{ij}$  correspond to the choice of the  $j$ th vertex of the graph as the  $i$ th element of the clique. For example, the hypergraph  $H_{4,2}$  looks as follows:



Denote by  $\mathbf{v}$ ,  $\mathbf{u}$ ,  $\mathbf{w}$ ,  $\mathbf{g}$  and  $\mathbf{h}$  the vectors of the  $v_i$ ,  $u_{jj'}$ ,  $w_{jj'}$ ,  $g^{ij}$  and  $h^{jj'}$ , respectively. Thus, we can think of the hypergraph function for  $H_{n,k}$  as having the form  $f_{H_{n,k}}(\mathbf{v}, \mathbf{u}, \mathbf{w}, \mathbf{g}, \mathbf{h})$ .

**Theorem 3.**  $\text{CLIQUE}_{n,k}(e) = f_{H_{n,k}}(\mathbf{0}, \mathbf{1}, e, \mathbf{1}, \mathbf{1})$ , for all  $e \in \{0, 1\}^{n(n-1)/2}$ .

## 4 The Size of Rewritings over TBoxes of Depth 1 and 2

We use these complexity-theoretic results on the hypergraph functions to investigate the size of rewritings over TBoxes of depth 1 and 2.



**Theorem 4.** *For any CQ  $\mathbf{q}$  and any TBox  $\mathcal{T}$  of depth 1, there is an NDL-rewriting of  $\mathbf{q}$  and  $\mathcal{T}$  of polynomial size.*

The proof of this theorem is based on the following facts. First, for any CQ  $\mathbf{q}$  and TBox  $\mathcal{T}$  of depth 1, the hypergraph  $H_{\mathcal{T}}^{\mathbf{q}}$  is of degree  $\leq 2$ , and the number of distinct tree witnesses for  $\mathbf{q}$  and  $\mathcal{T}$  does not exceed the number of variables in  $\mathbf{q}$ . Second, each hypergraph function given by a hypergraph of degree  $\leq 2$  is computed by a polynomial-size monotone NBP, and so by a polynomial size monotone Boolean circuit. Thus, it remains to use Theorem 2.

**Theorem 5.** *There is a sequence of CQs  $\mathbf{q}_n$  and TBoxes  $\mathcal{T}_n$  of depth 1 such that any PE-rewriting of  $\mathbf{q}_n$  and  $\mathcal{T}_n$  (over  $H$ -complete ABoxes) is of size  $2^{\Omega(\log^2 n)}$ .*

In the proof of this result, we use a sequence of Boolean functions  $f_n(\mathbf{x})$  that are computable by polynomial-size monotone NBPs, but any monotone Boolean formulas computing  $f_n$  are of size  $\geq 2^{\Omega(\log^2 n)}$ . (For example, Grigni and Sipser [9] consider  $f_n(\mathbf{x})$  that takes the adjacency matrix of a directed graph of  $n$  vertices with a distinguished vertex  $s$  as input and returns 1 iff there is a directed path from  $s$  to some vertex of outdegree at least two.) Since any NBP corresponds to a hypergraph of degree  $\leq 2$  of a polynomial size, we obtain a sequence  $H_n$  of polynomial hypergraphs of degree  $\leq 2$  such that  $f_n = f_{H_n}$ .

Now, we call a hypergraph  $H$  *representable* if there are a CQ  $\mathbf{q}_H$  and a TBox  $\mathcal{T}_H$  such that  $H$  is isomorphic to  $H_{\mathcal{T}_H}^{\mathbf{q}_H}$ . To show that any hypergraph of degree  $\leq 2$  is representable, take any such  $H = (V, E)$  and assume, to simplify notation, that any vertex in  $H$  belongs to exactly 2 hyperedges, so  $H$  comes with two fixed maps  $i_1, i_2: V \rightarrow E$  such that  $i_1(v) \neq i_2(v)$ ,  $v \in i_1(v)$  and  $v \in i_2(v)$ , for any  $v \in V$ . For every  $e \in E$ , we take an individual variable  $z_e$  and denote by  $\mathbf{z}$  the vector of all such variables. For every  $v \in V$ , we take a role name  $R_v$  and set:

$$\mathbf{q}_H = \exists \mathbf{z} \bigwedge_{v \in V} R_v(z_{i_1(v)}, z_{i_2(v)}),$$

For every hyperedge  $e \in E$ , let  $A_e$  be a concept name and  $S_e$  a role name. Consider the TBox  $\mathcal{T}_H$  with the following inclusions, for  $e \in E$ :

