Proceedings of the Sixteenth International Conference on
Principles of Knowledge Representation and Reasoning (KR 2018)

Loop Restricted Existential Rules and First-Order Rewritability for Query Answering

Vernon Asuncion, [a] **Yan Zhang,** [a, c] **Heng Zhang,** [b] **Yun Bai,** [a] **Weisheng Si** [a]
[a] School of Computing Engineering and Mathematics
Western Sydney University, Australia
[b] School of Computer Software, Tianjin University, China
[c] Huazhong University of Science & Technology, China

Introduction

Under the language of TGDs, queries are answered against an ontology represented by a set of TGDs and an input database. In particular, given a database instance D, a finite set Σ of TGDs, and a query q, we want to decide whether $D \cup \Sigma \models q$. However, this problem is undecidable generally, due to the potential cyclic applications of TGDs in Σ

In recent years, considerable research has been carried out to identify various expressive decidable classes of TGDs. Among all these decidable classes, some are of special interests for OBDA, i.e., the classes of first-order rewritable TGDs, where conjunctive query answering can be reduced to the evaluation of a first-order query over the database. So far, several useful first-order rewritable classes of TGDs have been discovered: *acyclic* (AC), *domain restricted* (DR) (Baget et al. 2011), *aGRD* (Baget et al. 2011), *linear* and *multi-linear* (ML), *sticky* and *sticky-join* (SJ), while multi-linear and sticky-join generalise linear TGDs and sticky TGDs, respectively (Calì, Gottlob, and Pieris 2012). Civili and Rosati (Civili and Rosati 2012) further identified another first-order rewritable class called *weakly recursive* TGDs, and showed that by restricting to simple TGDs, weakly recursive class contains all other first-order rewritable classes. Unfortunately, there are still real life scenarios that are simple and intuitive but not syntactically recognisable by any of the existing first-order rewritable TGDs classes, which this work on loop restricted TGDs tries to address.

Then main contributions of this paper are summarised here:

1. We define notations of *derivation paths* and *derivation trees* for query answering over TGDs (existential rules), and provide a precise characterisation for the traditional TGDs chase procedure through the corresponding derivation tree.
2. Based on the concept of derivation paths, we introduce a new class called *loop restricted* (LR) TGDs, which are TGDs with certain restrictions on the loops embedded in the underlying rule set.
3. Under our derivation tree framework, we show that the conjunctive query answering (CQA) under LR TGDs satisfies a property called bounded derivation tree depth property (BDTDP). We further prove that BDTDP implies the well-known bounded derivation-depth property (BDDP). This result implies that conjunctive query answering under LR TGDs is not only decidable but also first-order rewritable.
4. We further extend LR TGDs to *generalised loop restricted* (GLR) TGDs, and prove that the class of GLR TGDs is also first-order rewritable and contains most of other first-order rewritable TGD classes discovered in the literature so far.

Preliminaries

A tuple-generating dependency (TGD) σ, also called *existential rule*, over a schema $\mathcal{R}$ is a first-order formula of the form

$$\sigma : \forall \mathbf{X}\mathbf{Y}\varphi(\mathbf{X}, \mathbf{Y}) \rightarrow \exists \mathbf{Z}\psi(\mathbf{X}, \mathbf{Z}), \tag{1}$$

where $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z} \subset \Gamma \cup \Gamma_V$, φ and ψ are conjunctions of atoms over $\mathcal{R}$. When there is no confusion, we usually omit the universal quantifiers from (1). In this case, we also use $\mathsf{head}(\sigma)$ and $\mathsf{body}(\sigma)$ to denote formulas $\exists \mathbf{Z}\psi(\mathbf{X}, \mathbf{Z})$ and $\varphi(\mathbf{X}, \mathbf{Y})$ respectively. In this case, we also use $\mathsf{head}(\sigma)$ and $\mathsf{body}(\sigma)$ to denote formulas $\exists \mathbf{Z}\psi(\mathbf{X}, \mathbf{Z})$ and $\varphi(\mathbf{X}, \mathbf{Y})$ respectively.

