Model-theoretic Characterizations of Existential Rule Languages

Heng Zhang,[1] **Yan Zhang,**[2,4] **Guifei Jiang** [3]
[1]College of Intelligence and Computing, Tianjin University, Tianjin, China
[2]School of Computer, Data and Mathematical Sciences, Sydney, Australia
[3]College of Software, Nankai University, Tianjin, China
[4]School of Computer Sci. & Tech., Huazhong University of Sci. & Tech., Wuhan, China
heng.zhang@tju.edu.cn, yan.zhang@westernsydney.edu.au, g.jiang@nankai.edu.cn

Abstract

Existential rules, a.k.a. dependencies in databases, and Datalog+/- in knowledge representation and reasoning recently, are a family of important logical languages widely used in computer science and artificial intelligence. Towards a deep understanding of these languages in model theory, we establish model-theoretic characterizations for a number of existential rule languages such as (disjunctive) embedded dependencies, tuple-generating dependencies (TGDs), (frontier-)guarded TGDs and linear TGDs. All these characterizations hold for the class of arbitrary structures, and most of them also work on the class of finite structures. As a natural application of these results, complexity bounds for the rewritability of above languages are also identified.

1 Introduction

Existential rule languages, a family of languages that extend Datalog by allowing existential quantifiers in the rule head, had been initially introduced in databases in 1970s to specify the semantics of data stored in a database [Abiteboul *et al.*, 1995]. Since then, existential rule languages such as tuple-generating dependencies (TGDs), embedded dependencies and equality-generating dependencies have been extensively studied. These languages have been recently rediscovered as languages for data exchange [Fagin *et al.*, 2005], data integration [Lenzerini, 2002] and ontology-mediated query answering [Calì *et al.*, 2010]. Towards tractable reasoning, many restricted classes of these languages have been proposed, including linear and guarded TGDs [Calì *et al.*, 2012], as well as frontier-guarded TGDs [Baget *et al.*, 2011]. As a family of important logical languages, their model theory has not been fully investigated yet. In this work we aim at characterizing existential rule languages in a model-theoretic approach.

Model-theoretic characterizations, which assert that a sentence in a language is definable in another language if, and only if, it enjoys some semantic properties, play a key role in the study of logic [Chang and Keisler, 1990]. We are interested in semantic properties that are simple and manageable. Model-theoretic characterizations based on such properties thus provide a natural tool for identifying the expressibility of a language, i.e., determining which knowledge or ontology can be expressed in the language.

Besides the major position in model theory and the key role on understanding expressiveness, model-theoretic characterizations also have many potential implications. For example, model-theoretic characterizations provide a natural way for developing algorithms to identify language rewritability, i.e., to decide whether a given theory or ontology can be rewritten in a simpler language. Such algorithms may play important roles in implementing systems for ontology-mediated query answering. Moreover, we are also interested in understanding why the guarded-based restrictions make existential rule languages tractable. We hope our characterizations give an alternative explanation on this question, which may provide a new insight to exploit new tractable languages.

Model-theoretic characterizations over the class of finite structures for full TGDs (i.e., TGDs without existential quantifiers) and equality-generating dependencies had been studied in [Makowsky and Vardi, 1986], which are established by involving infinite sets of dependencies. To remedy the finite expressibility, some conditions had been proposed, including Hull's *finite-rank* notion [1984] and Makowsky and Vardi's *locality* [1986]. Yet both of them are not very natural. Over finite structures, even for full TGDs, a natural model-theoretic characterization remains open [ten Cate and Kolaitis, 2014]. For arbitrary structures, except for some simple classes of dependencies such as full TGDs and negative constraints, to the best of our knowledge, no model-theoretic characterization is known for expressive existential rule languages such as TGDs and its guarded-based restrictions.

In this work, we characterize existential rule languages by some natural semantic properties. The addressed languages consist of (disjunctive) embedded dependencies, TGDs, and several important restricted classes of TGDs such as frontier-guarded TGDs, guarded TGDs and linear TGDs, three of the main languages for ontology-mediated query answering [Calì *et al.*, 2010]. All the semantic properties involved in our characterizations are algebraic relationships among structures, incuding variants of homomorphisms and unions, as well as direct products. Interestingly, except the characterizations w.r.t. first-order logic, all other characterizations hold for both finite structures and arbitrary structures. As a natural application, we also use the obtained characterizations to identify the complexity of rewritability among the above languages.

For proof details please refer to a long version of this paper, which is available at https://arxiv.org/abs/2001.08688.

2 Preliminaries

2.1 Notations and Conventions

All *signatures* involved in this paper are *relational*, consisting of a set of *constant symbols* and a set of *relation symbols*, each of which is armed with a natural number, its *arity*. Each *term* is either a variable or a constant symbol. Given a signature τ, *atomic formulas*, *(first-order) formulas* and *sentences* over τ are defined as usual. An atomic formula is *relational* if it is of the form $R(\vec{t})$ where R is a relation symbol other than the equality symbol $=$. Given a formula φ, we write $\varphi(\vec{x})$ if every *free variable* of φ appears in $\vec{x}$.

Fix τ as a signature. Every *structure* $\mathcal{A}$ *over* τ (or simply τ*-structure*) consists of a nonempty set A called its *domain*, a relation $R^{\mathcal{A}} \subseteq A^n$ for each n-ary relation symbol $R \in \tau$, and a constant $c^{\mathcal{A}} \in A$ for each constant symbol $c \in \tau$. A structure is *finite* if its domain is finite, and *infinite* otherwise.

Let $\mathcal{A}$ be a τ-structure, and X a subset of A such that $c^{\mathcal{A}} \in X$ for all constant symbols $c \in \tau$. The *substructure* of $\mathcal{A}$ *induced* by a set $X \subseteq A$, denoted $\mathcal{A}|_X$, is a τ-structure with domain X which interprets each relation symbol $R \in \tau$ as $R^{\mathcal{A}}|_X$, and interprets each constant symbol $c \in \tau$ as $c^{\mathcal{A}}$. A structure $\mathcal{B}$ is called a *substructure* of $\mathcal{A}$, or equivalently, $\mathcal{A}$ is called an *extension* of $\mathcal{B}$, if $\mathcal{B} = \mathcal{A}|_X$ for some set $X \subseteq A$. Let ν be a signature such that $\tau \subseteq \nu$. A ν-structure $\mathcal{B}$ is called a ν*-expansion* of $\mathcal{A}$ if they have the same domain and share the same interpretation on every symbol in τ. Suppose $a_1, \dots, a_k \in A$, by $(\mathcal{A}, a_1, \dots, a_k)$ we denote the expansion of $\mathcal{A}$ that assigns each constant a_i to a fresh constant symbol.