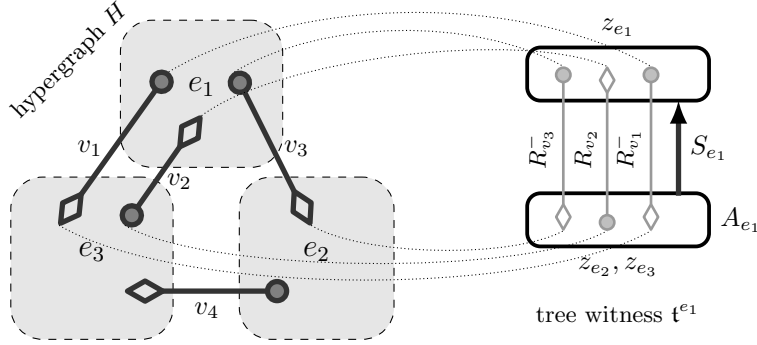
$$\begin{aligned} A_e &\equiv \exists S_e, \\ S_e &\sqsubseteq R_v^-, && \text{for } v \in V \text{ with } i_1(v) = e, \\ S_e &\sqsubseteq R_v, && \text{for } v \in V \text{ with } i_2(v) = e. \end{aligned}$$

We claim that  $H$  is isomorphic to  $H_{\mathcal{T}_H}^{\mathbf{q}_H}$  and illustrate the proof by an example.

*Example 3.* Let  $H = (V, E)$  with  $V = \{v_1, v_2, v_3, v_4\}$  and  $E = \{e_1, e_2, e_3\}$ , where  $e_1 = \{v_1, v_2, v_3\}$ ,  $e_2 = \{v_3, v_4\}$ ,  $e_3 = \{v_1, v_2, v_4\}$ . Suppose also that

$$\begin{aligned} i_1: v_1 \mapsto e_1, \quad v_2 \mapsto e_3, \quad v_3 \mapsto e_1, \quad v_4 \mapsto e_2, \\ i_2: v_1 \mapsto e_3, \quad v_2 \mapsto e_1, \quad v_3 \mapsto e_2, \quad v_4 \mapsto e_3. \end{aligned}$$

The hypergraph  $H$  is shown below, where each vertex  $v_i$  is represented by an edge:  $i_1(v_i)$  is indicated by the circle-shaped end of the edge and  $i_2(v_i)$  by the diamond-shaped one; the hyperedges  $e_j$  are shown as large grey squares:



$$\mathbf{q}_H = \exists z_{e_1} z_{e_2} z_{e_3} (R_{v_1}(z_{e_1}, z_{e_3}) \wedge R_{v_2}(z_{e_3}, z_{e_1}) \wedge R_{v_3}(z_{e_1}, z_{e_2}) \wedge R_{v_4}(z_{e_2}, z_{e_3}))$$

and the TBox  $\mathcal{T}_H$  contains the following inclusions:

$$\begin{array}{llll} A_{e_1} \equiv \exists S_{e_1}, & S_{e_1} \sqsubseteq R_{v_1}^-, & S_{e_1} \sqsubseteq R_{v_2}, & S_{e_1} \sqsubseteq R_{v_3}^-, \\ A_{e_2} \equiv \exists S_{e_2}, & S_{e_2} \sqsubseteq R_{v_3}, & S_{e_2} \sqsubseteq R_{v_4}^-, & \\ A_{e_3} \equiv \exists S_{e_3}, & S_{e_3} \sqsubseteq R_{v_1}, & S_{e_3} \sqsubseteq R_{v_2}^-, & S_{e_3} \sqsubseteq R_{v_4}. \end{array}$$

The canonical model  $\mathcal{C}_{\mathcal{T}_H}^{S_{e_1}}(a)$  is shown on the right-hand side of the picture above. We observe now that each variable  $z_e$  uniquely determines the tree witness  $t^e$  with  $\mathbf{q}_{t^e} = \{R_v(z_{i_1(v)}, z_{i_2(v)}) \mid v \in e\}$ ;  $\mathbf{q}_{t^e}$  and  $\mathbf{q}_{t^{e'}}$  are consistent iff  $e \cap e' \neq \emptyset$ . It follows that  $H$  is isomorphic to  $H_{\mathcal{T}_H}^{\mathbf{q}_H}$ .