Definition 1 (**Derivation path**). *Let Σ be a set of TGDs. A* derivation path *P of Σ is a finite sequence of pairs of an atom and a rule: (α_1, ρ_1), $\cdots$, (α_n, ρ_n), such that*

- *for each $1 \leq i \leq n$, $\alpha_i = \mathsf{head}(\rho_i)$;*
- *for each $1 \leq i \leq n$, $\rho_i = \sigma_i\theta_i$ for some $\sigma_i \in \Sigma$ and substitution θ_i;*
- *for each $1 \leq i < n$, $\alpha_{i+1} \in \mathsf{body}(\rho_i)$;*
- *for each $1 \leq i \leq n$, if a null $\mathsf{n} \in \mathsf{head}(\alpha_i)$ is introduced due to the elimination of existentially quantified variable, then this n must not occur in ρ_j, for all $j \in \{i+1, \ldots, n\}$.*

A *conjunctive query (CQ)* q of arity n over a schema $\mathcal{R}$ has the form $p(\mathbf{X}) \leftarrow \exists \mathbf{Y}\varphi(\mathbf{X}, \mathbf{Y})$, where $\varphi(\mathbf{X}, \mathbf{Y})$ is a conjunction of atoms with the variables $\mathbf{X}$ and $\mathbf{Y}$ from Γ_V and constants from Γ, but without nulls, and p is an n-ary predicate not occurring in $\mathcal{R}$. We allow $\varphi(\mathbf{X}, \mathbf{Y})$ to contain equalities but no inequalities. When $\varphi(\mathbf{X}, \mathbf{Y})$ is just a single atom, then we say that the CQ q is *atomic*. A *Boolean Conjunctive Query (BCQ)* over $\mathcal{R}$ is a CQ of zero arity. In this case, we can simply write a BCQ q as $\exists\mathbf{Y}\varphi(\mathbf{Y})$. A CQ answering problem, or called CQA problem, defined to be the *answer* to a CQ q with n arity over an instance I, denoted as $q(I)$, is the set of all n-tuples $\mathbf{t} \in \Gamma^n$ for which there exists a homomorphism $h : \mathbf{X} \cup \mathbf{Y} \to \Gamma \cup \Gamma_V$ such that $h(\varphi(\mathbf{X}, \mathbf{Y})) \subseteq I$ and $h(\mathbf{X}) = \mathbf{t}$. The answer to a BCQ is *positive* over I, denoted as $I \models q$, if $\langle\rangle \in q(I)$.

Generalised Loop Restricted TGDs

First we introduce the notion of the bounded derivation depth property [1].

Definition 2 (**BDDP**). *A class $\mathcal{C}$ of TGDs satisfies the* bounded derivation-depth property *(BDDP) if for each BCQ q over a schema $\mathcal{R}$, for every input database D for $\mathcal{R}$ and for every set $\Sigma \in \mathcal{C}$ over $\mathcal{R}$, $D \cup \Sigma \models q$ implies that there exists some $k \geq 0$ which only depends on q and Σ such that* $\mathsf{chase}^k(D, \Sigma) \models q$.

It has been shown that the BDDP implies the first-order rewritability (Calì, Gottlob, and Lukasiewicz 2012; Calì, Gottlob, and Pieris 2012). Formally, the BCQA problem is *first-order rewritable* for a class $\mathcal{C}$ of sets of TGDs if for each $\Sigma \in \mathcal{C}$, and each BCQ q, there exists a first-order query q_Σ such that $D \cup \Sigma \models q$ iff $D \models q_\Sigma$, for every input database D. In this case, we also simply say that the class $\mathcal{C}$ of TGDs is *first-order rewritable*.

Definition 3 (**BDTDP**). *A class $\mathcal{C}$ of TGDs satisfies the* bounded derivation tree depth property *(BDTDP) if for each $\Sigma \in \mathcal{C}$, there exists some $k \geq 0$ such that for every BCQ query $\exists\mathbf{Z}p(\mathbf{Z})$ and every database D, $D \cup \Sigma \models \exists\mathbf{Z}p(\mathbf{Z})$ iff $T(D, \Sigma) \models p(\mathbf{n})$ for some instantiated derivation tree $T(D, \Sigma)$ and atom $p(\mathbf{n})$, where $\mathsf{depth}(T(D, \Sigma)) \leq k$ and $h(\mathbf{Z}) = \mathbf{n}$ for some homomorphism h.*

Basically, Definition 3 says that if a class of TGDs satisfies BDTDP, then its every BCQ query answering problem can be always decided within a fixed number k of derivation steps with respect to the corresponding instantiated derivation trees.

Now we are ready to formally define the notion of generalised loop restricted patterns. Let A be a set of atoms, we use $\mathsf{var}(A)$ (resp., $\mathsf{nulls}(A)$) to denote the set of all variables (resp., nulls) occurring in A.

