THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Model-theoretic Characterizations of Rule-based Ontologies

Citation for published version:
Console, M, Kolaitis, PG & Pieris, A 2021, Model-theoretic Characterizations of Rule-based Ontologies. in *Proceedings of the 40th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems.* ACM Association for Computing Machinery, pp. 416-428, ACM SIGMOD/PODS International Conference on Management Data , Xi'an, China, 20/06/21. https://doi.org/10.1145/3452021.3458310

Digital Object Identifier (DOI):
10.1145/3452021.3458310

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Proceedings of the 40th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 17. May. 2022

Model-theoretic Characterizations of Rule-based Ontologies

Marco Console
University of Edinburgh
mconsole@inf.ed.ac.uk

Phokion G. Kolaitis
University of California, Santa Cruz &
IBM Research - Almaden
kolaitis@ucsc.edu

Andreas Pieris
University of Edinburgh
apieris@inf.ed.ac.uk

ABSTRACT

An ontology specifies an abstract model of a domain of interest via a formal language that is typically based on logic. Although description logics are popular formalisms for modeling ontologies, it is generally agreed that tuple-generating dependencies (tgds), originally introduced as a unifying framework for database integrity constraints, and later on used in data exchange and integration, are well suited for modeling ontologies that are intended for data-intensive tasks. The reason is that, unlike description logics, tgds can easily handle higher-arity relations that naturally occur in relational databases. In recent years, there has been an extensive study of tgd-ontologies and of their applications to several different data-intensive tasks. However, the fundamental question of whether the expressive power of tgd-ontologies can be characterized in terms of model-theoretic properties remains largely unexplored. We establish several characterizations of tgd-ontologies, including characterizations of ontologies specified by such central classes of tgds as full, linear, guarded, and frontier-guarded tgds. Our characterizations use the well-known notions of critical instance and direct product, as well as a novel locality property for tgd-ontologies. We further use this locality property to decide whether an ontology expressed by frontier-guarded (respectively, guarded) tgds can be expressed by tgds in the weaker class of guarded (respectively, linear) tgds, and effectively construct such an equivalent ontology if one exists.

ACM Reference Format:
Marco Console, Phokion G. Kolaitis, and Andreas Pieris. 2020. Model-theoretic Characterizations of Rule-based Ontologies. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Model theory is the study of the interaction between formulas in some logical formalism and their models, that is, structures that satisfy the formulas. There are two directions in this interaction, namely, from syntax to semantics and from semantics to syntax. The first direction aims to identify structural properties possessed by all models of formulas having common syntactical features. For example, it is easy to show that every universal first-order sentence is preserved under substructures. The second direction aims to characterize formulas in terms of their structural properties. For example, the well known Łos-Tarski Theorem asserts that if a first-order sentence is preserved under substructures, then it is logically equivalent to a universal first-order sentence. In general, establishing results in the second direction is a much harder task than establishing results in the first. In other words, obtaining model-theoretic characterizations of formulas or of classes of formulas is a far greater challenge than identifying structural properties possessed by all models of formulas with common syntactic features.

Makowsky and Vardi [14] were the first to obtain model-theoretic characterizations of classes of database dependencies expressed in suitable fragments of first-order logic. Furthermore, they classified the work on model-theoretic characterizations into two distinct approaches, which they called the *preservation approach* and the *axiomatizability approach*. In the preservation approach, one considers two logical formalisms $\mathcal{L}$ and $\mathcal{L}'$, where $\mathcal{L}'$ is typically a proper fragment of $\mathcal{L}$. The goal is to obtain model-theoretic characterizations of the form: a set Σ of $\mathcal{L}$-formulas is equivalent to a set Σ' of $\mathcal{L}'$-formulas if and only if the models of Σ satisfy certain structural properties. For example, the aforementioned Łos-Tarski Theorem is a prototypical example in the preservation approach. In the axiomatizability approach, one considers a logical formalism $\mathcal{L}$ and the goal is to obtain model-characterizations of the form: a class C of structures is axiomatizable by a set of $\mathcal{L}$-formulas if and only if the structures in C satisfy certain structural properties. Makowsky and Vardi [14] further distinguished the special case of *finite axiomatizability*, where the goal is to obtain model-theoretic characterizations of when a class of structures is axiomatizable by a finite set of $\mathcal{L}$-formulas. They then obtained model-theoretic characterizations of axiomatizability and finite axiomatizability of various classes of database dependencies.

As is well known, database dependencies were originally used to formalize integrity constraints in databases with much of the early work in this area focusing on the implication problem between database dependencies (see [10] for a survey). The two most prominent classes of databases dependencies are the class of tuple-generating dependencies (tgds) and the class of equality-generating dependencies (egds). By definition, a tgd is a first-order sentence of the form $\forall\bar{x}\forall\bar{y}\,(\phi(\bar{x},\bar{y}) \rightarrow \exists\bar{z}\,\psi(\bar{x},\bar{z}))$, where $\phi(\bar{x},\bar{y})$ is a (possibly empty) conjunction of atoms over a schema S, and $\psi(\bar{x},\bar{z})$ is a (non-empty) conjunction of atoms over S. Similarly, an egd is a first-order sentence of the form $\forall\bar{x}\,\left(\phi(\bar{x}) \rightarrow x_i = x_j\right)$, where $\phi(\bar{x})$ is a non-empty conjunction of atoms over a schema S and x_i, x_j are variables from $\bar{x}$. Later on, tgds and egds found numerous uses in other data management tasks. In particular, they have been used as schema-mapping specification languages and, as such, have been (and still are) successfully deployed in the study of data exchange [1, 9] and data integration [11]. More recently, tgds and egds have been extensively used in ontologies as we describe next.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference'17, July 2017, Washington, DC, USA
© 2020 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

An ontology specifies an abstract model of a domain of interest using a suitable logical formalism. In particular, description logics have been extensively used as ontology languages [2]. Description logic axioms typically involve unary and binary predicates that correspond, respectively, to concepts in the ontology and roles between two different concepts in the ontology. As it turns out, many axioms used in description logics can be expressed as tgds or egds over relational schemas consisting of unary and binary predicates. Since data-intensive tasks, such as ontology-mediated query answering [4], typically involve predicates of higher arities, it is now generally agreed that tgds and egds are well-suited as specification languages for ontologies. As a matter of fact, tgds have been extensively studied in the context of ontologies under different names, such as Datalog+/- [7] and existential rules [15].

Let us now summarize the known model-theoretic characterizations of tgds, egds, and description logics. Concerning the preservation approach, Lutz et al. [12] obtained model-theoretic characterizations of description logics, i.e., they characterized when a finite set of axioms in certain expressive description logics is equivalent to a finite set of axioms in some less expressive description logic. They also characterized when a first-order sentence is equivalent to a finite set of axioms in a certain description logic. Quite recently, Zhang et al. [17] obtained model-theoretic characterizations of existential rule languages in the preservation approach. In particular, they characterized when a finite set of tgds is equivalent to (i) a finite set of frontier-guarded tgds; (ii) a finite set of guarded tgds; and (iii) a finite set of linear tgds. The preservation approach between two logical formalisms $\mathcal{L}$ and $\mathcal{L}'$ gives rise to a natural decision problem, which we will denote by $\mathsf{Rewrite}(\mathcal{L}, \mathcal{L}')$: given a finite set of Σ of $\mathcal{L}$-formulas, is there a finite set Σ' of $\mathcal{L}'$ formulas such that Σ is equivalent to Σ'? The papers [12] and [17] contain a number of complexity results about the decision problem $\mathsf{Rewrite}(\mathcal{L}, \mathcal{L}')$ for various logical formalisms $\mathcal{L}$ and $\mathcal{L}'$.

Concerning the axiomatizability/finite axiomatizability approach, we mentioned earlier that Makowsky and Vardi [14] obtained model-theoretic characterizations of database dependencies. Specifically, the main results in [14] are model-theoretic characterizations of sets of full tgds and egds. Also in the axiomatizability/finite axiomatizability approach, Kolaitis and ten Cate [16] obtained model theoretic characterizations of source-to-target tgds, which are the tgds used to formalize data exchange between a source schema and a target schema. These results notwithstanding, however, the study of model-theoretic characterizations of sets of arbitrary tgds in the axiomatizability/finite axiomatizability approach has remained largely unexplored so far.

Summary of Results. Motivated by the preceding state of affairs, we embark here on a systematic investigation of model-theoretic characterizations of classes of tgds in the finite axiomatizability approach. This investigation is carried out in the context of ontologies. From a syntactic point of view, an ontology is specified by a set of formulas in some formalism. From a semantic point of view, an ontology can be identified with the set of all structures (finite or infinite) that satisfy the formulas specifying the ontology. Thus, as a semantic object, an ontology is an isomorphism-closed class of structures (finite or infinite) over some fixed relational schema.

Our goal is to answer the following question: what are necessary and sufficient conditions for an ontology (an isomorphism-closed class of structures) to be specified by a set of tgds? The main outcome of this investigation is to characterize the ontologies that are finitely axiomatizable by a finite set of arbitrary tgds or by a finite set of tgds that belong to one of the main subclasses of tgds, namely, full, frontier-guarded, guarded, and linear tgds.

Our model-theoretic characterizations make use of structural properties encountered in earlier model-theoretic characterizations, such as the ontology being closed under direct products and containing critical structures of every finite cardinality, where a structure is *critical* if each of its relations contains all possible tuples from the domain of the structure. The main innovation, however, is the introduction and use of the notion of (n, m)-*locality*, where, intuitively, n represents the number of universal quantifiers in the tgds and m represents the number of existential quantifiers in the tgds. Several different notions of locality have been used in earlier model-theoretic characterizations of restricted classes of tgds (e.g., in [14, 16]), yet none of them can be used in characterizations of arbitrary sets of tgds. Our first main result asserts that an ontology O is axiomatizable by a finite set of tgds if and only if O is closed under direct products, contains critical structures of every finite cardinality, and is (n, m)-local for some non-negative integers n and m. The notion of (n, m)-locality turns turns out to be delicate, yet flexible. Indeed, this notion can be tailored to other classes of tgds, so that it gives rise to the refined notions of frontier-guarded (n, m)-locality, guarded (n, m)-locality, and linear (n, m)-locality. Using these refined notions, we obtain model theoretic characterizations of ontologies that are axiomatizable by a finite set of, respectively, frontier-guarded, guarded, and linear tgds.

Finally, we investigate the decision problems $\mathsf{Rewrite}(\mathcal{L}, \mathcal{L}')$, where $\mathcal{L}$ is the class of frontier-guarded tgds and $\mathcal{L}'$ is the class of guarded tgds, or $\mathcal{L}$ is the class of guarded tgds and $\mathcal{L}'$ is the class of linear tgds. In both these cases, we obtain complexity results that significantly sharpen the results established in [17].