We take the sequence of CQs  $\mathbf{q}_n$  and TBoxes  $\mathcal{T}_n$  associated with the hypergraphs of degree  $\leq 2$  for the sequence  $f_n$  of Boolean functions chosen above. Then we show that a PE-rewriting  $\mathbf{q}'_n$  of  $\mathbf{q}_n$  and  $\mathcal{T}_n$  can be transformed (using quantifier elimination over single-individual ABoxes [13]) into monotone Boolean formulas  $\chi_n$  computing  $f_n$  and having size  $\leq |\mathbf{q}'_n|$ . It remains to recall that  $|\mathbf{q}'_n| \geq |\chi_n| \geq 2^{\Omega(\log^2 n)}$ .

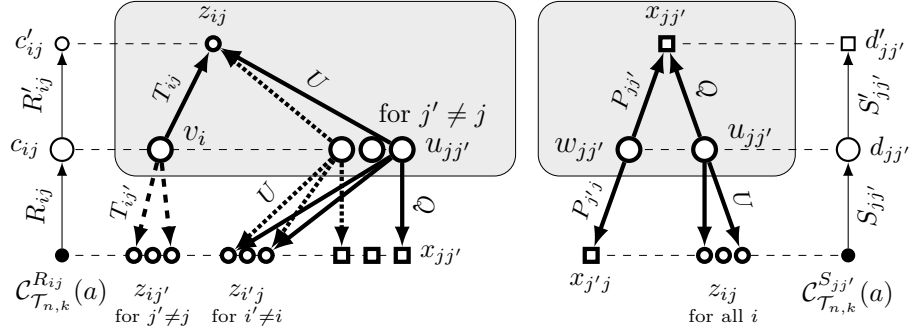
**Theorem 6.** *There exists a sequence of CQs  $\mathbf{q}_n$  and TBoxes  $\mathcal{T}_n$  of depth 2 such that any PE- and NDL-rewriting of  $\mathbf{q}_n$  and  $\mathcal{T}_n$  is of exponential size, while any FO-rewriting of  $\mathbf{q}_n$  and  $\mathcal{T}_n$  is of superpolynomial size (unless  $\text{NP} \subseteq \text{P/poly}$ ).*

To prove this theorem, we show that the hypergraphs  $H_{n,k}$  computing  $\text{CLIQUE}_{n,k}$  are representable as subgraphs of  $H_{\mathcal{T}_{n,k}}^{\mathbf{q}_{n,k}}$  for suitable  $\mathbf{q}_{n,k}$  and  $\mathcal{T}_{n,k}$ , and then use quantifier elimination over single-individual ABoxes as above.

The CQ  $\mathbf{q}_{n,k}$  and the TBox  $\mathcal{T}_{n,k}$  of polynomial size (in  $n$ ) that ‘compute’  $\text{CLIQUE}_{n,k}$  are defined as follows. The Boolean CQ  $\mathbf{q}_{n,k}$  contains the atoms:

$$\begin{array}{ll} T_{ij}(v_i, z_{ij}), & \text{for } 1 \leq i \leq k \text{ and } 1 \leq j \leq n, \\ P_{jj'}(w_{jj'}, x_{jj'}), P_{j'j}(w_{jj'}, x_{j'j}), & \text{for } 1 \leq j < j' \leq n, \\ Q(u_{jj'}, x_{jj'}), U(u_{jj'}, z_{ij}), & \text{for } 1 \leq j \neq j' \leq n \text{ and } 1 \leq i \leq k. \end{array}$$

The picture below illustrates the fragments of  $\mathbf{q}_{n,k}$  centred around  $z_{ij}$  and  $x_{jj'}$ :



The TBox  $\mathcal{T}_{n,k}$  mimics the arrangement of atoms in the layers depicted above and contains the following inclusions: for  $1 \leq i \leq k$  and  $1 \leq j \neq j' \leq n$ ,

$$\begin{aligned}
A_{ij} &\equiv \exists R_{ij}, & R_{ij} &\sqsubseteq T_{ij}^-, & R_{ij} &\sqsubseteq U^-, & R_{ij} &\sqsubseteq Q^-, & \exists R_{ij}^- &\sqsubseteq A'_{ij} \\
A'_{ij} &\equiv \exists R'_{ij}, & R'_{ij} &\sqsubseteq T_{ij}, & R'_{ij} &\sqsubseteq U, \\
B_{jj'} &\equiv \exists S_{jj'}, & S_{jj'} &\sqsubseteq P_{jj'}^-, & S_{jj'} &\sqsubseteq U^-, & \exists S_{jj'}^- &\sqsubseteq B'_{jj'}, \\
B'_{jj'} &\equiv \exists S'_{jj'}, & S'_{jj'} &\sqsubseteq P_{jj'}, & S'_{jj'} &\sqsubseteq Q.
\end{aligned}$$