Let $\mathcal{A}$ and $\mathcal{B}$ be τ-structures. If $\mathcal{A}$ and $\mathcal{B}$ have the same interpretations on constant symbols then let $\mathcal{A} \cup \mathcal{B}$ denote the *union* of $\mathcal{A}$ and $\mathcal{B}$, which is a τ-structure with domain $A \cup B$, interpreting R as $R^{\mathcal{A}} \cup R^{\mathcal{B}}$ for each relation symbol $R \in \tau$, and interpreting c as $c^{\mathcal{A}}$ for each constant symbol $c \in \tau$. We say $\mathcal{A}$ is *homomorphic to* $\mathcal{B}$, written $\mathcal{A} \to \mathcal{B}$, if there is a function $h : A \to B$ such that (i) $h(c^{\mathcal{A}}) = c^{\mathcal{B}}$ for all constant symbols $c \in \tau$, and (ii) $h(R^{\mathcal{A}}) \subseteq R^{\mathcal{B}}$ for all relation symbols $R \in \tau$. We write $\mathcal{A} \rightleftarrows \mathcal{B}$ if both $\mathcal{A} \to \mathcal{B}$ and $\mathcal{B} \to \mathcal{A}$ hold.

Let $\mathcal{A}$ be a structure. An *assignment* in $\mathcal{A}$ is a function from a set of variables to A. Given a tuple $\vec{a}$ of constants in A and a tuple $\vec{x}$ of variables of the same length, we let $\vec{a}/\vec{x}$ denote the assignment that maps the i-component of $\vec{x}$ to the i-component of $\vec{a}$ for $1 \leq i \leq |\vec{x}|$, where $|\vec{x}|$ denotes the length of $\vec{x}$. Let s be an assignment in $\mathcal{A}$ and $\varphi(\vec{x})$ be a first-order formula. By $\mathcal{A} \models \varphi[s]$ we mean that φ is *satisfied* by s in $\mathcal{A}$. In particular, if φ is a sentence, we simply write $\mathcal{A} \models \varphi$, and say φ is *satisfied* in $\mathcal{A}$, or equivalently, $\mathcal{A}$ is a *model* of φ. If the assignment $\vec{a}/\vec{x}$ is clear from the context, we simply use $\varphi[\vec{a}]$ to denote $\varphi[\vec{a}/\vec{x}]$. Let Σ be a set of sentences, $\mathcal{A}$ is a *model* of Σ if $\mathcal{A} \models \varphi$ for all $\varphi \in \Sigma$. Given a sentence ψ, we write $\Sigma \vDash \psi$ (resp., $\Sigma \vDash_{\text{fin}} \psi$) if every model (resp., finite model) of Σ is also a model of ψ.

2.2 Existential Rule Languages

A *generalized dependency* (GD) is a sentence σ of the form

$$\forall \vec{x}(\phi(\vec{x}) \to \exists \vec{y}(\psi_1(\vec{x}, \vec{y}) \vee \cdots \vee \psi_n(\vec{x}, \vec{y})) \qquad (1)$$

where $n \geq 0$, and $\phi, \psi_1, \dots, \psi_n$ are conjunctions of atomic formulas. The left-hand (resp., right-hand) side of the implication is called the *body* (resp., *head*). Variables among $\vec{x}$ and $\vec{y}$ are called *universal*, and *existential*, respectively. A *frontier variable* is a universal variable that occurs in the head. In particular, σ is called *nondisjunctive* if $n \leq 1$, and called a *negative constraint* if $n = 0$. In the latter case, we write σ as

$$\forall \vec{x}(\phi(\vec{x}) \to \bot). \qquad (2)$$

For simplicity, we will omit the universal quantifiers and the brackets appearing outside the atoms if no confusion occurs.

Furthermore, a GD σ is called *safe* if every frontier variable of σ has at least one occurrence in some relational atomic formula in the body of σ. Every *disjunctive embedded dependency* (DED) is a safe generalized dependency which is not a negative constraint. Every *embedded dependency* (ED) is a nondisjunctive DED. In addition, an ED is called an *tuple-generating dependency* (TGD) if it is equality-free.

We will also address several important classes of restricted TGDs. A TGD σ is called *frontier-guarded* (resp., *guarded*) if there is a relational atomic formula α in its body that contains all the frontier (resp., universal) variables of σ. In either case, α is called the *guard* of σ. Moreover, σ is *linear* if the body of σ consists of exactly one conjunct. Note that all linear TGDs are guarded and all guarded TGDs are frontier-guarded.

3 Model-theoretic Characterizations

In this section, we address the model-theoretic characterizations of existential rule languages mentioned above.

3.1 Generalized Dependencies

We first give some notions. Let $\mathcal{A}$ and $\mathcal{B}$ be structures over a signature τ. By a *tuple* on $\mathcal{A}$ we mean a finite sequence of constants in A. We say that $\mathcal{A}$ is globally-homomorphic to $\mathcal{B}$, written $\mathcal{A} \Rightarrow \mathcal{B}$, if there is a function π that maps each tuple $\vec{a}$ on $\mathcal{A}$ to a tuple $\pi(\vec{a})$ on $\mathcal{B}$ such that $(\mathcal{A}, \vec{a}) \rightleftarrows (\mathcal{B}, \pi(\vec{a}))$; in this case, we call π a *global homomorphism* from $\mathcal{A}$ to $\mathcal{B}$, and call $\mathcal{A}$ a *globally-homomorphic preimage* of $\mathcal{B}$.

Given a first-order sentence φ over τ, we say that φ is *preserved under globally-homomorphic preimages [in the finite]* if for all [finite] τ-structures $\mathcal{A}$ and $\mathcal{B}$, if $\mathcal{A}$ is globally homomorphic to $\mathcal{B}$ and $\mathcal{B}$ is a model φ, then $\mathcal{A}$ is also a model of φ. Notice that every sentence preserved under globally-homomorphic preimages is also preserved under globally-homomorphic preimages in the finite, but not vice versa.

By a routine check, it is easy to prove the following:

Proposition 1. *Every set of GDs is preserved under globally-homomorphic preimages [in the finite].*

To establish the desired characterization, we hope that the preservation under globally-homomorphic preimages is not too powerful. The following is a very simple example which is slightly beyond the class of GDs but already not preserved under globally-homomorphic preimages in the finite.