Definition 4 (**Generalised loop restricted (GLR) patterns**). *Let Σ be a set of TGDs. Σ is* generalised loop restricted *(GLR), if each loop pattern $L = (\alpha_1, \rho_1) \cdots (\alpha_n, \rho_n)$ of Σ falls into one of the following four types:*

Type I *For each pair (α_i, ρ_i) in L ($1 \leq i < n$), $\mathsf{body}(\rho_i)$ can be separated into two disjoint parts $\mathsf{body}(\rho_i) = \mathsf{body_h}(\rho_i) \cup \mathsf{body_b}(\rho_i)$ such that the following three conditions holds:*

1. $\mathsf{body_h}(\rho_i) \cap \mathsf{body_b}(\rho_i) = \emptyset$,
2. $\alpha_{i+1} \in \mathsf{body_b}(\rho_i)$,
3. $\mathsf{var}\big(\{\alpha_i\} \cup \mathsf{body_h}(\rho_i)\big) \cap \mathsf{var}\big(\mathsf{body_b}(\rho_i)\big) = \bigcap_{j=1}^{n} \mathsf{var}(\alpha_j)$;

Type II *There exists a pair (α_i, ρ_i) in L ($1 \leq i < n$) such that $\mathsf{body}(\rho_i)$ can be separated into two disjoint parts $\mathsf{body}(\rho_i) = \mathsf{body_h}(\rho_i) \cup \mathsf{body_b}(\rho_i)$, where the following three conditions hold:*

1. $\mathsf{body_h}(\rho_i) \cap \mathsf{body_b}(\rho_i) = \emptyset$,
2. $\alpha_{i+1} \in \mathsf{body_b}(\rho_i)$,
3. $\mathsf{var}\big(\{\alpha_i\} \cup \mathsf{body_h}(\rho_i)\big) \cap \mathsf{var}\big(\mathsf{body_b}(\rho_i)\big) = \emptyset$;

Type III *For each pair (α_i, ρ_i) in L ($1 \leq i < n$) and each $\beta \in \mathsf{body}(\rho_i)$, $\mathsf{var}(\rho_i) \subseteq \mathsf{var}(\beta)$;*

Type IV *For each pair (α_i, ρ_i) in L ($1 \leq i < n$) and each $\beta \in \mathsf{body}(\rho_i) \setminus \{\alpha_{i+1}\}$, $\big(\mathsf{var}(\alpha_{i+1}) \cap \mathsf{var}(\beta)\big) \neq \emptyset$ implies $\big(\mathsf{var}(\alpha_{i+1}) \cap \mathsf{var}(\beta)\big) \subseteq \bigcap_{j=1}^{i} \mathsf{var}(\alpha_j)$;*

Type V *There exists a pair (α_i, ρ_i) in L ($1 \leq i < n$), such that $\mathsf{body}(\rho_i)$ can be separated into two disjoint parts $\mathsf{body}(\rho_i) = \mathsf{body_h}(\rho_i) \cup \mathsf{body_b}(\rho_i)$, where the following three conditions hold:*

1. $\mathsf{body_h}(\rho_i) \cap \mathsf{body_b}(\rho_i) = \emptyset$,
2. $(\bigcup_{j=i+1}^{n}(\alpha_j)) \cap \mathsf{body_h}(\rho_i) = \emptyset$,
3. $\mathsf{null}(\mathsf{body_h}(\rho_i)) \neq \emptyset$.

[1]For space reasons, the full notions of the BDDP, BDTDP and loop patterns can be found in our archived conference version of the paper: https://arxiv.org/abs/1804.07099.

Main Results

Theorem 1. *If a class $\mathcal{C}$ of TGDs satisfies BDTDP then $\mathcal{C}$ also satisfies BDDP. Therefore, since we can also prove that the class GLR of TGDs satisfies BDTDP, then it follows that the class GLR is also first-order rewritable.*

Theorem 2. *Consider the BCQA problem for a given set of GLR TDGs. Its data complexity is in* AC^0, *and its combined complexity is* EXPTIME *complete.*

Theorem 3. *Deciding whether a set of TGDs is generalised loop restricted is* PSPACE *complete.*

GLR actually captures a large class of first-order rewritable TGDs. In fact, we have the following result.

Proposition 1. *Let GLR be the class of generalised loop restricted TGDs defined in Definition 4. Then we have that: (1)* AC $\subsetneq$ GLR*; (2)* ML $\subsetneq$ GLR*; (3)* SJ $\subsetneq$ GLR*; (4)* aGRD $\subsetneq$ GLR*; (5)* DR $\subsetneq$ GLR.

References

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

Calì, A.; Gottlob, G.; and Lukasiewicz, T. 2012. A general datalog-based framework for tractable query answering over ontologies. *J. Web Sem.* 14:57–83.

Calì, A.; Gottlob, G.; and Pieris, A. 2012. Towards more expressive ontology languages: The query answering problem. *Artif. Intell.* 193:87–128.

Civili, C., and Rosati, R. 2012. A broad class of first-order rewritable tuple-generating dependencies. In *Proceedings of the 2nd International Conference on Datalog in Academia and Industry (Datalog-2012)*, 68–80.