The picture above also shows the elements and ‘generating roles’ of the models  $\mathcal{C}_{\mathcal{T}_{n,k}}^{R_{ij}}(a)$  and  $\mathcal{C}_{\mathcal{T}_{n,k}}^{S_{jj'}}(a)$ . The horizontal dashed lines show possible ways of embedding the fragments of  $\mathbf{q}_{n,k}$  into the respective canonical models. These embeddings give rise to the following tree witnesses for  $\mathbf{q}_{n,k}$  and  $\mathcal{T}_{n,k}$ :

- $\mathfrak{t}^{ij} = (\mathfrak{t}_r^{ij}, \mathfrak{t}_i^{ij})$  generated by  $R_{ij}$ , for  $1 \leq i \leq k$  and  $1 \leq j \leq n$ , where
$$\begin{aligned}
\mathfrak{t}_r^{ij} &= \{z_{ij'}, x_{jj'} \mid 1 \leq j' \leq n, j' \neq j\} \cup \{z_{i'j} \mid 1 \leq i' \leq k, i' \neq i\}, \\
\mathfrak{t}_i^{ij} &= \{v_i, z_{ij}\} \cup \{u_{jj'} \mid 1 \leq j' \leq n, j' \neq j\};
\end{aligned}$$
- $\mathfrak{s}^{jj'} = (\mathfrak{s}_r^{jj'}, \mathfrak{s}_i^{jj'})$  and  $\mathfrak{s}^{j'j} = (\mathfrak{s}_r^{j'j}, \mathfrak{s}_i^{j'j})$ , generated by  $S_{jj'}$  and  $S_{j'j}$ , where
$$\begin{aligned}
\mathfrak{s}_r^{jj'} &= \{x_{j'j}\} \cup \{z_{ij} \mid 1 \leq i \leq k\}, & \mathfrak{s}_r^{j'j} &= \{x_{jj'}\} \cup \{z_{ij'} \mid 1 \leq i \leq k\}, \\
\mathfrak{s}_i^{jj'} &= \{w_{jj'}, u_{jj'}, x_{jj'}\}, & \mathfrak{s}_i^{j'j} &= \{w_{j'j}, u_{j'j}, x_{j'j}\}.
\end{aligned}$$

The tree witnesses  $\mathfrak{t}^{ij}$ ,  $\mathfrak{s}^{jj'}$  and  $\mathfrak{s}^{j'j}$  are uniquely determined by their most remote (from the root) variable,  $z_{ij}$ ,  $x_{jj'}$  and  $x_{j'j}$ , respectively, and correspond to the hyperedges  $f^{ij}$ ,  $h^{jj'}$ ,  $h^{j'j}$  of the hypergraph  $H_{n,k}$ ; their internal variables of the form  $v_i$ ,  $w_{jj'}$  and  $u_{jj'}$  correspond to the vertices in the respective hyperedge.

## 5 Tree-Shaped Queries and TBoxes of Depth 1

In this section, we consider one particular equivalent transformation of the tree-witness rewriting. Let  $\mathbf{q}$  be a CQ and  $\mathbf{z}$  a subset of the set of existentially quantified variables in  $\mathbf{q}$ . By a  $\mathbf{z}$ -partition of  $\mathbf{q}$  we understand any disjoint sets of atoms  $\mathbf{q}_1, \dots, \mathbf{q}_k$ , called  $\mathbf{z}$ -components, that cover  $\mathbf{q}$  and are such that if, for

$i = 1, 2$ , each of  $S_i(\mathbf{z}_i) \in \mathbf{q}$  contains a variable  $z_i \in \mathbf{z}$  and  $z_1$  is connected to  $z_2$  by a path coming through variables in  $\mathbf{z}$  only, then both  $S_i(\mathbf{z}_i)$  are in the same component. Note that, for any  $\mathbf{z}$ -partition of  $\mathbf{q}$ , every tree witness for  $\exists \mathbf{z} \mathbf{q}$  and any  $\mathcal{T}$  is contained in some  $\mathbf{z}$ -component of the partition.