Example 2. *Let* ψ *denote* $\exists x \neg Q(x)$ *and* $\tau = \{Q\}$*. Let* $\mathcal{A}$ *be a* τ*-structure with* $A = \{a, b\}$ *and* $Q^{\mathcal{A}} = \{a\}$*. Let* $\mathcal{B}$ *be the substructure of* $\mathcal{A}$ *induced by* $\{a\}$*. Clearly,* $\mathcal{B}$ *is globally homomorphic to* $\mathcal{A}$*. It is also easy to see that* $\mathcal{A}$ *is a model of*

ψ, but $\mathcal{B}$ is not, which implies that ψ is not preserved under globally-homomorphic preimages even in the finite.

The following theorem establishes the desired characterizations for the class of finite sets of GDs.

Theorem 3. *A first-order sentence is equivalent to a finite set of GDs iff it is preserved under globally-homomorphic preimages.*

To prove this theorem, we need some notions and lemmas. Let $\mathcal{A}$ and $\mathcal{B}$ be structures over a signature τ. Given a class $\mathcal{C}$ of sentences over τ, we write $\mathcal{A} \preceq_{\mathcal{C}} \mathcal{B}$ if for all sentences $\varphi \in \mathcal{C}$, $\mathcal{A} \models \varphi$ implies $\mathcal{B} \models \varphi$. For simplicity, we simply drop the subscript $\mathcal{C}$ if $\mathcal{C}$ is the class of all first-order sentences over τ. We write $\mathcal{A} \equiv \mathcal{B}$ if both $\mathcal{A} \preceq \mathcal{B}$ and $\mathcal{B} \preceq \mathcal{A}$ hold.

We write $\Gamma(x)$ to denote a set of formulas with exactly one free variable x. We say that $\Gamma(x)$ is *realized* in a structure $\mathcal{A}$ if there is some $a \in A$ such that $\mathcal{A} \models \vartheta[a/x]$ for all formulas $\vartheta(x) \in \Gamma(x)$. By $Th(\mathcal{A})$ we denote the class of all first-order sentences satisfied in $\mathcal{A}$. We say that $\mathcal{A}$ is *ω-saturated* if for every finite set $X \subseteq A$, every set $\Gamma(x)$ of formulas consistent with $Th((\mathcal{A}, a)_{a\in X})$ is realized in $(\mathcal{A}, a)_{a\in X}$. It is well-known [Chang and Keisler, 1990] that for every structure $\mathcal{A}$ there is an ω-saturated structure $\mathcal{B}$ such that $\mathcal{A} \equiv \mathcal{B}$.

Every *existential-positive* formula is a first-order formula built on atomic formulas and negated atomic formulas by using connectives $\wedge, \vee$ and the quantifier $\exists$. Let $\exists^+$ denote the class of existential-positive sentence. It is easy to prove:

Lemma 4. *Let $\mathcal{A}$ and $\mathcal{B}$ be structures over the same signature. Then both of the following are true:*

1. *If $\mathcal{A} \rightarrow \mathcal{B}$ then $\mathcal{A} \preceq_{\exists^+} \mathcal{B}$.*
2. *If $\mathcal{A} \preceq_{\exists^+} \mathcal{B}$ and $\mathcal{B}$ is ω-saturated then $\mathcal{A} \rightarrow \mathcal{B}$.*

Let GD denote the class of finite sets fo generalized dependencies. With Lemma 4, we are able to prove the following:

Lemma 5. *Let $\mathcal{A}$ and $\mathcal{B}$ be ω-saturated structures over the same signature. If $\mathcal{B} \preceq_{\mathsf{GD}} \mathcal{A}$ then $\mathcal{A} \Rightarrow \mathcal{B}$.*

Proof. Assume $\mathcal{B} \preceq_{\mathsf{GD}} \mathcal{A}$. We need to prove $\mathcal{A} \Rightarrow \mathcal{B}$. By Lemma 4, it suffices to show that for each tuple $\vec{a}$ on $\mathcal{A}$ there is a tuple $\pi(\vec{a})$ such that $(\mathcal{B}, \pi(\vec{a})) \preceq_{\mathsf{GD}} (\mathcal{A}, \vec{a})$. Note that, by Proposition 5.1.1 in [Chang and Keisler, 1990], $(\mathcal{B}, \pi(\vec{a}))$ and $(\mathcal{A}, \vec{a})$ are ω-saturated; so Lemma 4 is applicable.

The desired statement can be done by an induction on the length of $\vec{a}$. It is trivial for the case where $|\vec{a}| = 0$. Assume as induction hypothesis that the desired statement holds for $|\vec{a}| = k \geq 0$, we need to prove that it also holds for the case where $|\vec{a}| = k + 1$. Suppose $\vec{a} = (\vec{a}_0, a)$. By inductive hypothesis, there is a tuple $\vec{b}_0$ such that

$$(\mathcal{B}, \vec{b}_0) \preceq_{\mathsf{GD}} (\mathcal{A}, \vec{a}_0). \tag{3}$$

Let $\Gamma(x)$ be the class of existential-positive formulas and their negations such that $(\mathcal{A}, \vec{a}_0) \models \varphi[a/x]$ for all $\varphi(x) \in \Gamma(x)$. To prove the existence of a constant $b \in B$ such that

$$(\mathcal{B}, \vec{b}_0, b) \preceq_{\mathsf{GD}} (\mathcal{A}, \vec{a}_0, a), \tag{4}$$

by the ω-saturatedness of $\mathcal{B}$, it suffices to show that every finite subset of $\Gamma(x)$ is realized in $(\mathcal{B}, \vec{b}_0)$. Let $\Gamma_0(x)$ be any finite subset of $\Gamma(x)$. Let $\varphi(x)$ denote the conjunction of all formulas in $\Gamma_0(x)$, and let $\psi = \neg\exists x\varphi(x)$. Clearly, ψ is equivalent to a finite set of GDs and $(\mathcal{A}, \vec{a}_0) \not\models \psi$. By the inductive assumption (3), we know $(\mathcal{B}, \vec{b}_0) \not\models \psi$, or equivalently, there exists a constant $b' \in B$ such that $(\mathcal{B}, \vec{b}_0) \models \varphi[b'/x]$. Consequently, $\Gamma_0(x)$ is realized in $(\mathcal{B}, \vec{b}_0)$, which is as desired. □

Now we are able to prove the desired theorem.

Proof of Theorem 3. (Only-if) By Proposition 1.