2 PRELIMINARIES

Let $\mathbf{C}$ and $\mathbf{V}$ be disjoint countably infinite sets of constants and variables, respectively. For $n > 0$, let $[n]$ be the set $\{1, \ldots, n\}$.

Relational Instances. A *schema* $\mathbf{S}$ is a finite set of relation symbols (or predicates) with associated (positive) arity; we write $\mathsf{ar}(R)$ for the arity of R. An *instance*[1] I over $\mathbf{S} = \{R_1, \ldots, R_n\}$, or $\mathbf{S}$-*instance*, is a tuple $(\mathsf{dom}(I), R_1^I, \ldots, R_n^I)$, where $\mathsf{dom}(I) \subseteq \mathbf{C}$ is a (finite or infinite) domain, and $R_1^I, \ldots, R_n^I$ are relations over $\mathsf{dom}(I)$, i.e., $R_i^I \subseteq \mathsf{dom}(I)^{\mathsf{ar}(R_i)}$ for $i \in [n]$. A *fact* of I is an expression of the form $R_i(\bar{c})$, where $\bar{c} \in R_i^I$; let $\mathsf{facts}(I)$ be the set of facts of I. For an $\mathbf{S}$-instance $J = (\mathsf{dom}(J), R_1^J, \ldots, R_n^J)$, we write $J \subseteq I$ if $\mathsf{facts}(J) \subseteq \mathsf{facts}(I)$. We say that J is a *subinstance* of I, denoted $J \preceq I$, if $\mathsf{dom}(J) \subseteq \mathsf{dom}(I)$, and $R^J = R^I_{|\mathsf{dom}(J)}$ for each $R \in \mathbf{S}$ with $R^I_{|\mathsf{dom}(J)}$ being the *restriction of* R^I *over* $\mathsf{dom}(J)$, i.e., the relation $\{\bar{c} \in R^I \mid \bar{c} \in \mathsf{dom}(J)^{\mathsf{ar}(R)}\}$. Note that $J \preceq I$ implies $J \subseteq I$, but the other direction does not hold. A *homomorphism* from I to J is a function $h : \mathsf{dom}(I) \to \mathsf{dom}(J)$ such that, for each $i \in [n]$, $(c_1, \ldots, c_m) \in R_i^I$ implies $(h(c_1), \ldots, h(c_m)) \in$

[1] In mathematical logic literature, instances are called *structures*, which is the term we used in the Introduction. For the remainder of this paper, we adopt the term *instances*.

R_i^J. We write $h : I \rightarrow J$ for the fact that h is a homomorphism from I to J. We also write $h(\text{facts}(I))$ for the set $\{R(h(\bar{c})) \mid R(\bar{c}) \in \text{facts}(I)\}$. Finally, we say that I and J are *isomorphic*, written $I \simeq J$, if there is an 1-1 homomorphism from I to J such that $h^{-1} : J \rightarrow I$.

Tuple-Generating Dependencies. An atom over $\mathbf{S}$ is an expression of the form $R(\bar{v})$, where $R \in \mathbf{S}$ and $\bar{v}$ is an $\text{ar}(R)$-tuple of variables from $\mathbf{V}$. A *tuple-generating dependency* (tgd) σ over a schema $\mathbf{S}$ is a constant-free first-order sentence $\forall\bar{x}\forall\bar{y}\,(\phi(\bar{x},\bar{y}) \rightarrow \exists\bar{z}\,\psi(\bar{x},\bar{z}))$, where $\bar{x},\bar{y},\bar{z}$ are tuples of variables of $\mathbf{V}$, $\phi(\bar{x},\bar{y})$ is a (possibly empty) conjunction of atoms over $\mathbf{S}$, and $\psi(\bar{x},\bar{z})$ is a (non-empty) conjunction of atoms over $\mathbf{S}$. For brevity, we write σ as $\phi(\bar{x},\bar{y}) \rightarrow \exists\bar{z}\,\psi(\bar{x},\bar{z})$, and use comma instead of $\wedge$ for joining atoms. We refer to $\phi(\bar{x},\bar{y})$ and $\psi(\bar{x},\bar{z})$ as the *body* and *head* of σ, denoted $\text{body}(\sigma)$ and $\text{head}(\sigma)$, respectively. By abuse of notation, we may treat a tuple of variables as a set of variables, and a conjunction of atoms as a set of atoms. An instance I satisfies a tgd σ as the one above, written $I \models \sigma$, if the following holds: whenever there exists a function $h : \bar{x} \cup \bar{y} \rightarrow \text{dom}(I)$ such that $h(\phi(\bar{x},\bar{y})) \subseteq \text{facts}(I)$ (as usual, we write $h(\phi(\bar{x},\bar{y}))$ for the set $\{R(h(\bar{v})) \mid R(\bar{v}) \in \phi(\bar{x},\bar{y})\}$), then there exists an extension h' of h such that $h'(\psi(\bar{x},\bar{z})) \subseteq \text{facts}(I)$. The instance I satisfies a set Σ of tgds, written $I \models \Sigma$, in which case we say that I is a *model* of Σ, if $I \models \sigma$ for each $\sigma \in \Sigma$.

We write $\mathsf{TGD}_{n,m}$ for the family of all possible finite sets of tgds with at most $n \geq 0$ universally, and at most $m \geq 0$ existentially quantified variables.[2] We also write TGD for the family of all possible finite sets of TGDs, i.e., $\mathsf{TGD} = \bigcup_{n\geq 0, m\geq 0} \mathsf{TGD}_{n,m}$.

Classes of Tuple-Generating Dependencies. In our analysis, we will consider the following central classes of tgds:

Full. A tgd σ is *full* if it has no existentially quantified variables. The class of full tgds, i.e., the family of all possible finite sets of full tgds, is denoted FTGD. Notice that FTGD coincides with the class $\bigcup_{n>0} \mathsf{TGD}_{n,0}$.

Linear. A tgd σ is *linear* if $\text{body}(\sigma)$ has at most one atom. The class of linear tgds is denoted LTGD.

Guarded. A tgd σ is *guarded* if $\text{body}(\sigma)$ is either empty, or has an atom that contains all the universally quantified variables of σ. The class of guarded tgds is denoted GTGD.

Frontier-Guarded. The *frontier* of a tgd σ, denoted $\text{fr}(\sigma)$, is the set of universally quantified variables occurring in $\text{head}(\sigma)$. A tgd σ is *frontier-guarded* if $\text{body}(\sigma)$ is either empty, or has an atom that contains all the variables of $\text{fr}(\sigma)$. The class of frontier-guarded tgd is denoted FGTGD.

It is not difficult to verify that

$$\mathsf{LTGD} \subsetneq \mathsf{GTGD} \subsetneq \mathsf{FGTGD} \neq \mathsf{FTGD}.$$

Given a class C of finite sets of tgds, i.e., $\mathsf{C} \subseteq \mathsf{TGD}$, we can naturally define the class $\mathsf{C}_{n,m}$, for $n,m \geq 0$, as the class $\mathsf{C} \cap \mathsf{TGD}_{n,m}$. For example, $\mathsf{GTGD}_{n,m}$ is the class of guarded tgds with at most $n \geq 0$ universally, and at most $m \geq 0$ existentially quantified variables.

Ontologies. An *ontology* O over a schema $\mathbf{S}$ is a (finite or infinite) set of $\mathbf{S}$-instances closed under isomorphisms, i.e., if $I \in O$ and J is an $\mathbf{S}$-instance such that $I \simeq J$, then $J \in O$. Given a class C of tgds, O is a C-*ontology* if there is a set $\Sigma \in \mathsf{C}$ such that $I \in O$ iff $I \models \Sigma$.

[2] A tgd has at least one variable, and thus, $\mathsf{TGD}_{0,0}$ contains only the empty set of tgds.

3 MODEL-THEORETIC PROPERTIES

We proceed to introduce three model-theoretic properties of ontologies that will play a crucial role in our characterizations. In fact, will turn out that those three properties are enough to characterize when an ontology is a TGD-ontology. The first two properties rely on the well-known notions of critical instance and direct product, which have been used in several different contexts (see, e.g., [14]), whereas the third one relies on a novel locality property, which we consider as one of the main conceptual contributions of the present work. In the rest, we fix an arbitrary schema $\mathbf{S} = \{R_1, \ldots, R_\ell\}$.

3.1 Criticality

An $\mathbf{S}$-instance $I = (\text{dom}(I), R_1^I, \ldots, R_\ell^I)$ is called k-*critical*, for some $k > 0$, if $|\text{dom}(I)| = k$, i.e., the domain of I consists of k distinct constants from $\mathbf{C}$, and $R_i^I = \text{dom}(I)^{\text{ar}(R_i)}$ for each $i \in [\ell]$, or, in other words, $\text{facts}(I) = \{R_i(\bar{c}) \mid i \in [\ell] \text{ and } \bar{c} \in \text{dom}(I)^{\text{ar}(R_i)}\}$. For example, the $\{R\}$-instance I, with R being a binary relation, such that $\text{dom}(I) = \{c,d\}$ and $\text{facts}(I) = \{R(c,c), R(c,d), R(d,c), R(d,d)\}$ is 2-critical. An ontology O over $\mathbf{S}$ is k-critical if it contains a k-critical $\mathbf{S}$-instance. We can now define the notion of critical ontology.

Definition 3.1 (**Criticality**). An ontology is called *critical* if it is k-critical for each integer $k > 0$ ■

It is easy to show that:

Lemma 3.2. *Every* TGD-*ontology is critical.*

Proof. Consider a TGD-ontology O over $\mathbf{S}$. By definition, there exists a finite set Σ of tgds over $\mathbf{S}$ such that, for every $\mathbf{S}$-instance I, we have that $I \in O$ iff $I \models \Sigma$. Fix an arbitrary integer $k > 0$, and a k-critical $\mathbf{S}$-instance I. We proceed to show that $I \models \Sigma$, which in turn implies that $I \in O$, and thus O is critical. Consider a tgd $\sigma \in \Sigma$ of the form $\phi(\bar{x},\bar{y}) \rightarrow \exists\bar{z}\,\psi(\bar{x},\bar{z})$. It is clear that there exists a function $h : \bar{x} \cup \bar{y} \rightarrow \text{dom}(I)$ such that $h(\phi(\bar{x},\bar{y})) \subseteq \text{facts}(I)$. Let $c \in \text{dom}(I)$, and consider the extension h' of h such that $h(z) = c$ for each $z \in \bar{z}$. Since I is k-critical, it is straightforward to verify that $h'(\psi(\bar{x},\bar{z})) \subseteq \text{facts}(I)$, which implies that $I \models \sigma$, as needed. □

3.2 Closure Under Direct Products

The *direct product* of the $\mathbf{S}$-instances $I = (\text{dom}(I), R_1^I, \ldots, R_\ell^I)$ and $J = (\text{dom}(J), R_1^J, \ldots, R_\ell^J)$, denoted $I \otimes J$, is the $\mathbf{S}$-instance $K = (\text{dom}(K), R_1^K, \ldots, R_\ell^K)$, where $\text{dom}(K) = \text{dom}(I) \times \text{dom}(J)$, and

$$R_i^K = \Big\{\Big((a_1,b_1), \ldots, (a_{\text{ar}(R_i)}, b_{\text{ar}(R_i)})\Big) \mid \Big(a_1, \ldots, a_{\text{ar}(R_i)}\Big) \in R_i^I \text{ and } \Big(b_1, \ldots, b_{\text{ar}(R_i)}\Big) \in R_i^J\Big\},$$

for each $i \in [\ell]$. The model-theoretic property of interest follows:

Definition 3.3 (**$\otimes$-closure**). An ontology O over $\mathbf{S}$ is *closed under direct products* if, for every two $\mathbf{S}$-instances $I, J \in O$, $I \otimes J \in O$. ■

It is not difficult to show the following, which is actually implicit in [8], but we provide a proof for the sake of completeness.