Given a TBox  $\mathcal{T}$  and a CQ  $\mathbf{q}(\mathbf{x}) = \exists \mathbf{y} \varphi(\mathbf{x}, \mathbf{y})$ , we define a CQ  $\mathbf{q}^\dagger(\mathbf{x}) = \exists \mathbf{y} \mathbf{q}^\mathbf{y}$ , where  $\mathbf{q}^\mathbf{z}$  is computed recursively as follows: we take a  $\mathbf{z}$ -partition  $\mathbf{q}_1, \dots, \mathbf{q}_k$  of  $\mathbf{q}$ , take  $\mathbf{z}_j$  to be the set of the variables in  $\mathbf{z}$  that occur in  $\mathbf{q}_j$ , for each  $1 \leq j \leq k$ , ( $\mathbf{z}_1, \dots, \mathbf{z}_k$  form a partition of  $\mathbf{z}$ ) and set  $\mathbf{q}^\mathbf{z} = \mathbf{q}_1^{*\mathbf{z}_1} \wedge \dots \wedge \mathbf{q}_k^{*\mathbf{z}_k}$ , where

$$\mathbf{q}_j^{*\mathbf{z}_j} = \begin{cases} (\mathbf{q}_j)^{\mathbf{z}_j \setminus \{z_j\}} \vee \bigvee_{\substack{\text{tree witness } \mathfrak{t} \text{ for } \exists \mathbf{z}_j \mathbf{q}_j \text{ and } \mathcal{T} \\ \text{with } z_j \in \mathfrak{t}_i}} (\text{tw}_{\mathfrak{t}} \wedge (\mathbf{q}_j \setminus \mathfrak{q}_{\mathfrak{t}})^{\mathbf{z}_j \setminus (\mathfrak{t}_r \cup \mathfrak{t}_i)}), & \text{if there is } z_j \in \mathbf{z}_j, \\ \mathbf{q}_j, & \text{otherwise.} \end{cases}$$

Note that  $\mathbf{q}^\mathbf{z}$  depends on the choice of  $\mathbf{q}_1, \dots, \mathbf{q}_k$  and  $z_j \in \mathbf{z}_j$ , which can be arbitrary. Intuitively, the first disjunct of  $\mathbf{q}^\mathbf{z}$  reflects the situation where  $z_j$  is mapped to an ABox element; so we treat  $z_j$  as a free variable when rewriting  $\mathbf{q}_j$ . The other disjuncts reflect the case when  $z_j$  is mapped to the non-ABox part of the canonical model, in which case  $z_j$  belongs to the internal part  $\mathfrak{t}_i$  of a tree witness  $\mathfrak{t} = (\mathfrak{t}_r, \mathfrak{t}_i)$  for  $\exists \mathbf{z}_j \mathbf{q}_j$  and  $\mathcal{T}$ . As the variables in  $\mathfrak{t}_r$  must be mapped to ABox elements, this leaves the set  $\mathbf{q}_j \setminus \mathfrak{q}_{\mathfrak{t}}$  of atoms with existentially quantified  $\mathbf{z}_j \setminus (\mathfrak{t}_r \cup \mathfrak{t}_i)$  for further rewriting (this set of variables does not contain  $z_j$ ).

**Theorem 7.** *For any ABox  $\mathcal{A}$  that is H-complete with respect to  $\mathcal{T}$  and any  $\mathbf{a} \subseteq \text{ind}(\mathcal{A})$ , we have  $\mathcal{C}_{\mathcal{T}, \mathcal{A}} \models \mathbf{q}(\mathbf{a})$  iff  $\mathcal{A} \models \mathbf{q}^\dagger(\mathbf{a})$ .*

*Example 4.* Take  $\mathbf{q}$  and  $\mathcal{T}$  from Example 1. The only  $\{y_2, y_3\}$ -component of  $\mathbf{q}$  is  $\mathbf{q}$ . Then we pick, say  $y_2$ , and obtain  $\mathbf{q}^{\{y_2, y_3\}} = \mathbf{q}^{\{y_3\}} \vee (\text{tw}_{\mathfrak{t}^1} \wedge R(y_3, x_4)^\emptyset)$ . Now,  $\mathbf{q}$  has two  $\{y_3\}$ -components,  $\{R(x_1, y_2)\}$  and  $\mathbf{q}_1 = \{R(y_3, y_2), R(y_3, x_4)\}$ . The former gives  $R(x_1, y_2)$ , while in the latter we have to pick  $y_3$  and obtain  $\mathbf{q}_1^\emptyset \vee \text{tw}_{\mathfrak{t}^2}$ , assuming that the empty set of atoms is  $\top$ . This gives the rewriting:

$$\mathbf{q}^\dagger(x_1, x_4) = \exists y_2, y_3 \left[ \left( R(x_1, y_2) \wedge ((R(y_3, y_2) \wedge R(y_3, x_4)) \vee \text{tw}_{\mathfrak{t}^2}) \right) \vee \left( \text{tw}_{\mathfrak{t}^1} \wedge R(y_3, x_4) \right) \right].$$