(If) We assume that φ is a first-order sentence preserved under globally-homomorphic preimages. Let $\mathrm{con}(\varphi)$ denote the class of all GDs that are logical consequences of φ. We want to show that $\mathrm{con}(\varphi)$ is equivalent to φ, which implies the desire result by compactness. Let $\mathcal{A}$ be any model of $\mathrm{con}(\varphi)$. It suffices to show that $\mathcal{A}$ is also a model of φ. Let

$$\Sigma = \{\neg\gamma : \gamma \in \mathsf{GD} \,\&\, \mathcal{A} \models \neg\gamma\}.$$

Now we prove the following property:

Claim. $\Sigma \cup \{\varphi\}$ is satisfiable.

Let Σ_0 be an arbitrary finite subset of Σ. To show the claim, by compactness, it suffices to show that $\Sigma_0 \cup \{\varphi\}$ is satisfiable. Towards a contradiction, assume that this is not the case. Suppose $\Sigma_0 = \{\neg\gamma_1, \ldots, \neg\gamma_n\}$, and let ψ denote the formula $\gamma_1 \vee \cdots \vee \gamma_n$. Then we must have $\varphi \models \psi$. It is not difficult to see that ψ is equivalent to a GD (by renaming the individual variables and lifting the universal quantifiers, and then by a routine transformation). Thus, $\mathcal{A}$ should be a model of ψ. This implies that there is some integer $i : 1 \leq i \leq n$ such that $\mathcal{A} \models \gamma_i$, which contradicts with $\gamma_i \in \Sigma$ and the definition of Σ. So, we obtain the claim.

Applying the above claim, there is thus a model, say $\mathcal{B}$, of $\Sigma \cup \{\varphi\}$. Consequently, we have $\mathcal{B} \preceq_{\mathsf{GD}} \mathcal{A}$. Let $\mathcal{A}^+$ and $\mathcal{B}^+$ be ω-saturated structures such that $\mathcal{A} \equiv \mathcal{A}^+$ and $\mathcal{B} \equiv \mathcal{B}^+$. Then $\mathcal{B}^+ \preceq_{\mathsf{GD}} \mathcal{A}^+$ is clearly true, and $\mathcal{B}^+$ is a model of φ. By Lemma 5, $\mathcal{A}^+$ is then globally homomorphic to $\mathcal{B}^+$. Since by assumption φ is preserved under globally-homomorphic preimages, $\mathcal{A}^+$ should be a model of φ. So, $\mathcal{A}$ is also a model of φ. This thus completes the proof of Theorem 3. □

Note that the above argument only works on the class of arbitrary structures. Over finite structures, the characterization is in general not true, as shown by the following proposition.

Theorem 6. *There is a first-order sentence that is preserved under globally-homomorphic preimages in the finite but is not equivalent to any finite set of GDs over finite structures.*

This can prove by constructing an example, which can be done by a slight modification to Gurevich and Shelah's counterexample (see, e.g., Theorem 2.1.1 in [Rosen, 2002]).

3.2 Disjunctive Embedded Dependencies

According to the definition, DEDs are safe GDs that are not negative constraints. So, to characterize DEDs, we need some properties to assure the safeness and to avoid occurrences of negative constraints. To do the latter, we use a technique called *trivial structure*, which was used in [Makowsky and Vardi, 1986] to characterize full TGDs.

We first recall the notion of trivial structure. A structure $\mathcal{A}$ is called *trivial* if the domain of $\mathcal{A}$ consists of exactly one element and every relation symbol in the signature is interpreted by $\mathcal{A}$ as the full relation on the domain of a proper arity.

To capture the safeness of a DED, we propose a similar notion. A structure $\mathcal{A}$ is called *sharp* if all the following hold:

- the domain of $\mathcal{A}$ consists of exactly two distinct constants, say $*$ and $\circ$;
- for each constant symbol c in the signature, $c^{\mathcal{A}} = *$;
- for each relation symbol R in the signature, $R^{\mathcal{A}}$ consists of exactly a single tuple $(*, \dots, *)$ of a proper length.

The following example shows that the sharp models are able to separate the class of DEDs from the class of GDs:

Example 7. *Let σ be a DED of the following form:*

$$P(x) \wedge R(x, y) \to Q(y). \tag{5}$$

Let $\tau = \{P, Q, R\}$, and let $\mathcal{A}$ be a τ-structure with the domain $\{a, b\}$, interpreting both P and Q as $\{a\}$, and interpreting R as $\{(a, a)\}$. Clearly, $\mathcal{A}$ is a sharp model of σ.

Let σ_0 denote the GD obtained from σ by replacing $R(x, y)$ with $R(x, x)$. Clearly, σ_0 is a GD that is not satisfied in $\mathcal{A}$.

The following result can be shown by a routine check:

Proposition 8. *Let Σ be a finite set of GDs. Then all the following properties are equivalent:*

1. *Σ is equivalent to a finite set of DEDs;*
2. *Σ is equivalent to a finite set of DEDs over finite structures;*
3. *Σ has both a trivial model and a sharp model.*

Note that both "φ has a trivial model" and "φ has a sharp model" can be regarded as trivial preservation properties.

3.3 Embedded Dependencies

To characterize EDs, we use the notion of direct products. Let $\mathcal{A}$ and $\mathcal{B}$ be structures over a signature τ. The *direct product* of $\mathcal{A}$ and $\mathcal{B}$, denoted $\mathcal{A} \times \mathcal{B}$, is a τ-structure defined as follows:

- the domain of $\mathcal{A} \times \mathcal{B}$ is $A \times B$;
- for all constant symbols $c \in \tau$, $c^{\mathcal{A} \times \mathcal{B}} = \langle c^{\mathcal{A}}, c^{\mathcal{B}} \rangle$;
- for all k-ary relation symbols $R \in \tau$, all tuples $\vec{a}$ on $\mathcal{A}$, and all tuples $\vec{b}$ on $\mathcal{B}$, $(\langle a_1, b_1 \rangle, \dots, \langle a_k, b_k \rangle) \in R^{\mathcal{A} \times \mathcal{B}}$ if $\vec{a} \in R^{\mathcal{A}}$ and $\vec{b} \in R^{\mathcal{B}}$, where a_i and b_i denote the i-th component of $\vec{a}$ and $\vec{b}$, respectively.

We say a sentence φ is *preserved under direct products [in the finite]* if, for any two [finite] models $\mathcal{A}$ and $\mathcal{B}$ of φ, $\mathcal{A} \times \mathcal{B}$ is also a model of φ.

The following can be shown by a routine check.