Lemma 3.4. *Every* TGD-*ontology is closed under direct products.*

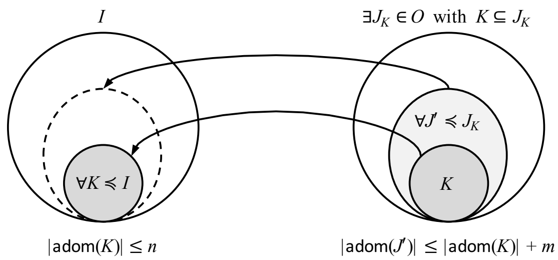

Figure 1: O **is** (n, m)**-locally embeddable in** I.

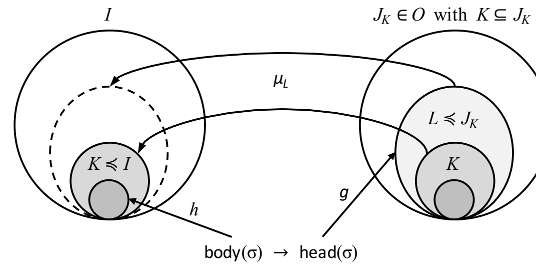

Figure 2: The function $\lambda = \mu_L \circ g$ **in the proof of Lemma 3.6.**

Proof. Consider a TGD-ontology O over $\mathbf{S}$. By definition, there exists a finite set Σ of tgds over $\mathbf{S}$ such that, for every $\mathbf{S}$-instance I, $I \in O$ iff $I \models \Sigma$. Consider two instances $I, J \in O$, and let $K = I \otimes J$. Our goal is to show that $K \in O$, or, equivalently, $K \models \Sigma$. Consider a tgd $\sigma \in \Sigma$ of the form $\phi(\bar{x}, \bar{y}) \rightarrow \exists \bar{z}\, \psi(\bar{x}, \bar{z})$, and assume that there exists a function $h : \bar{x} \cup \bar{y} \rightarrow \mathsf{dom}(K)$ such that $h(\phi(\bar{x}, \bar{y})) \subseteq \mathsf{facts}(K)$. We show that there exists an extension h' of h such that $h'(\psi(\bar{x}, \bar{z})) \subseteq \mathsf{facts}(K)$. We first observe that $h_I : K \rightarrow I$ and $h_J : K \rightarrow J$, where, for each $(a, b) \in \mathsf{dom}(K)$, $h_I((a, b)) = a$ and $h_J((a, b)) = b$. This holds since $(a_1, b_1), \ldots, (a_n, b_n) \in R^K$ iff $(a_1, \ldots, a_n) \in R^I$ and $(b_1, \ldots, b_n) \in R^J$, for every $R \in \mathbf{S}$. Therefore, $h_I \circ h$ maps $\phi(\bar{x}, \bar{y})$ to $\mathsf{facts}(I)$, and $h_J \circ h$ maps $\phi(\bar{x}, \bar{y})$ to $\mathsf{facts}(J)$. Since $I \models \Sigma$ and $J \models \Sigma$, there is an extension h'_I of $h_I \circ h$, and an extension h'_J of $h_J \circ h$ such that $h'_I(\psi(\bar{x}, \bar{z})) \subseteq \mathsf{facts}(I)$ and $h'_J(\psi(\bar{x}, \bar{z})) \subseteq \mathsf{facts}(J)$. Consider now the extension h' of h such that $h'(z) = (h'_I(z), h'_J(z))$, for every $z \in \bar{z}$. We proceed to show that $h'(\psi(\bar{x}, \bar{z})) \subseteq \mathsf{facts}(K)$. Consider $R(w_1, \ldots, w_n) \in \psi(\bar{x}, \bar{z})$, and assume that $h'(R(w_1, \ldots, w_n)) = R((a_1, b_1), \ldots, (a_n, b_n))$. By the definition of h', $R(a_1, \ldots, a_n) \in \mathsf{facts}(I)$ and $R(b_1, \ldots, b_n) \in \mathsf{facts}(J)$. Therefore, by construction, $R((a_1, b_1), \ldots, (a_n, b_n)) \in \mathsf{facts}(K)$. $\square$

3.3 Locality

We now proceed to introduce our new locality property of ontologies, which in turn relies on the notion of local embedding of an ontology in an instance. Roughly speaking, an ontology O over $\mathbf{S}$ is locally embeddable in an $\mathbf{S}$-instance I if, for every subinstance K of I with a bounded number of active domain elements (i.e., domain elements that occur in $\mathsf{facts}(K)$), we can find an instance $J_K \in O$ such that every local neighbour of K in J_K (i.e., subinstances of J_K that contain K and have a bounded number of additional active domain elements not in $\mathsf{facts}(K)$), can be embedded in I while preserving K. We call the ontology O local if every $\mathbf{S}$-instance that is locally embeddable in O is a member of O. We proceed to formalize the above high-level description.

The *active domain* of an instance I, denoted $\mathsf{adom}(I)$, is the set of elements of $\mathsf{dom}(I)$ that occur in at least one fact of I. Consider an $\mathbf{S}$-instance J and a finite set of constants $F \subseteq \mathsf{adom}(J)$. For an integer $m \geq 0$, the *m-neighbourhood of F in J* is the set of $\mathbf{S}$-instances

$$\{J' \mid F \subseteq \mathsf{adom}(J'),\ J' \preceq J \ \text{ and } \ |\mathsf{adom}(J')| \leq |F| + m\},$$

i.e., all the subinstances of J such that their facts contain constants from F and at most m additional elements not occurring in F. Furthermore, given an $\mathbf{S}$-instance $K \subseteq J$, the *m-neighbourhood of K in J* is defined as the m-neighbourhood of $\mathsf{adom}(K)$ in J, that is, all the subinstances of J that contain K and their facts mention at most m additional elements not occurring in the facts of K.

Consider an ontology O over $\mathbf{S}$, and an $\mathbf{S}$-instance I. For integers $n, m \geq 0$, we say that O *is (n, m)-locally embeddable in I* if, for every $K \preceq I$ with $|\mathsf{adom}(K)| \leq n$, there is $J_K \in O$ such that $K \subseteq J_K$, and for every J' in the m-neighbourhood of K in J_K, there is a function $h : \mathsf{adom}(J') \rightarrow \mathsf{adom}(I)$, which is the identity on $\mathsf{adom}(K)$, such that $h(\mathsf{facts}(J')) \subseteq \mathsf{facts}(I)$. An illustration of when O is (n, m)-locally embeddable in I is shown in Figure 1; the circles represent the set of facts of the instances. The key property of locality follows.

Definition 3.5 (**Locality**). An ontology O over $\mathbf{S}$ is (n, m)*-local*, for $n, m \geq 0$, if, for every $\mathbf{S}$-instance I, the following holds: O is (n, m)-locally embeddable in I implies $I \in O$. We further say that O is *local* if there exist $n, m \geq 0$ such that O is (n, m)-local. ■

The next result states that every TGD-ontology is local. Actually, it shows a stronger claim since it relates the integers n, m that witness (n, m)-locality with the number of universally and existentially quantified variables, respectively, that can occur in the tgds.

Lemma 3.6. *For $n, m \geq 0$, every $\mathsf{TGD}_{n,m}$-ontology is (n, m)-local.*

Proof. Consider a $\mathsf{TGD}_{n,m}$-ontology O over $\mathbf{S}$. By definition, there is a set $\Sigma \in \mathsf{TGD}_{n,m}$ (i.e., each tgd of Σ mentions at most n universally and m existentially quantified variables), such that, for every $\mathbf{S}$-instance I, $I \in O$ iff $I \models \Sigma$. Consider an $\mathbf{S}$-instance I, and assume that O is (n, m)-locally embeddable in I. We proceed to show that $I \in O$, or, equivalently, $I \models \Sigma$. Consider a tgd $\sigma \in \Sigma$ of the form $\phi(\bar{x}, \bar{y}) \rightarrow \exists \bar{z}\, \psi(\bar{x}, \bar{z})$, and assume that there exists a function $h : \bar{x} \cup \bar{y} \rightarrow \mathsf{dom}(I)$ such that $h(\phi(\bar{x}, \bar{y})) \subseteq \mathsf{facts}(I)$. We show that there exists an extension λ of h such that $\lambda(\psi(\bar{x}, \bar{z})) \subseteq \mathsf{facts}(I)$; the existence of λ is illustrated in Figure 2. Let $K = (\mathsf{dom}(K), R_1^K, \ldots, R_\ell^K)$ where $\mathsf{dom}(K)$ is the set of constants occurring in $h(\phi(\bar{x}, \bar{y}))$, and, for each $i \in [\ell]$, $R_i^K = R^I_{i\,|K}$. It is clear that $K \preceq I$ with $|\mathsf{adom}(K)| \leq n$ since $\phi(\bar{x}, \bar{y})$ mentions at most n variables. Since, by hypothesis, O is (n, m)-locally embeddable in I, we conclude that there exists $J_K \in O$ such that $K \subseteq J_K$, and, for every J' in the m-neighbourhood of K in J_K, there is a function $\mu_{J'} : \mathsf{adom}(J') \rightarrow \mathsf{adom}(I)$, which is the identity on $\mathsf{adom}(K)$, such that $\mu_{J'}(\mathsf{facts}(J')) \subseteq \mathsf{facts}(I)$. It is clear that $h(\phi(\bar{x}, \bar{y})) \subseteq \mathsf{facts}(J_K)$. Since $J_K \in O$, or, equivalently, $J_K \models \Sigma$, there exists an extension g of h such that $g(\psi(\bar{x}, \bar{z})) \subseteq \mathsf{facts}(J_K)$. Let $L = (\mathsf{dom}(L), R_1^L, \ldots, R_\ell^L)$ where $\mathsf{dom}(L)$ are the constants occurring in $h(\phi(\bar{x}, \bar{y})) \cup g(\psi(\bar{x}, \bar{z}))$, and, for each $i \in [\ell]$, $R_i^L = R_i^{J_K}{}_{|\mathsf{dom}(L)}$. It is clear that L is in the m-neighbourhood of K in J_K since $\bar{z}$ has at most m variables. Therefore, there is a function $\mu_L : \mathrm{adom}(L) \to \mathrm{adom}(I)$, which is the identity on $\mathrm{adom}(K)$, such that $\mu_L(\mathrm{facts}(L)) \subseteq \mathrm{facts}(I)$. Consider the function $\lambda = \mu_L \circ g$. Since g is an extension of h, and μ_L is the identity on the elements occurring in $h(\phi(\bar{x}, \bar{y}))$, we get that $\lambda(v) = h(v)$ for each variable v in $\phi(\bar{x}, \bar{y})$, and thus, λ is an extension of h. Moreover, since $g(\psi(\bar{x}, \bar{z})) \subseteq \mathrm{facts}(L)$, we get that $\lambda(\psi(\bar{x}, \bar{z})) \subseteq \mathrm{facts}(I)$. □

Domain independence is another central property of ontologies, which will play a crucial role in our technical development, and, as we shall see below, it is guaranteed by locality.