A CQ  $\mathbf{q}$  is said to be *tree-shaped* if its primal graph (whose vertices are the variables in  $\mathbf{q}$  and edges are pairs  $\{t, t'\}$  such that  $R(t, t') \in \mathbf{q}$ ) is a tree. In each component  $\mathbf{q}_j$  of a tree-shaped CQs  $\mathbf{q}$ , we can choose a variable  $z_j$  that splits it in half. More formally, we have the following:

**Proposition 1.** *For any tree  $T = (V, E)$ , there is a vertex  $v \in V$  such that each connected component obtained by removing  $v$  from  $T$  contains  $\leq |V|/2$  vertices.*

By Proposition 1, any tree-shaped CQ can be split into components, each of which contains less than a half of the existentially quantified variables of the CQ. By applying this argument recursively and using the fact that, if a TBox  $\mathcal{T}$  is of depth 1 then, for any variable  $z$  in  $\mathbf{q}$ , the number of tree witnesses  $\mathfrak{t} = (\mathfrak{t}_r, \mathfrak{t}_i)$  for  $\mathbf{q}$  and  $\mathcal{T}$  with  $z \in \mathfrak{t}_i$  does not exceed 2, we obtain our final result:

**Theorem 8.** *Any tree-shaped CQ over any TBox of depth 1 has a polynomial PE-rewriting.*

## References

1. N. Alon and R. Boppana. The monotone circuit complexity of Boolean functions. *Combinatorica*, 7(1):1–22, 1987.
2. S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.
3. A. Artale, D. Calvanese, R. Kontchakov, and M. Zakharyashev. The DL-Lite family and relations. *J. Artif. Intell. Res. (JAIR)*, 36:1–69, 2009.
4. J. Avigad. Eliminating definitions and Skolem functions in first-order logic. In *Proc. of 16th Annual IEEE Symposium on Logic in Computer Science (LICS)*, pages 139–146, 2001.
5. F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.
6. E. Börger, E. Grädel, and Y. Gurevich. *The Classical Decision Problem*. Perspectives in Mathematical Logic. Springer, 1997.
7. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Automated Reasoning*, 39(3):385–429, 2007.
8. G. Gottlob and T. Schwentick. Rewriting ontological queries into small nonrecursive datalog programs. In *Proc. of the 24th Int. Workshop on Description Logics (DL 2011)*, vol. 745 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2011.
9. M. Grigni and M. Sipser. Monotone separation of logarithmic space from logarithmic depth. *J. Comput. Syst. Sci.*, 50(3):433–437, 1995.
10. S. Jukna. *Boolean Function Complexity — Advances and Frontiers*, volume 27 of *Algorithms and Combinatorics*. Springer, 2012.
11. S. Kikot, R. Kontchakov, and M. Zakharyashev. On (In)Tractability of OBDA with OWL 2 QL. In *Proc. of the 24th Int. Workshop on Description Logics (DL 2011)*, vol. 745 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2011.
12. S. Kikot, R. Kontchakov, and M. Zakharyashev. Conjunctive query answering with OWL 2 QL. In *Proc. of the 13th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR 2012)*. AAAI Press, 2012.
13. S. Kikot, R. Kontchakov, V. Podolskii, and M. Zakharyashev. Exponential lower bounds and separation for query rewriting. In *Proc. of the 39th Int. Colloquium on Automata, Languages, and Programming (ICALP 2012)*, vol. 7392 of LNCS, pages 263–274. Springer, 2012.
14. S. Kikot, R. Kontchakov, V. Podolskii, and M. Zakharyashev. Long rewritings, short rewritings. In *Proc. of the 2012 Int. Workshop on Description Logics (DL 2012)*, vol. 846 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2012.
15. R. Raz and A. Wigderson. Monotone circuits for matching require linear depth. *J. ACM*, 39(3):736–744, 1992.
16. A. Razborov. Lower bounds for the monotone complexity of some Boolean functions. *Dokl. Akad. Nauk SSSR*, 281(4):798–801, 1985.
17. A. Razborov. Lower bounds for deterministic and nondeterministic branching programs. In *Proc. of the 8th Int. Symposium on Fundamentals of Computation Theory (FCT'91)*, vol. 529 of LNCS, pages 47–60. Springer, 1991.