Proposition 9. *Every set of EDs is preserved under direct products [in the finite].*

In general, the direct product preservation fails for DEDs. A simple counterexample is given as follows:

Example 10. *Let σ denote the DED $R \to S \vee T$ where R, S and T are nullary relation symbols. Let τ be the signature $\{R, S, T\}$. Let $\mathcal{A}$ and $\mathcal{B}$ be τ-structures such that*

- *$\mathcal{A}$ and $\mathcal{B}$ have the same domain $\{a\}$;*
- *$R^{\mathcal{A}} = R^{\mathcal{B}} = S^{\mathcal{A}} = T^{\mathcal{B}} = true$, $S^{\mathcal{B}} = T^{\mathcal{A}} = false$.*

Clearly, both $\mathcal{A}$ and $\mathcal{B}$ are models of σ, but $\mathcal{A} \times \mathcal{B}$ is not. Thus, σ is not preserved under direct products even in the finite.

The following result shows that the property of direct product preservation exactly captures the class of DEDs in which the disjunctions can be eliminated. This works over the class of finite structures as well as the class of arbitrary structures.

Theorem 11. *A finite set of DEDs is equivalent to a finite set of EDs [over finite structures] iff it is preserved under direct products [in the finite].*

3.4 Tuple-generating Dependencies

Let $\mathcal{A}$ and $\mathcal{B}$ be structures over a signature τ. A *strict homomorphism from $\mathcal{A}$ into (resp., onto) $\mathcal{B}$* is a function h from A into (resp., onto) B such that

- for every relation symbol $R \in \tau$ and every tuple $\vec{a}$ on $\mathcal{A}$ of a proper length, we have $\vec{a} \in R^{\mathcal{A}}$ iff $h(\vec{a}) \in R^{\mathcal{B}}$, and
- for every constant symbol $c \in \tau$, we have $h(c^{\mathcal{A}}) = c^{\mathcal{B}}$.

If such a strict homomorphism exists, we say $\mathcal{B}$ is a *strictly-homomorphic image* of $\mathcal{A}$, and say $\mathcal{A}$ is, conversely, a *strictly-homomorphic preimage* of $\mathcal{B}$. A sentence φ is said to be *preserved under strictly-homomorphic (pre)images [in the finite]* if, for every [finite] model $\mathcal{A}$ of φ and every [finite] strictly-homomorphic (pre)image $\mathcal{B}$ of $\mathcal{A}$, $\mathcal{B}$ is also a model of φ.

The following gives us the desired characterazations:

Theorem 12. *A finite set of EDs is equivalent to a finite set of TGDs [over finite structures] iff it is preserved under both strictly-homomorphic images and preimages [in the finite].*

Interestingly, it is not difficult to show that, if no constant symbol is involved, the strictly-homomorphic image preservation can be removed from the characterization.

3.5 Frontier-guarded TGDs

To characterize frontier-guarded TGDs, we first define some notations. Let $\mathcal{A}$ be a structure. We define $\{\mathcal{A}_X : X \subseteq A\}$ as a family of structures over the same signature such that

- for all $X \subseteq A$, there is an isomorphism p_X from $\mathcal{A}$ to $\mathcal{A}_X$ such that $p_X(a) = a$ for all $a \in X$;
- for all $X, Y \subseteq A$, $A_X \cap A_Y = X \cap Y$, where A_X and A_Y denote the domains of $\mathcal{A}_X$ and $\mathcal{A}_Y$, respectively.

Every *guarded set* of $\mathcal{A}$ is defined as a finite subset X of A that contains all interpretations of constant symbols in $\mathcal{A}$. A sentence φ is said to be *preserved under isomorphic unions [in the finite]* if, for all [finite] models $\mathcal{A}$ of φ and all finite sets $\mathbb{G}$ of guarded sets of A, $\bigcup_{X \in \mathbb{G}} \mathcal{A}_X$ is also a model of φ.

Example 13. *Let τ denote $\{R\}$ where R is a binary relation symbol. Let $\mathcal{A}$ be a τ-structure defined as follows:*

- *the domain A consists of two distinct constants a and b;*
- *the relation symbol R is interpreted as $A \times A$.*

Let $X = \{a\}$, $Y = \{a, b\}$, and $\mathbb{G} = \{X, Y\}$. Then $\mathcal{A}_X, \mathcal{A}_Y$ and $\bigcup_{Z \in \mathbb{G}} \mathcal{A}_Z$ are τ-structures illustrated by Figure 1.

By a routine check, one can prove the following property:

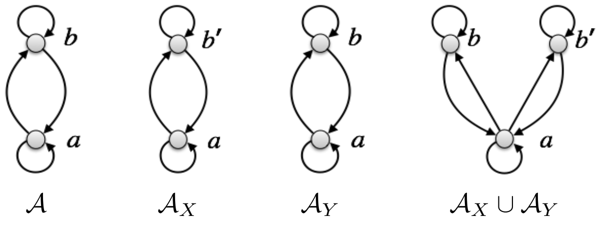

Figure 1: Isomorphic Union in Example 13

Proposition 14. *Every set of frontier-guarded TGDs is preserved under isomorphic unions [in the finite].*

Now, a natural question arises as to whether the isomorphic union preservation is able to separate frontier-guarded TGDs from TGDs. The following example shows that it is true.

Example 15 (Example 13 cont.)**.** *Let σ denote the TGD*

$$R(x,y) \wedge R(y,z) \rightarrow R(x,z) \tag{6}$$

and let $\mathcal{A}$ be the structure defined in Example 13. Then it is easy to see that $\mathcal{A}$ is a model of σ but $\bigcup_{Z\in\mathbb{G}} \mathcal{A}_Z$ is not. So, σ is not preserved under isomorphic unions even in the finite.

The following result provides the desired characterization. Note that the characterization also holds over finite structures.

Theorem 16. *A finite set of TGDs is equivalent to a finite set of frontier-guarded TGDs [over finite structures] iff it is preserved under isomorphic unions [in the finite].*

Every *conjunctive query* (CQ) is a first-order formula of the form $\exists\vec{y}\vartheta(\vec{x},\vec{y})$ where ϑ is a conjunction of relational atomic formulas. Now we first present a lemma as follows:

Lemma 17. *Let $\phi(\vec{x})$ be a CQ, τ the signature τ of ϕ, $\mathcal{A}$ a τ-structure, $\vec{a}$ a tuple on $\mathcal{A}$ with $|\vec{a}| = |\vec{x}|$, and $\mathbb{G}$ a finite set of guarded sets of $\mathcal{A}$ such that every constant in $\vec{a}$ belongs to some $X \in \mathbb{G}$. If $\bigcup_{X\in\mathbb{G}} \mathcal{A}_X \models \phi[\vec{a}]$ then $\mathcal{A} \models \phi[\vec{a}]$.*

Now we are in the position to prove the theorem.