Definition 3.7 (**Domain Independence**). An ontology O over $\mathbf{S}$ is called *domain independent* if, for every two $\mathbf{S}$-instances I and J such that $\mathrm{facts}(I) = \mathrm{facts}(J)$, $I \in O$ iff $J \in O$. ■

In simple words, O is domain independent if, for every two $\mathbf{S}$-instances that have the same set of facts, but not necessarily the same domain, either are both in O, or none of them is in O. It is not difficult to show that locality implies domain independence.

Lemma 3.8. *Every local ontology is domain independent.*

Proof. Consider an ontology O over $\mathbf{S}$ that is (n, m)-local for $n, m \geq 0$, and two $\mathbf{S}$-instances I and J such that $\mathrm{facts}(I) = \mathrm{facts}(J)$. We show that $I \in O$ implies $J \in O$; the fact that $J \in O$ implies $I \in O$ is shown analogously. This is done by showing that O is (n, m)-locally embeddable in J, which in turn implies that $J \in O$ since, by hypothesis, O is (n, m)-local. Consider an $\mathbf{S}$-instance $K \preceq J$ with $|\mathrm{adom}(K)| \leq n$. Since $\mathrm{facts}(I) = \mathrm{facts}(J)$, it is clear that $K \subseteq I$. Since $I \in O$, it suffices to show that, for every J' in the m-neighbourhood of K in I, there is a function $h_{J'} : \mathrm{adom}(J') \to \mathrm{adom}(J)$, which is the identity on $\mathrm{adom}(K)$, such that $h_{J'}(\mathrm{facts}(J')) \subseteq \mathrm{facts}(J)$. Since $J' \subseteq I$, which means that $\mathrm{facts}(J') \subseteq \mathrm{facts}(I)$, we get that $\mathrm{facts}(J') \subseteq \mathrm{facts}(J)$. Therefore, $h_{J'}$ is simply the identity on $\mathrm{adom}(J')$. □

4 CHARACTERIZING TGD-ONTOLOGIES

Are the three main properties presented above, i.e., criticality, ⊗-closure, and locality, enough to characterize when an ontology is a TGD-ontology? We proceed to give a positive answer. Actually, we provide a more refined result in the sense that we can characterize when an ontology is a $\mathsf{TGD}_{n,m}$-ontology. Unsurprisingly, (n, m)-locality is the key property for such a refined characterization.

Theorem 4.1. *Given an ontology O, and integers $n, m \geq 0$, the following are equivalent:*

(1) O is a $\mathsf{TGD}_{n,m}$-ontology.
(2) O is critical, closed under direct products, and (n, m)-local.

From the above result, which is interesting in its own right, we obtain a characterization of when an ontology is a TGD-ontology.

Corollary 4.2. *Given an ontology O, the following are equivalent:*

(1) O is a TGD*-ontology.*
(2) O is critical, closed under direct products, and local.

The rest of this section is devoted to discussing the proof of Theorem 4.1. Actually, we focus on the non-trivial direction (2) ⇒ (1); the direction (1) ⇒ (2) follows from Lemmas 3.2, 3.4 and 3.6.

4.1 Some Preparation

We first need to introduce some auxiliary technical notions.

Existential Disjunctive Dependencies. In our proof, we are going to use existential disjunctive dependencies, which essentially generalize tgds with equality and disjunction in the head. More precisely, an *existential disjunctive dependency* (edd) δ over a schema $\mathbf{S}$ is a constant-free first-order sentence $\forall \bar{x}(\phi(\bar{x}) \to \bigvee_{i=1}^{k} \psi_i(\bar{x}_i))$, where $\bar{x}$ is a tuple of variables of $\mathbf{V}$, $\phi(\bar{x})$ is a (possibly empty) conjunction of atoms over $\mathbf{S}$, and, for each $i \in [k]$, $\bar{x}_i \subseteq \bar{x}$, and $\psi(\bar{x}_i)$ is either an equality expression $y = z$ with $\bar{x}_i = \{y, z\}$, or a constant-free formula $\exists \bar{y}_i \chi_i(\bar{x}_i, \bar{y}_i)$ with $\bar{y}_i$ being a tuple of variables from $\mathbf{V} \setminus \bar{x}$, and $\chi_i(\bar{x}_i, \bar{y}_i)$ a (non-empty) conjunction of atoms over $\mathbf{S}$. If $k = 1$ and $\psi_1(\bar{x}_1)$ is an equality expression, then δ is called an *equality-generating dependency* (egd). An instance I satisfies the edd δ, written $I \models \delta$, if, whenever there exists a function $h : \bar{x} \to \mathrm{dom}(I)$ such that $h(\phi(\bar{x})) \subseteq \mathrm{facts}(I)$, then there is $i \in [k]$ such that, if $\psi_i(\bar{y}_i)$ is $y = z$, then $h(y) = h(z)$; otherwise, if $\psi_i(\bar{y}_i)$ is of the form $\exists \bar{y}_i \chi_i(\bar{x}_i, \bar{y}_i)$, then there is an extension h' of h such that $h'(\chi_i(\bar{x}_i, \bar{y}_i)) \subseteq \mathrm{facts}(I)$. The instance I satisfies a set Σ of edds, written $I \models \Sigma$, i.e., I is a *model* of Σ, if $I \models \delta$ for each $\delta \in \Sigma$.

Relative Diagram of an Instance. We now proceed to introduce the diagram of an instance relative to another instance. This can be seen as a refinement of the standard notion of diagram of a relational structure in model theory (see, e.g., [8]). Consider two $\mathbf{S}$-instances I, K such that $K \preceq I$, and an integer $\ell \geq 0$. We are interested in the so-called ℓ-diagram of K relative to I, which we define now. Let $A_{K,\ell}$ be the set of all atomic formulas of the form $R(\bar{u})$ that can be formed using predicates from $\mathbf{S}$, constants from $\mathrm{dom}(K)$, and ℓ distinct variables $\star_1, \ldots, \star_\ell$ from $\mathbf{V}$, i.e., $R \in \mathbf{S}$ and $\bar{u} \in (\mathrm{dom}(K) \cup \{\star_1, \ldots, \star_\ell\})^{\mathrm{ar}(R)}$. Let $C_{K,\ell}$ be the set of all (possibly infinite) conjunctions of atomic formulas from $A_{K,\ell}$. Given a formula $\gamma(\bar{y}) \in C_{K,\ell}$, we can naturally talk about the satisfaction of the sentence $\exists \bar{y}\, \gamma(\bar{y})$ by an instance J, in which case we simply write $J \models \exists \bar{y}\, \gamma(\bar{y})$. The *$\ell$-diagram of K relative to I*, denoted $\Delta_{K,\ell}^{I}$, is the (possibly infinite) first-order formula

$$\bigwedge_{\alpha \in \mathrm{facts}(K)} \alpha \;\wedge \bigwedge_{\substack{c,d \in \mathrm{dom}(K),\\ c \neq d}} \neg(c = d) \;\wedge \bigwedge_{\substack{\gamma(\bar{y}) \in C_{K,\ell},\\ I \not\models \exists \bar{y}\, \gamma(\bar{y})}} \neg \exists \bar{y}\, \gamma(\bar{y}).$$

We are, actually, interested in the first-order formula $\Phi_{K,\ell}^{I}(\bar{x})$ obtained from $\Delta_{K,\ell}^{I}$ by replacing each constant element $c \in \mathrm{dom}(K)$ with a new variable $x_c \in \mathbf{V} \setminus \{\star_1, \ldots, \star_\ell\}$. As for the formulas of $C_{K,\ell}$, we can naturally talk about the satisfaction of $\exists \bar{x}\, \Phi_{K,\ell}^{I}(\bar{x})$ by an instance J, in which case we simply write $J \models \exists \bar{x}\, \Phi_{K,\ell}^{I}(\bar{x})$. It is straightforward to verify the following easy lemma:

Lemma 4.3. *Consider an $\mathbf{S}$-instance I. For each $\mathbf{S}$-instance $K \preceq I$ and integer $\ell \geq 0$, it holds that $I \models \exists \bar{x}\, \Phi_{K,\ell}^{I}(\bar{x})$.*

4.2 The Proof of (2) ⇒ (1)

We now have all the ingredients needed for giving the proof of the direction (2) ⇒ (1) in Theorem 4.1. Assume that the ontology O is over a schema $\mathbf{S}$. The proof proceeds in three main steps:

(1) We construct a finite set $\Sigma^{\vee}$ of edds over $\mathbf{S}$ that mention at most n universally, and at most m existentially quantified

variables, in such a way that O is precisely the set of models of $\Sigma^{\vee}$. This exploits the fact that O is 1-critical (since it is critical), and the fact that O is (n, m)-local.

(2) We then show that there exists a finite set $\Sigma^{\exists,=}$ of tgds and egds over S that is logically equivalent to $\Sigma^{\vee}$, i.e., $\Sigma^{\exists,=}$ and $\Sigma^{\vee}$ have exactly the same models, written $\Sigma^{\exists,=} \equiv \Sigma^{\vee}$; in fact, $\Sigma^{\exists,=}$ is simply the set of tgds and egds in $\Sigma^{\vee}$. This exploits the fact that O is closed under direct products.

(3) We finally show that there exists a finite set $\Sigma^{\exists}$ of tgds over S from $\mathsf{TGD}_{n,m}$ such that $\Sigma^{\exists} \equiv \Sigma^{\exists,=}$; in fact, $\Sigma^{\exists}$ consists of the tgds of $\Sigma^{\exists,=}$. This exploits the fact that O is critical.

We proceed to give further details on each of the above steps.