Sketched Proof of Theorem 16. (Only-if) By Proposition 14.

(If) Only address arbitrary structures. A slight modification to the following argument applies to finite structures.

Let Σ be a finite set of TGDs preserved under isomorphic unions. We first show that Σ is equivalent to a set of *diverse dependencies*, each of which is a sentence of the form

$$\forall\vec{x}(\lambda_{\mathrm{una}}(\vec{x}) \wedge \phi(\vec{x}) \rightarrow \exists\vec{y}\psi(\vec{x},\vec{y})) \tag{7}$$

where ϕ and ψ are conjunctions of relational atomic formulas, and $\lambda_{\mathrm{una}}(\vec{x})$ denotes $\bigwedge_{1\leq i<j\leq k} \neg t_i = t_j$ with $t_1,\ldots,t_k$ being an enumeration (without repetition) of all constant symbols and universal variables in ϕ and ψ. It is easy to show

Claim 1. Σ is equivalent to a finite set of diverse dependencies.

To present the proof, more notions are needed. Let σ be a diverse dependency of the form (7). The *graph* of σ is defined as an undirected graph with each conjunct of ψ as a vertex and with each pair of conjuncts of ψ that share some existential variable as an edge. We say that σ is *quasi-frontier-guarded* if, for every connected component δ of the graph of σ, the set of variables that occurs in both δ and $\vec{x}$ (the tuple of universal variables of σ) co-occur in some atomic formula in ϕ.

Let Γ be a finite set of diverse dependencies that is equivalent to Σ. Take $\gamma \in \Gamma$ as a diverse dependency of the form (7). Let S_γ denote the set of substitutions, which only map existential variables to some terms in γ, such that $s(\gamma)$ is a quasi-frontier-guarded diverse dependency. Let γ^* denote

$$\forall\vec{x}\left[\lambda_{\mathrm{una}}(\vec{x}) \wedge \phi(\vec{x}) \rightarrow \exists\vec{y} \bigvee_{s\in S_\gamma} s(\psi)(\vec{x},\vec{y})\right] \tag{8}$$

and let Γ^* be the set of γ^* for all $\gamma \in \Gamma$. We want to prove that Γ^* is equivalent to Σ. The direction $\Gamma^* \models \Sigma$ follows from the definition of Γ^*. To show the converse, it suffices to prove

Claim 2. $\Sigma \models \gamma^*$ for all $\gamma \in \Gamma$.

Proof. Let $\mathcal{A}$ be a model of Σ and $\vec{a}$ a tuple on $\mathcal{A}$ such that $\mathcal{A} \models \lambda_{\mathrm{una}}[\vec{a}]$ and $\mathcal{A} \models \phi[\vec{a}]$. Let C be the set of all interpretations of constant symbols in $\mathcal{A}$. Let $\mathbb{G}$ be the set of guarded sets of $\mathcal{A}$ such that if $X \in \mathbb{G}$ then all constants in $X \setminus C$ co-occur in an atomic formula in $\phi(\vec{a})$. Let $\mathcal{B} = \bigcup_{X\in\mathbb{G}} \mathcal{A}_X$. By definition we know $\mathcal{B} \models \lambda_{\mathrm{una}}[\vec{a}]$ and $\mathcal{B} \models \phi[\vec{a}]$. As Σ is preserved under isomorphic unions, $\mathcal{B}$ must be a model of Σ. Consequently, $\mathcal{B}$ is a model of γ. We thus have that $\mathcal{B} \models \exists\vec{y}\psi[\vec{a}/\vec{x}]$, i.e., there is a tuple $\vec{b}$ on $\mathcal{B}$ with $\mathcal{B} \models \psi[\vec{a},\vec{b}]$.

Define a substitution s as follows: Given $i = 1,\ldots,|\vec{y}|$, let $s(y_i) = c$ if for some constant symbol c with $b_i = c^{\mathcal{A}}$; if no such c then let $s(y_i) = x_j$ for some j with $b_i = a_j$; if no such j either then let $s(y_i) = y_i$, where a_i, b_i, x_i, y_i denote the i-th components of $\vec{a}, \vec{b}, \vec{x}, \vec{y}$, respectively. Clearly $\mathcal{B} \models s(\exists\vec{y}\psi)[\vec{a}/\vec{x}]$. By Lemma 17, we have $\mathcal{A} \models s(\exists\vec{y}\psi)[\vec{a}/\vec{x}]$. By a careful check, one can show $s \in S_\gamma$, i.e., $s(\gamma)$ is quasi-frontier-guarded as desired. We omit the proof here. □

With Claim 2, we then have that Γ^* is equivalent to Σ. Take $\gamma \in \Gamma$ and suppose γ^* is of the form (8). It is easy to see that γ^* can be equivalently rewritten as a sentence $\gamma^\dagger$ of the form

$$\forall\vec{x}\left[\phi(\vec{x}) \rightarrow \bigvee_{s\in S_\gamma} \exists\vec{y}s(\psi) \vee \bigvee_{1\leq i<j\leq k} x_i = x_j\right] \tag{9}$$

where $t_1,\ldots,t_k$ is an enumeration (without repetition) of all terms in γ. Let $\Gamma^\dagger$ consist of $\gamma^\dagger$ for all $\gamma \in \Gamma$, and let $\Delta(\gamma)$ be a set that consists of the TGD

$$\forall\vec{x}(\phi(\vec{x}) \rightarrow \exists\vec{y}s(\psi)) \tag{10}$$

for all $s \in S_\gamma$, and Δ the union of $\Delta(\gamma)$ for all $\gamma \in \Gamma$. Let $\mathrm{con}(\Gamma^\dagger)$ denote the set of TGDs $\sigma \in \Delta$ such that $\Gamma^\dagger \models \sigma$. It is easy to see that each TGD in $\mathrm{con}(\Gamma^\dagger)$ is equivalent to a finite number of frontier-guarded TGDs. To complete the proof, it is thus sufficient to show the following property:

Claim 3. $\mathrm{con}(\Gamma^\dagger)$ is equivalent to $\Gamma^\dagger$.

This can be proved by combining the direct-product argument that proves Theorem 11 with the strictly-homomorphic preimage preservation argument that proves Theorem 12. □

3.6 Guarded TGDs

We say that a sentence φ is *preserved under disjoint unions [in the finite]* if, for each pair of [finite] models $\mathcal{A}$ and $\mathcal{B}$ of φ, $\mathcal{A} \cup \mathcal{B}$ is also a model of φ if both the following hold: (i) $\mathcal{A}$ and $\mathcal{B}$ have the same interpretations on constant symbols, and (ii) if $X = A \cap B$ and $X \neq \emptyset$ then $\mathcal{A}|_X = \mathcal{B}|_X$.