Step 1: The finite set of edds $\Sigma^{\vee}$

Let $\mathsf{E}_{n,m}$ be the set that collects all the edds over S of the form $\forall \bar{x}(\phi(\bar{x}) \rightarrow \bigvee_{i=1}^{k} \psi_i(\bar{x}_i))$ such that $\bar{x}$ consists of at most n distinct variables, and, for each $i \in [k]$, $\psi_i(\bar{x}_i)$ mentions at most $n + m$ distinct variables. The latter means that, if $\psi_i(\bar{x}_i)$ is a formula of the form $\exists \bar{y}_i\, \chi_i(\bar{x}_i, \bar{y}_i)$ (i.e., is not an equality expression), then $\bar{y}_i$ consists of at most m distinct variables. It is important to observe that $\mathsf{E}_{n,m}$ is finite (up to logical equivalence); this is a consequence of the fact that S is finite, and the number of variables in each edd of $\mathsf{E}_{n,m}$ is finite. We then define the set $\Sigma^{\vee}$ of edds as the set of all edds from $\mathsf{E}_{n,m}$ that are satisfied by every instance of O, i.e.,

$$\Sigma^{\vee} \;=\; \left\{\delta \in \mathsf{E}_{n,m} \mid \text{ for each } I \in O, \text{ it holds that } I \models \delta\right\}.$$

It is clear that $\Sigma^{\vee}$ is finite (up to logical equivalence) since $\Sigma^{\vee} \subseteq \mathsf{E}_{n,m}$. We show that O is precisely the set of models of $\Sigma^{\vee}$.

Lemma 4.4. *For each* S*-instance I, $I \in O$ iff $I \models \Sigma^{\vee}$.*

The ($\Rightarrow$) direction of Lemma 4.4 holds by construction. We proceed to discuss the non-trivial direction ($\Leftarrow$). Consider an S-instance I such that $I \notin O$. By Lemma 3.8, O is domain independent, which allows us to assume that $\mathrm{dom}(I) = \mathrm{adom}(I)$, i.e., all the constants in $\mathrm{dom}(I)$ occur in at least one fact of I. The goal is to show that $I \not\models \Sigma^{\vee}$, which in turn establishes Lemma 4.4. We first show the following technical claim that involves $\Phi^{I}_{K,m}(\bar{x})$; recall that $\Phi^{I}_{K,m}(\bar{x})$ is the formula obtained from the m-diagram of K relative to I, which here is finite since the domain of K is finite.

Claim 4.5. *There exists an* S*-instance $K \preceq I$ with $|\mathrm{adom}(K)| \leq n$ such that, for each $J \in O$, it holds that $J \models \neg\exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$.*

Proof. By contradiction, assume that, for each S-instance $K \preceq I$ with $|\mathrm{adom}(K)| \leq n$, there exists $J \in O$ such that $J \models \exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$. We proceed to show that the ontology O is (n, m)-locally embeddable in I, which in turn implies that $I \in O$ since, by hypothesis, O is (n, m)-local. But this contradicts the fact that $I \notin O$.

Fix an arbitrary S-instance $K \preceq I$ with $|\mathrm{adom}(K)| \leq n$, and let $J \in O$ be the instance such that $J \models \exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$. We first observe that $J \models \exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$ implies the existence of an instance $J_K \subseteq J$ such that $K \simeq J_K$. We can therefore assume, w.l.o.g., that $K \subseteq J$. To show that O is (n, m)-locally embeddable in I, it suffices to show that for every S-instance J' in the m-neighbourhood of K in J, there exists a function $h : \mathrm{adom}(J') \rightarrow \mathrm{adom}(I)$, which is the identity on $\mathrm{adom}(K)$, such that $h(\mathrm{facts}(J')) \subseteq \mathrm{facts}(I)$.

By contradiction, assume that there is J' in the m-neighbourhood of K in J for which there is no function $h : \mathrm{adom}(J') \rightarrow \mathrm{adom}(I)$ that is the identity on $\mathrm{adom}(K)$ with $h(\mathrm{facts}(J')) \subseteq \mathrm{facts}(I)$. Let L be the S-instance defined as the difference between J' and K, i.e., L is such that $\mathrm{facts}(L) = \mathrm{facts}(J') \setminus \mathrm{facts}(K)$, while $\mathrm{dom}(L)$ consists of all the constants occurring in $\mathrm{facts}(J') \setminus \mathrm{facts}(K)$, i.e., $\mathrm{dom}(L) = \mathrm{adom}(L)$. Clearly, there is no function $h : \mathrm{adom}(L) \rightarrow \mathrm{adom}(I)$ that is the identity on $\mathrm{adom}(K)$ such that $h(\mathrm{facts}(L) \subseteq I$. Observe that $|\mathrm{adom}(L) \setminus \mathrm{adom}(K)| \leq m$; we assume that $\mathrm{adom}(L) \setminus \mathrm{adom}(K) = \{d_1, \ldots, d_{m'}\}$ for $m' \leq m$. Let $\gamma(\bar{y})$ be the formula obtained from $\bigwedge_{\alpha \in \mathrm{facts}(L)} \alpha$ after renaming each constant d_i to the variable $\star_i$; clearly, $\bar{y} = \star_1, \ldots, \star_{m'}$. Since there is no function $h : \mathrm{adom}(L) \rightarrow \mathrm{adom}(I)$ that is the identity on $\mathrm{adom}(K)$ such that $h(\mathrm{facts}(L) \subseteq I$, we can conclude that $I \not\models \exists \bar{y}\, \gamma(\bar{y})$. Observe now that, by construction, $\neg\exists \bar{y}\, \gamma(\bar{y})$ is a conjunct of $\Delta^{I}_{K,m}$, which in turn implies that the formula $\neg\exists \bar{z}\, \gamma(\bar{z})$ obtained from $\neg\exists \bar{y}\, \gamma(\bar{y})$ after renaming each constant $c \in \mathrm{dom}(K) = \mathrm{adom}(K)$ to the variable x_c is a conjunct of $\Phi^{I}_{K,m}(\bar{x})$. Since $L \subseteq J$, we conclude that $J \models \exists \bar{z}\, \gamma(\bar{z})$, which in turn implies that $J \not\models \exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$. But this contradicts the fact that $J \models \exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$, and the claim follows. □

Consider now the S-instance K provided by Claim 4.5. We can show that $\neg\exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$ is logically equivalent to an edd from $\mathsf{E}_{n,m}$.

Claim 4.6. *There is an edd $\delta \in \mathsf{E}_{n,m}$ such that $\delta \equiv \neg\exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$.*

Actually, the edd δ claimed above is obtained by simply converting $\neg\exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$ into an equivalent edd via the standard logical transformations; the formal construction can be found in the Appendix. However, we need to argue that, after this transformation, we indeed obtain an edd from $\mathsf{E}_{n,m}$, namely (i) the right-hand side is non-empty, (ii) each variable in the right-hand side is either existentially quantified, or appears in the left-hand side, and (iii) there are at most n universally and at most m existentially quantified variables. Item (i) holds since $\Phi^{I}_{K,m}(\bar{x})$ contains at least one negative conjunct; otherwise, the 1-critical instance in O (it exists since O is 1-critical) is a model of $\exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$, which contradicts Claim 4.5. Item (ii) is guaranteed by the fact that $\mathrm{dom}(K) = \mathrm{adom}(K)$ since $\mathrm{dom}(I) = \mathrm{adom}(I)$; recall that the latter relies on the fact that O is domain independence. Finally, item (iii) is ensured by the fact that $|\mathrm{dom}(K)| = |\mathrm{adom}(K)| \leq n$.

Having the above technical claims in place, we can now show that $I \not\models \Sigma^{\vee}$, which in turn completes the proof of Lemma 4.4. By Claims 4.5 and 4.6, we get that there exists an S-instance $K \preceq I$ with $|\mathrm{adom}(K)| \leq n$ such that (i) for every $J \in O$, $J \models \neg\exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$, and (ii) $\neg\exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$ is equivalent to an edd $\delta \in \mathsf{E}_{n,m}$. By the definition of $\Sigma^{\vee}$, we conclude that $\delta \in \Sigma^{\vee}$. On the other hand, by Lemma 4.3, $I \models \exists \bar{x}\, \Phi^{I}_{K,m}(\bar{x})$, i.e., $I \not\models \delta$, and thus $I \not\models \Sigma^{\vee}$, as needed.

Step 2: The set of tgds and egds $\Sigma^{\exists,=}$ such that $\Sigma^{\exists,=} \equiv \Sigma^{\vee}$

We define the set $\Sigma^{\exists,=}$ as the set of all tgds and egds in $\Sigma^{\vee}$, that is,

$$\Sigma^{\exists,=} \;=\; \left\{\delta \in \Sigma^{\vee} \mid \delta \text{ is either a tgd or an egd}\right\}.$$

We show that this is the desired set of tgds and egds

Lemma 4.7. *It holds that $\Sigma^{\exists,=} \equiv \Sigma^{\vee}$.*

It is clear that $\Sigma^{\vee} \models \Sigma^{\exists,=}$, that is, each model of $\Sigma^{\vee}$ is a model of $\Sigma^{\exists,=}$, since $\Sigma^{\exists,=}$ is a subset of $\Sigma^{\vee}$. We now discuss the non-trivial direction $\Sigma^{\exists,=} \models \Sigma^{\vee}$. Towards a contradiction, assume that $\Sigma^{\exists,=} \not\models \Sigma^{\vee}$. This implies that there is an edd $\delta \in \Sigma^{\vee}$ of the form $\forall \bar{x}(\phi(\bar{x}) \rightarrow \bigvee_{i=1}^{k} \psi_i(\bar{x}_i))$ such that $\Sigma^{\exists,=} \not\models \delta$. It is clear that, for each $j \in [k]$, $\sigma_j = \forall \bar{x}(\phi(\bar{x}) \rightarrow \psi_j(\bar{x}_j))$ does not belong to $\Sigma^{\exists,=}$; otherwise, $\Sigma^{\exists,=} \models \delta$ which is not the case. Therefore, by the definition of $\Sigma^{\vee}$, for each $j \in [k]$, there exists an S-instance $I_j \in O$ such that $I_j \not\models \sigma_j$, or, equivalently, $I_j \models \exists \bar{x}(\phi(\bar{x}) \wedge \neg\psi_j(\bar{x}_j))$. We define the S-instance

$$J \;=\; I_1 \otimes \cdots \otimes I_k.$$

Since, by hypothesis, O is closed under direct products, we get that $J \in O$. We can also show the following technical claim concerning the instance J; the proof is deferred to the Appendix.

Claim 4.8. *It holds that $J \not\models \delta$.*

By Lemma 4.4, $J \models \Sigma^{\vee}$, and thus $J \models \delta$ since $\delta \in \Sigma^{\vee}$. But this contradicts Claim 4.8. Consequently, $\Sigma^{\exists,=} \models \Sigma^{\vee}$, as needed.

Step 3: The set of tgds $\Sigma^{\exists}$ from $\mathsf{TGD}_{n,m}$ such that $\Sigma^{\exists} \equiv \Sigma^{\exists,=}$

We define the set $\Sigma^{\exists}$ as the set of all tgds in $\Sigma^{\exists,=}$, that is,

$$\Sigma^{\exists} \;=\; \left\{\delta \in \Sigma^{\exists,=} \mid \delta \text{ is a tgd}\right\}.$$

We show that this is the desired set of tgds

Lemma 4.9. *It holds that $\Sigma^{\exists} \in \mathsf{TGD}_{n,m}$ and $\Sigma^{\exists} \equiv \Sigma^{\exists,=}$.*

The fact that $\Sigma \in \mathsf{TGD}_{n,m}$ follows by construction since each edd of $\Sigma^{\exists}$ belongs to $\mathsf{E}_{n,m}$. It is also easy to see that $\Sigma^{\exists,=} \models \Sigma^{\exists}$ since $\Sigma^{\exists} \subseteq \Sigma^{\exists,=}$. For showing that $\Sigma^{\exists} \models \Sigma^{\exists,=}$, assume by contradiction that there is an egd $\delta \in \Sigma^{\exists,=}$ of the form $\forall \bar{x}(\phi(\bar{x}) \rightarrow y = z)$ such that $\Sigma^{\exists} \not\models \delta$. This implies that there exists an S-instance I such that $I \not\models \delta$, i.e., there is a function $h : \bar{x} \rightarrow \mathsf{dom}(I)$ such that $h(\phi(\bar{x})) \subseteq \mathsf{facts}(I)$ but $h(y) \neq h(z)$. Observe that $h(\phi(\bar{x})) \not\models \delta$. Let J be a k-critical instance, where k is the number of distinct variables in $\bar{x}$, such that $h(\phi(\bar{x})) \subseteq \mathsf{facts}(J)$. It is clear that $J \not\models \delta$. Since O is critical, $J \in O$. But this contradicts the fact that $\delta \in \Sigma^{\exists,=} \subseteq \Sigma^{\vee}$, which means that δ is satisfied by every instance in O.

It should be now clear that the non-trivial direction (2) $\Rightarrow$ (1) of Theorem 4.1 readily follows from Lemmas 4.4, 4.7 and 4.9.

5 CHARACTERIZING FTGD-ONTOLOGIES

The next natural question is whether we can characterize when an ontology can be expressed as a finite set of existential-free tgds, i.e., when an ontology is an FTGD-ontology. Since FTGD is the class $\bigcup_{n>0} \mathsf{TGD}_{n,0}$, an answer to this question can be obtained from the characterization established in the previous section, which exemplifies the usefulness of our new locality property. In particular, the following is an immediate consequence of Theorem 4.1.

Corollary 5.1. *Given an ontology O, the following are equivalent:*

(1) O is an FTGD-ontology.
(2) O is critical, closed under direct products, and $(n, 0)$-local for some integer $n > 0$.

Observe also that Theorem 4.1 provides a characterization of when an ontology can be expressed as a finite set of full tgds with at most $n > 0$ universally quantified variables. In particular, given an ontology O and integer $n > 0$, O is an $\mathsf{FTGD}_{n,0}$-ontology iff O is critical, is closed under direct products, and is $(n, 0)$-local.

5.1 An Alternative Characterization

At this point, it is worth noting that the question of whether we can provide a characterization for full tgds has been already considered back in the 1980s by Makowsky and Vardi in a slightly different context [14]. As described in the Introduction, the main goal of [14] was to characterize the expressive power of database dependencies in terms of model-theoretic properties. Among others, they studied the question of when a family of databases (i.e., instances with a *finite* domain) can be specified as a (finite or infinite) set of full tgds; note that they did not obtain results about tgds with existentially quantified variables. In principle, the characterization of [14] for full tgds can be lifted to our setting, where we consider finite sets of tgds and unrestricted instances. Unfortunately, the model-theoretic characterization of full tgds in Theorem 3 of [14] turns out to be incorrect since one of the closure properties used, called *closure under duplicating extensions*, does not serve its purpose. In fact, Lemma 7 of [14] states that tgds are closed under duplicating extensions, which, however, is not the case, as we discuss next.

In [14], an instance J is a *duplicating extension* of an instance I if there are constants $c \in \mathsf{dom}(I)$ and $d \notin \mathsf{dom}(I)$ such that

$$\mathsf{dom}(J) = \mathsf{dom}(I) \cup \{d\} \quad \text{and} \quad \mathsf{facts}(J) = \mathsf{facts}(I) \cup h(\mathsf{facts}(I))$$

with $h : \mathsf{dom}(I) \rightarrow \mathsf{dom}(I) \cup \{d\}$ being the identity on $\mathsf{dom}(I) \setminus \{c\}$ and $h(c) = d$.[3] Intuitively speaking, J is obtained from I by adding d to $\mathsf{dom}(I)$, and by adding to $\mathsf{facts}(I)$ a copy of itself after renaming c to d. An ontology O over S is *closed under duplicating extensions* if, for every $I \in O$, and S-instance J that is a duplicating extension of I, we have that $J \in O$. Lemma 7 of [14] states that tgds are closed under duplicating extensions, and this has been used to prove the characterization of full tgds in Theorem 3 of [14]. However, we can show that there is an FTGD-ontology *not* closed under duplicating extensions, and thus, Lemma 7 in [14] is incorrect.

Example 5.2 (**Counterexample**). Consider the full tgd

$$\sigma \;=\; R(x,y), S(y,z) \rightarrow T(x,z).$$

It is clear that the instance I with

$$\mathsf{dom}(I) = \{a, b\} \quad \text{and} \quad \mathsf{facts}(I) = \{R(a,b), S(b,a), T(a,a)\}$$

satisfies σ. It is easy now to verify that the instance J with

$$\mathsf{dom}(J) = \mathsf{dom}(I) \cup \{c\} \quad \text{and} \quad \mathsf{facts}(J) = \mathsf{facts}(I) \cup \{R(c,b), S(b,c), T(c,c)\}$$

is a duplicating extension of I due to the function $h : \mathsf{dom}(I) \rightarrow \mathsf{dom}(I) \cup \{c\}$ with $h(a) = c$ and $h(b) = b$. However, $J \not\models \sigma$; indeed, there is a function $h : \{x, y, z\} \rightarrow \mathsf{dom}(J)$ such that $h(\mathsf{body}(\sigma)) = \{R(a,b), S(b,c)\}$, but $h(T(x,z)) = T(a,c) \notin \mathsf{facts}(J)$. ■

In what follows in this section, we obtain an alternative model-theoretic characterization of full tgds that uses a different version of closure under duplicating extensions. The problem with the definition of closure under duplicating extensions in [14] is that it does not distinguish the different occurrences of the special constant, which is mapped to the new constant, appearing in a single fact.

[3] Note that the definition in [14] is for databases, but it can be transferred to instances.

Going back to our counterexample, since the different occurrences of the constant a in $T(a,a)$ are not distinguished, there is no way to obtain an atom of the form $T(a,c)$ in the duplicating extension. A valid duplicating extension of I would be the instance J with

$$\mathsf{dom}(J) = \mathsf{dom}(I) \cup \{c\} \quad \text{and}$$
$$\mathsf{facts}(J) = \mathsf{facts}(I) \cup \{R(c,b), S(b,c), T(a,c), T(c,a), T(c,c)\}.$$

This is achieved by the following refined definition. Consider two S-instances I and J. We say that J is a *non-oblivious duplicating extension of* I if there are constants $c \in \mathsf{dom}(I)$ and $d \notin \mathsf{dom}(I)$ such that, for every $R \in \mathbf{S}$ and tuple $\bar{t} \in (\mathsf{dom}(I) \cup \{d\})^{\mathsf{ar}(R)}$, $R(\bar{t}) \in J$ iff $h(R(\bar{t})) \in I$ with $h : \mathsf{dom}(I) \cup \{d\} \rightarrow \mathsf{dom}(I)$ being the identity on $\mathsf{dom}(I)$ and $h(d) = c$. The term non-oblivious refers to the fact that now the definition is not oblivious to the different occurrences of the constant c. The desired closure property follows.

Definition 5.3 (**Duplicating Extensions Closure**). An ontology O over S is *closed under non-oblivious duplicating extensions* if, for every $I \in O$, and S-instance J that is a non-oblivious duplicating extension of I, it holds that $J \in O$. ■

Our alternative characterization of FTGD-ontologies relies on closure under non-oblivious duplicating extensions, 1-criticality, domain independence, and two properties that we have not considered before, namely modularity and closure under intersections. Let us introduce those two properties; let $\mathbf{S} = \{R_1, \ldots, R_\ell\}$.

Modularity. We start with n-modularity, for $n \geq 0$, which actually provides a small witness instance with at most n domain elements of why an instance does not belong to an ontology. Let us clarify that in [14] there is an analogous notion for classes of databases called n-locality. However, here we adopt the name n-modularity, which has been already used in [16] for a similar notion in the context of schema mappings, in order to avoid any confusion with the new notion of (n,m)-locality introduced in this work. Indeed, (n,m)-locality (even when $m = 0$) is inherently different than the notion of n-modularity, or the notion of n-locality from [14].

Definition 5.4 (**Modularity**). An ontology O over S is *n-modular*, for some integer $n \geq 0$, if, for every S-instance $I \notin O$, there exists an S-instance $J \preceq I$ with $|\mathsf{dom}(J)| \leq n$ such that $J \notin I$. ■

Closure Under Intersections. We now recall closure under intersections. Given two S-instances $I = (\mathsf{dom}(I), R_1^I, \ldots, R_\ell^I)$ and $J = (\mathsf{dom}(J), R_1^J, \ldots, R_\ell^J)$, the *intersection of* I *and* J, denoted $I \cap J$, is the instance $(\mathsf{dom}(I) \cap \mathsf{dom}(J), R_1^I \cap R_1^J, \ldots, R_\ell^I \cap R_\ell^J)$. Then:

Definition 5.5 (**$\cap$-closure**). An ontology O over S is *closed under intersections* if, for every two S-instances $I, J \in O$, $I \cap J \in O$. ■

The Characterization. It is now possible to obtain the alternative characterization of when an ontology is an FTGD-ontology by using the notion of non-oblivious duplicating extension. The proof, which is similar in spirit to that of Theorem 4.1, is in the Appendix.

Theorem 5.6. *Given an ontology O, the following are equivalent:*

(1) O is an FTGD*-ontology.*
(2) O is 1*-critical, domain independent, n-modular for some integer* $n \geq 0$*, closed under intersections, and closed under non-oblivious duplicating extensions.*

Let us conclude this section by clarifying that the problematic Theorem 3 in [14] does not use the property of modularity since it follows the axiomatizability approach, where infinite sets of full tgds are allowed. However, here we adopt the finite axiomatizability approach, which means that we are interested only in finite sets of full tgds, and this is the reason why modularity is needed.