Proposition 18. *Every set of guarded TGDs is preserved under disjoint unions [in the finite].*

The following example shows that the above property separates guarded TGDs from frontier-guarded TGDs.

Example 19. *Let σ be the following frontier-guarded TGD:*

$$E(x,y) \wedge E(y,z) \rightarrow C(y) \tag{11}$$

and let $\tau = \{C, E\}$. Let $\mathcal{A}$ and $\mathcal{B}$ be τ-structures defined by:

- *the domain of $\mathcal{A}$ is $\{a,b\}$ and the domain of $\mathcal{B}$ is $\{b,c\}$;*
- *$C^{\mathcal{A}} = C^{\mathcal{B}} = \emptyset$, $E^{\mathcal{A}} = \{(a,b)\}$, and $E^{\mathcal{B}} = \{(b,c)\}$.*

Let $X = A \cap B = \{b\}$. Clearly, $\mathcal{A}|_X = \mathcal{B}|_X$. By definition, $\mathcal{A} \cup \mathcal{B}$ is a τ-structure with $\{a,b,c\}$ as domain, interpreting C as $\emptyset$, and interpreting E as $\{(a,b),(b,c)\}$. It is easy to see that both $\mathcal{A}$ and $\mathcal{B}$ are models of σ, but $\mathcal{A} \cup \mathcal{B}$ is not. So, σ is not preserved under disjoint unions even in the finite.

Now, let us present the desired characterization.

Theorem 20. *A finite set of TGDs is equivalent to a finite set of guarded TGDs [over finite structures] iff it is preserved under disjoint unions [in the finite].*

The general idea of proving the hard direction is as follows: First show that every finite set of frontier-guarded TGDs preserved under disjoint unions [in the finite] is equivalent to a finite set of guarded TGDs [over finite structures]. As the disjoint union preservation always implies the isomorphic union preservation, by Theorem 16, we then have the desired result.

3.7 Linear TGDs

Every sentence φ is said to be *preserved under unions [in the finite]* if, for all [finite] models $\mathcal{A}$ and $\mathcal{B}$ of φ with the same interpretations on constant symbols, $\mathcal{A} \cup \mathcal{B}$ is a model of φ.

The following theorem was obtained by ten Cate *et al.*:

Theorem 21 ([ten Cate *et al.*, 2015])**.** *A finite set of TGDs is equivalent to a finite set of linear TGDs over finite structures iff it is preserved under unions in the finite.*

To separate the class of linear TGDs from guarded TGDs, a simple example is presented as follows:

Example 22. *Let σ denote the following guarded TGD:*

$$P(x) \wedge Q(x) \rightarrow R(x). \tag{12}$$

Let τ denote $\{P,Q,R\}$. Let $\mathcal{A}$ and $\mathcal{B}$ be τ-structures with the same domain $\{a\}$ such that $P^{\mathcal{A}} = Q^{\mathcal{B}} = R^{\mathcal{A}} = R^{\mathcal{B}} = \emptyset$ and $P^{\mathcal{B}} = Q^{\mathcal{A}} = \{a\}$. Then it is obvious that both $\mathcal{A}$ and $\mathcal{B}$ are models of σ. However, $\mathcal{A} \cup \mathcal{B}$ does not satisfy σ. Therefore, σ is not preserved under unions even in the finite.

It is worth noting that ten Cate *et al.*'s proof of Theorem 21 does not work over arbitrary structures. Fortunately, thanks to Theorem 16 and the finite model property of frontier-guarded TGDs, we are able to show the following characterization:

Theorem 23. *A finite set of TGDs is equivalent to a finite set of linear TGDs iff it is preserved under unions.*

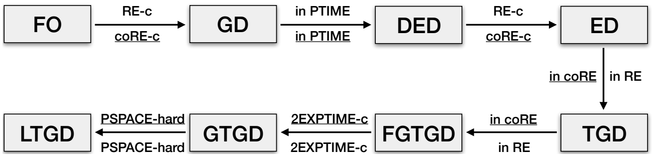

Figure 2: Complexity of Rewritability

4 Application: Complexity of Rewritability

As a direct application, we use the obtained model-theoretic characterizations to identify complexity bounds of language rewritability. Let PTIME (resp., PSPACE, 2EXPTIME) denote the class of languages accepted by some deterministic Turing machine in polynomial time (resp., polynomial space, double-exponential time). By [CO]RE we mean [the complement of] the class of recursively enumerable languages.

Let FO denote the class of all first-order sentences. Let GD (resp., DED, ED, TGD, FGTGD, GTGD and LTGD) denote the class of all finite sets of GDs (resp., DEDs, EDs, TGDs, frontier-guarded TGDs, guarded TGDs and linear TGDs).

Suppose $\mathcal{C}$ and $\mathcal{C}'$ are classes of first-order sentences, and $\mathcal{K}$ a complexity class. A sentence $\varphi \in \mathcal{C}$ is called *rewritable to $\mathcal{C}'$ [in the finite]* if there is a sentence $\psi \in \mathcal{C}'$ such that φ is equivalent to ψ [over finite strutures]. We say that the rewritability of $\mathcal{C}$ to $\mathcal{C}'$ [in the finite] is in $\mathcal{K}$ if there is a Turing machine M in $\mathcal{K}$ such that, given a sentence $\varphi \in \mathcal{C}$ as input, M accepts φ if and only if φ is rewritable to $\mathcal{C}'$ [in the finite].

Theorem 24. *The complexity of rewritability for the above existential rule languages is illustrated in Figure 2, where, along each arrow, the bound without underline indicates the complexity over arbitrary structures, and the bound with underline indicates the complexity over finite structures.*

To prove the above theorem, we only explain the idea of proving the 2EXPTIME-completeness of the rewritability of FGTGD to GTGD. By Theorem 20, it suffices to prove that recognizing the preservation of FGTGD under disjoint unions is 2EXPTIME-complete, which is proved in Statement 6 of Theorem 25. So, it remains to prove the following theorem:

Theorem 25.