6 CHARACTERIZING LTGD-ONTOLOGIES

We now focus our attention on linear tgds, and ask whether we can characterize when an ontology is an LTGD-ontology. In particular, our objective is to obtain characterizations for linear tgds in the spirit of Theorem 4.1 and Corollary 4.2, based on our new locality property. Interestingly, this can be done by replacing locality with a refined version of it, which we call linear locality, whereas criticality and closure under direct products remain untouched.

6.1 Linear Locality

We first refine the notion local embedding by taking into account the fact that linear tgds have at most one body atom. As one might expect, now an ontology O over S is locally embeddable in an S-instance I if, for every $K \subseteq I$ with at most one fact (instead of every subinstance of I) that mentions a bounded number of constants, we can find an instance $J_K \in O$ such that every local neighbour of K in J_K can be embedded in I while preserving K. We proceed to formalize this intuitive description.

Consider an ontology O over a schema S, and an S-instance I. For $n,m \geq 0$, we say that O *is linearly* (n,m)*-locally embeddable in* I if, for every S-instance $K \subseteq I$ such that $|\mathsf{facts}(K)| \leq 1$ and $|\mathsf{adom}(K)| \leq n$, there exists $J_K \in O$ such that $K \subseteq J_K$, and for every J' in the m-neighbourhood of K in J_K, there is a function $h : \mathsf{adom}(J') \rightarrow \mathsf{adom}(I)$, which is the identity on $\mathsf{adom}(K)$, such that $h(\mathsf{facts}(J')) \subseteq \mathsf{facts}(I)$. The property of linear locality follows.

Definition 6.1 (**Linear Locality**). An ontology O over S is *linear* (n,m)*-local*, for $n,m \geq 0$, if, for every S-instance I, O is linearly (n,m)-locally embeddable in I implies $I \in O$. We also say that O is *linear local* if there are $n,m \geq 0$ such that O is linear (n,m)-local. ■

It is important to observe that linear locality implies locality as this will be crucial for obtaining our main characterization.

Lemma 6.2. *Consider an ontology O that is linear* (n,m)*-local for some* $n,m \geq 0$*. It holds that O is* (n,m)*-local.*

Proof. Assume that O is over the schema S. Consider an arbitrary S-instance I such that O is (n,m)-locally embeddable in I. Therefore, by definition, O is linearly (n,m)-locally embeddable in I. Since, by hypothesis, O is linear (n,m)-local, we conclude that $I \in O$, which in turn implies that O is (n,m)-local, as needed. □

6.2 The Characterization

We proceed to show that indeed linear locality is the right notion for obtaining our main characterization. To this end, we first present a technical lemma, dubbed *Linearization Lemma*, that essentially characterizes when a TGD-ontology can be expressed as a finite set of linear tgds; the proof is deferred to the Appendix.

Lemma 6.3 (Linearization). *Given a* $\mathsf{TGD}_{n,m}$*-ontology O, for some* $n,m \geq 0$*, the following are equivalent:*

(1) O is an LTGD*-ontology.*
(2) O is an $\text{LTGD}_{n,m}$*-ontology.*
(3) O is linear (n, m)*-local.*

Let us stress that the Linearization Lemma, apart from characterizing when a $\text{TGD}_{n,m}$-ontology is an $\text{LTGD}_{n,m}$-ontology via linear (n, m)-locality, it also tells us the following: if a finite set Σ of tgds can be equivalently rewritten as a finite set Σ_L of linear tgds, then each tgd of Σ_L does not have to use more universally or existentially quantified variables than the tgds of Σ (direction (1) $\Rightarrow$ (2)). This is an interesting fact that, although is not essential for our main characterization concerning linear tgds, it will serve as the basis of the procedure for checking whether a finite set of guarded tgds can be rewritten as a finite set of linear tgds (see Section 9).

Having the Linearization Lemma in place, it is now not difficult to obtain the desired characterizations for linear tgds.

Theorem 6.4. *Given an ontology O, and integers* $n, m \geq 0$*, the following are equivalent:*

(1) O is an $\text{LTGD}_{n,m}$*-ontology.*
(2) O is critical, closed under direct prod., and linear (n, m)*-local.*

Proof. The fact that (1) $\Rightarrow$ (2) follows from Lemmas 3.2 and 3.4, and the direction (2) $\Rightarrow$ (3) of Lemma 6.3. For (2) $\Rightarrow$ (1), since O is linear (n, m)-local, Lemma 6.2 implies that O is (n, m)-local. Since, by hypothesis, O is also critical and closed under direct products, we get from Theorem 4.1 that O is a $\text{TGD}_{n,m}$-ontology, which in turn allows us to apply the Linearization Lemma (direction (3) $\Rightarrow$ (2)), and get that O is an $\text{LTGD}_{n,m}$-ontology, as needed. □

We conclude this section by observing that Theorem 6.4 provides a characterization of when an ontology is an LTGD-ontology.

Corollary 6.5. *Given an ontology O, the following are equivalent:*

(1) O is an LTGD*-ontology.*
(2) O is critical, is closed under direct products, and is linear local.

7 CHARACTERIZING GTGD-ONTOLOGIES

Let us now proceed with guarded tgds, and perform a similar analysis as in the previous section for linear tgds. Our objective here is to obtain characterizations for guarded tgds in the spirit of Theorem 4.1 and Corollary 4.2, by relying on a refined version of our locality property, which we call guarded locality, whereas criticality and closure under direct products remain in place.

7.1 Guarded Locality

We first refine the notion of local embedding by taking into account the fact that guarded tgds have either an empty body, or a body atom that contains all the universally quantified variables. As one might guess, now an ontology O over $\mathbf{S}$ is locally embeddable in an $\mathbf{S}$-instance I if, for every guarded subinstance K of I (i.e., K has a fact that contains all the active domain elements), that mentions a bounded number of constants, we can find an instance $J_K \in O$ such that every local neighbour of K in J_K can be embedded in I while preserving the instance K. The formal definition follows.

An instance I is called *guarded* if either $\text{facts}(I) = \emptyset$, or there exists a fact $R(c_1, \ldots, c_k) \in \text{facts}(I)$ such that $\text{adom}(I) = \{c_1, \ldots, c_k\}$. Consider now an ontology O over a schema $\mathbf{S}$, and an $\mathbf{S}$-instance I. For $n, m \geq 0$, we say that O *is guardedly* (n, m)*-locally embeddable in* I if, for every guarded $\mathbf{S}$-instance $K \leq I$ with $|\text{adom}(K)| \leq n$, there is $J_K \in O$ such that $K \subseteq J_K$, and for every J' in the m-neighbourhood of K in J_K, there exists a function $h : \text{adom}(J') \to \text{adom}(I)$, which is the identity on $\text{adom}(K)$, such that $h(\text{facts}(J')) \subseteq \text{facts}(I)$. The property of guarded locality is defined as expected.

Definition 7.1 (**Guarded Locality**). An ontology O over $\mathbf{S}$ is *guarded* (n, m)*-local*, for $n, m \geq 0$, if, for every $\mathbf{S}$-instance I, the following holds: O is guardedly (n, m)-locally embeddable in I implies $I \in O$. We also say that O is *guarded local* if there are $n, m \geq 0$ such that O is guarded (n, m)-local. ■

As for linear locality, we can show that guarded locality implies locality; the proof, which is omitted, is similar to that of Lemma 6.2

Lemma 7.2. *Consider an ontology O that is guarded* (n, m)*-local for some* $n, m \geq 0$*. It holds that O is* (n, m)*-local.*

7.2 The Characterization

We now show that guarded locality is the right notion for obtaining our main characterization for guarded tgds. To this end, we first present a technical lemma in the spirit of the Linearization Lemma, called *Guardedization Lemma*; the proof is deferred to the Appendix.

Lemma 7.3 (Guardedization). *Given a* $\text{TGD}_{n,m}$*-ontology O, for some* $n, m \geq 0$*, the following are equivalent:*

(1) O is a GTGD*-ontology.*
(2) O is a $\text{GTGD}_{n,m}$*-ontology.*
(3) O is guarded (n, m)*-local.*

Let us note that, similarly to the Linearization Lemma, the Guardedization Lemma, apart from characterizing when a $\text{TGD}_{n,m}$-ontology is a $\text{GTGD}_{n,m}$-ontology via guarded (n, m)-locality, it also tells that, if a finite set Σ of tgds can be equivalently rewritten as a finite set Σ_G of guarded tgds, then each tgd of Σ_G does not have to use more universally or existentially quantified variables than the tgds of Σ ((1) $\Rightarrow$ (2)). This interesting fact will be crucial for the procedure that checks whether a finite set of frontier-guarded tgds can be rewritten as a finite set of guarded tgds (see Section 9).

Having the Guardedization Lemma in place, it is now easy to obtain the desired characterizations for guarded tgds.

Theorem 7.4. *Given an ontology O, and integers* $n, m \geq 0$*, the following are equivalent:*

(1) O is a $\text{GTGD}_{n,m}$*-ontology.*
(2) O is critical, closed under direct prod., and guarded (n, m)*-local.*

Proof. The fact that (1) $\Rightarrow$ (2) follows from Lemmas 3.2 and 3.4, and the direction (2) $\Rightarrow$ (3) of Lemma 7.3. For (2) $\Rightarrow$ (1), since O is guarded (n, m)-local, Lemma 7.2 implies that O is (n, m)-local. Since, by hypothesis, O is also critical and closed under direct products, we get from Theorem 4.1 that O is a $\text{TGD}_{n,m}$-ontology, which in turn allows us to apply the Guardedization Lemma (direction (3) $\Rightarrow$ (2)), and get that O is a $\text{GTGD}_{n,m}$-ontology, as needed. □

We conclude this section by observing that Theorem 7.4 provides a characterization of when an ontology is a GTGD-ontology.

Corollary 7.5. *Given an ontology O, the following are equivalent:*

(1) O is a GTGD*-ontology.*
(2) O is critical, is closed under direct prod., and is guarded local.

8 CHARACTERIZING FGTGD-ONTOLOGIES

We now concentrate on frontier-guarded tgds, and provide characterizations similar to those established in the previous sections. This is achieved by exploiting the so-called frontier-guarded locality property (another refinement of our locality property), as well as criticality and closure under direct products.

8.1 Frontier-Guarded Locality

The refined versions of the notion of local embedding for linear and guarded tgds were essentially emerged from the syntactic shape of the tgd bodies. This is somehow also the case for frontier-guarded local embeddings. Frontier-guardedness can be seen as a relativized version of guardedness. Indeed, a frontier-guarded body is essentially a guarded body relative to the frontier, i.e., only the frontier should satisfy the guardedness condition. In the same spirit, one can define frontier-guarded instances, which are essentially guarded instances relative to a certain set of active domain elements.