1. *Determining whether a given first-order sentence is preserved under globally-homomorphic preimages [in the finite] is [co]RE-complete.*
2. *Determining whether a given finite set of GDs has both a trivial model and a sharp model is in PTIME.*
3. *Determining whether a given finite set of DEDs is preserved under direct products [in the finite] is [co]RE-complete.*
4. *Determining whether a given finite set of EDs is preserved under both strictly-homomorphic images and preimages [in the finite] is in [co]RE.*
5. *Determining whether a given finite set of EDs is preserved under isomorphic unions [in the finite] is in [co]RE.*
6. *Determine whether a given finite set of frontier-guarded TGDs is preserved under disjoint unions [in the finite] is 2EXPTIME-complete.*

7. *Determining whether a given finite set of GTGDs is preserved under unions [in the finite] is PSPACE-hard.*

Sketched proof. Only explain the general idea of proving the 2EXPTIME-membership for the complexity in Statement 6. To yield the desired membership, it suffices to prove that determining whether a given set Σ of frontier-guarded TGDs is preserved under disjoint unions [in the finite] is in 2EXPTIME. We implement this by constructing a first-order sentence φ_Σ such that Σ is preserved under disjoint unions [in the finite] iff φ_Σ is unsatisfiable [over finite structures]. Thanks to the simplicity of the disjoint-union-preservation property, φ_Σ can be expressed in the guarded negation logic, a fragment of first-order logic whose [finite] satisfiability problem is proved to be 2EXPTIME-complete [Bárány *et al.*, 2015]. □

5 Conclusion and Related Work

We have established model-theoretic characterizations for several important classes of existential rules. Very interestingly, our characterizations show that the guarded-based notions are exactly captured by union-like preservations. Since union-like preservations can be regarded as modular properties in a certain sense, this work also provides alternative perspective on why guarded-based existential rule languages enjoy good computational properties. We believe this may shed new insight on identifying new tractable languages.

There have been a number of earlier works related to ours. Over finite structures, Makowsky and Vardi [1986] established several characterizations for full TGDs (i.e., TGDs without existential quantifiers) and equality-generating dependencies; ten Cate *et al.* [2015] observed that the union preservation captures the definability of TGDs by linear TGDs. Over arbitrary structures, Lutz *et al.* [2011] established characterizations for description logics $\mathcal{EL}$ and $DL\text{-}Lite_{horn}$. Note that both $\mathcal{EL}$ and $DL\text{-}Lite_{horn}$ are sublanguages of existential rule languages. Bárány *et al.* [2013] proved that every TGDs-defined first-order sentence in the guarded negation fragment is definable by frontier-guarded TGDs. Moreover, in the setting of schema mapping, ten Cate and Kolaitis [2010] estashlished a number of characterizations for source-to-target TGDs (a class of acyclic TGDs) and its subclasses; in the setting of ontology-mediated query answering, Zhang *et al.* [2016] characterizes the class of DEDs by using both complexity- and model-theoretic properties.

Acknowledgments

We are deeply indebted to Professor Carsten Lutz for insightful discussions. This work was partially supported by the National Natural Science Foundation of China (No. 61806102).

References

[Abiteboul *et al.*, 1995] Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of Databases*. Addison-Wesley, 1995.

[Baget *et al.*, 2011] Jean-François Baget, Michel Leclère, Marie-Laure Mugnier, and Eric Salvat. On rules with existential variables: Walking the decidability line. *Artif. Intell.*, 175(9-10):1620–1654, 2011.

[Bárány *et al.*, 2013] Vince Bárány, Michael Benedikt, and Balder ten Cate. Rewriting guarded negation queries. In *Proceedings of MFCS-2013*, pages 98–110, 2013.

[Bárány *et al.*, 2015] Vince Bárány, Balder ten Cate, and Luc Segoufin. Guarded negation. *J. ACM*, 62(3):22, 2015.

[Beeri and Vardi, 1981] Catriel Beeri and Moshe Y. Vardi. The implication problem for data dependencies. In *Proceedings of ICALP-1981*, pages 73–85, 1981.

[Calì *et al.*, 2010] Andrea Calì, Georg Gottlob, Thomas Lukasiewicz, Bruno Marnette, and Andreas Pieris. Datalog+/-: A family of logical knowledge representation and query languages for new applications. In *Proceedings of LICS-2010*, pages 228–242, 2010.

[Calì *et al.*, 2012] Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. A general datalog-based framework for tractable query answering over ontologies. *J. Web Sem.*, 14:57–83, 2012.

[Calì *et al.*, 2013] Andrea Calì, Georg Gottlob, and Michael Kifer. Taming the infinite chase: Query answering under expressive relational constraints. *J. Artif. Intell. Res.*, 48:115–174, 2013.

[Chang and Keisler, 1990] C. C. Chang and H. J. Keisler. *Model Theory*. Studies in Logic and the Foundations of Mathematics. Elsevier Science, 1990.

[Fagin *et al.*, 2005] Ronald Fagin, Phokion Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: Semantics and query answering. *Theor. Comput. Sci.*, 336(1):89–124, 2005.

[Gottlob and Papadimitriou, 2003] Georg Gottlob and Christos H. Papadimitriou. On the complexity of single-rule datalog queries. *Inf. Comput.*, 183(1):104–122, 2003.

[Hull, 1984] Richard Hull. Finitely specifiable implicational dependency families. *J. ACM*, 31(2):210–226, 1984.

[Lenzerini, 2002] Maurizio Lenzerini. Data integration: A theoretical perspective. In *Proceedings of PODS-2002*, pages 233–246, 2002.

[Lutz *et al.*, 2011] Carsten Lutz, Robert Piro, and Frank Wolter. Description logic tboxes: Model-theoretic characterizations and rewritability. In *Proceedings of IJCAI-2011*, pages 983–988, 2011.

[Makowsky and Vardi, 1986] Johann A. Makowsky and Moshe Y. Vardi. On the expressive power of data dependencies. *Acta Inf.*, 23(3):231–244, 1986.

[Rosen, 2002] Eric Rosen. Some aspects of model theory and finite structures. *Bulletin of Symbolic Logic*, 8(3):380–403, 2002.

[ten Cate and Kolaitis, 2010] Balder ten Cate and Phokion Kolaitis. Structural characterizations of schema-mapping languages. *Commun. ACM*, 53(1):101–110, 2010.

[ten Cate and Kolaitis, 2014] Balder ten Cate and Phokion Kolaitis. Schema mappings: A case of logical dynamics in database theory. In *Johan van Benthem on Logic and Information Dynamics*, pages 67–100, 2014.

[ten Cate *et al.*, 2015] Balder ten Cate, Gaëlle Fontaine, and Phokion Kolaitis. On the data complexity of consistent query answering. *Theory Comput. Syst.*, 57(4):843–891, 2015.

[Zhang *et al.*, 2016] Heng Zhang, Yan Zhang, and Jia-Huai You. Expressive completeness of existential rule languages for ontology-based query answering. In *Proceedings of IJCAI-2016*, pages 1330–1337, 2016.