Consider an instance I, and a finite set $F \subseteq \text{adom}(I)$. We say that I is *guarded relative to F*, or simply *F-guarded*, if either $\text{facts}(I) = \emptyset$, or there is a fact $R(c_1, \ldots, c_k) \in \text{facts}(I)$ such that $F \subseteq \{c_1, \ldots, c_k\}$. Consider now an ontology O over a schema $\mathbf{S}$, and an $\mathbf{S}$-instance I. For $n, m \geq 0$, we say that O *is fr-guardedly (n, m)-locally embeddable in I* if, for every finite set $F \subseteq \text{adom}(I)$ and F-guarded $\mathbf{S}$-instance $K \leq I$ with $|\text{adom}(K)| \leq n$, there exists $J_K^F \in O$ such that $K \subseteq J_K^F$, and for every J' in the m-neighbourhood of F in J_K^F, there exists a function $h : \text{adom}(J') \to \text{adom}(I)$, which is the identity on F, such that $h(\text{facts}(J')) \subseteq \text{facts}(I)$. The property of frontier-guarded locality is defined as expected.

Definition 8.1 (**Frontier-Guarded Locality**). An ontology O over $\mathbf{S}$ is *frontier-guarded (n, m)-local*, for $n, m \geq 0$, if, for every $\mathbf{S}$-instance I, O is fr-guardedly (n, m)-locally embeddable in I implies $I \in O$. We further say that O is *frontier-guarded local* if there are $n, m \geq 0$ such that O is frontier-guarded (n, m)-local. ■

As for the properties of linear and guarded locality, we can show that frontier-guarded locality implies locality; the proof, which is omitted, is similar to that of Lemma 6.2

Lemma 8.2. *Consider an ontology O that is frontier-guarded (n, m)-local for some $n, m \geq 0$. It holds that O is (n, m)-local.*

8.2 The Characterization

We now show that frontier-guarded locality is an appropriate notion towards our characterization for frontier-guarded tgds. To this end, we first present a technical lemma in the spirit of the Linearization and Guardedization Lemmas; the proof is deferred to the Appendix.

Lemma 8.3. *Given a $\text{TGD}_{n,m}$-ontology O, for some $n, m \geq 0$, the following are equivalent:*

(1) O is an $\text{FGTGD}_{n,m}$-ontology.
(2) O is frontier-guarded (n, m)-local.

Having Lemma 8.3 in place, it is now easy to obtain the desired characterizations for frontier-guarded tgds.

Theorem 8.4. *Given an ontology O, and integers $n, m \geq 0$, the following are equivalent:*

(1) O is an $\text{FGTGD}_{n,m}$-ontology.
(2) O is critical, closed under direct products, and frontier-guarded (n, m)-local.

Proof. The fact that (1) ⇒ (2) follows from Lemmas 3.2, 3.4, and 8.3. For (2) ⇒ (1), since O is frontier-guarded (n, m)-local, Lemma 8.2 implies that O is (n, m)-local. Since, by hypothesis, O is also critical and closed under direct products, we get from Theorem 4.1 that O is a $\text{TGD}_{n,m}$-ontology, which in turn allows us to apply the Frontier-Guardedization Lemma (direction (2) ⇒ (1)), and get that O is an $\text{FGTGD}_{n,m}$-ontology, as needed. □

We conclude this section by observing that Theorem 8.4 provides a characterization of when an ontology is an FGTGD-ontology.

Corollary 8.5. *Given an ontology O, the following are equivalent:*

(1) O is an FGTGD*-ontology.*
(2) O is critical, is closed under direct products, and is frontier-guarded local.

9 RELATIVE EXPRESSIVENESS AND REWRITABILITY

By exploiting our new locality properties, and, in particular, the Linearization and Guardedization Lemmas, we can easily separate, in terms of expressive power, LTGD, GTGD and FGTGD. More precisely, we can devise a finite set of guarded (respectively, frontier-guarded) tgds that cannot be equivalently rewritten as a finite set of linear (respectively, guarded) tgds. Actually, those separations were folklore, and made explicit in the recent work [17]. The value of our analysis is that it provides further insights, using the linear and guarded locality properties, on why those separations hold.

9.1 Semantic Separations

Linear vs. Guarded. Let us first separate LTGD from GTGD. To this end, we need to devise a set $\Sigma_G \in$ GTGD that is provably not equivalent to a set $\Sigma_L \in$ LTGD. Consider the singleton set

$$\Sigma_G = \{R(x), P(x) \to T(x)\}.$$

By the Linearization Lemma (directions (1) ⇔ (3)), there exists $\Sigma_L \in$ LTGD such that $\Sigma_G \equiv \Sigma_L$ iff Σ_G is linear $(1, 0)$-local.[4] But, we can show that Σ_G is *not* linear $(1, 0)$-local, and thus, such a set Σ_L does not exist. To this end, we devise the $\{R, P, T\}$-instance I with

$$\text{dom}(I) = \{c\} \quad \text{and} \quad \text{facts}(I) = \{R(c), P(c)\}.$$

for which it is easy to verify that Σ_G is linearly $(1, 0)$-locally embeddable in I, but $I \not\models \Sigma_G$. Therefore, Σ_G is not linear $(1, 0)$-local.

Guarded vs. Frontier-Guarded. Let us now separate GTGD from FGTGD. We need to devise a set $\Sigma_F \in$ FGTGD that is provably not equivalent to a set $\Sigma_G \in$ GTGD. Consider the singleton set

$$\Sigma_F = \{R(x), P(y) \to T(x)\}.$$

By the Guardedization Lemma (directions (1) ⇔ (3)), there is $\Sigma_G \in$ GTGD such that $\Sigma_F \equiv \Sigma_G$ iff Σ_F is guarded $(2, 0)$-local. But, we can

[4]By abuse of terminology, we say that Σ_G is linear $(1, 0)$-local meaning that the ontology consisting of the models of Σ_G is linear $(1, 0)$-local .

Input: A set $\Sigma \in \mathsf{GTGD}_{n,m}$ for $n, m \geq 0$ over $\mathbf{S}$
Output: A set $\Sigma' \in \mathsf{LTGD}$ such that $\Sigma \equiv \Sigma'$, if one exists; otherwise, $\bot$

$\Sigma' := \{\sigma \mid \sigma \text{ is over } \mathbf{S}, \{\sigma\} \in \mathsf{LTGD}_{n,m} \text{ and } \Sigma \models \sigma\}$
if $\Sigma' \neq \emptyset$ and $\Sigma' \models \Sigma$ **then**
  return Σ'
else
  return $\bot$

Algorithm 1: G-to-L

show that Σ_F is *not* guarded $(2, 0)$-local, and thus, such a set Σ_G does not exist. To this end, we devise the $\{R, P, T\}$-instance I with

$$\mathrm{dom}(I) = \{c\} \quad \text{and} \quad \mathrm{facts}(I) = \{R(c), P(d)\}.$$

for which it is easy to verify that Σ_F is guardedly $(2, 0)$-locally embeddable in I, but $I \not\models \Sigma_F$. Hence, Σ_F is not guarded $(2, 0)$-local.

9.2 Rewritability

Having the above semantic separations in place, the next natural question is whether we can decide if a finite set of guarded (respectively, frontier-guarded) tgds can be equivalently rewritten as a finite set of linear (respectively, guarded) tgds. This brings us to the following decision problem; let C_1 and C_2 be classes of tgds:

PROBLEM :	$\mathsf{Rewrite}(\mathsf{C}_1, \mathsf{C}_2)$
INPUT :	A set of tgds $\Sigma \in \mathsf{C}_1$.
QUESTION :	Is there a set $\Sigma' \in \mathsf{C}_2$ such that $\Sigma \equiv \Sigma'$?

The rest of this section is devoted to studying the problems Rewrite(GTGD, LTGD) and Rewrite(FGTGD, GTGD).

From Guarded to Linear. The problem Rewrite(GTGD, LTGD) has been already considered in the recent work [17]. It was shown to be PSPACE-hard, but no decision procedure was provided. We proceed to pinpoint the complexity of this problem by exploiting the results of this work, in particular, the Linearization Lemma.

THEOREM 9.1. *The following hold:*

(1) Rewrite(GTGD, LTGD) *is* 2EXPTIME*-complete, and* EXPTIME*-complete for schemas of bounded arity.*
(2) Given a set $\Sigma \in \mathsf{GTGD}$*, a set* $\Sigma' \in \mathsf{LTGD}$ *such that* $\Sigma \equiv \Sigma'$*, if one exists, can be computed in double exponential time, and in exponential time in the case of schemas of bounded arity.*

The complexity lower bounds claimed in item (1) are shown via a reduction from conjunctive query answering under guarded tgds, which is 2EXPTIME-hard in general, and EXPTIME-hard for schemas of bounded arity [5]; the proof is given in the Appendix. Consider now a set $\Sigma \in \mathsf{GTGD}$ over a schema $\mathbf{S}$, where each tgd of Σ has at most $n \geq 0$ universally and at most $m \geq 0$ existentially quantified variables; it is clear that $\Sigma \in \mathsf{GTGD}_{n,m}$. By the Linearization Lemma, we know that the following statements are equivalent:

- There exists $\Sigma' \in \mathsf{LTGD}$ over $\mathbf{S}$ such that $\Sigma \equiv \Sigma'$.
- There exists $\Sigma' \in \mathsf{LTGD}_{n,m}$ over $\mathbf{S}$ such that $\Sigma \equiv \Sigma'$.

This essentially means that, even though there are infinitely many finite sets of linear tgds over $\mathbf{S}$, it suffices to search only the fragment of $\mathsf{LTGD}_{n,m}$ over $\mathbf{S}$, which is finite, to find a set Σ' that is equivalent to Σ. This leads to the very simple procedure depicted in Algorithm 1. It first collects in Σ' all the linear tgds over $\mathbf{S}$ with at most n universally, and at most m existentially quantified variables, that are entailed by the input set of tgds Σ, and then checks whether Σ' is non-empty and entails Σ; the latter is actually done by checking whether $\Sigma' \models \sigma$ for each $\sigma \in \Sigma$. We proceed to show that the algorithm G-to-L runs in double exponential time in general, and in single exponential time in the case of schemas of bounded arity, which in turn implies items (1) and (2) of Theorem 9.1.

We first observe that the total number of linear tgds over $\mathbf{S}$ with at most n universally quantified variables, and at most m existentially quantified variables, is bounded by

$$\underbrace{|\mathbf{S}| \cdot n^{\mathrm{ar}(\mathbf{S})}}_{\geq \text{ \# of linear bodies}} \cdot \underbrace{2^{|\mathbf{S}| \cdot (n+m)^{\mathrm{ar}(\mathbf{S})}}}_{\geq \text{ \# of heads}}$$

where $\mathrm{ar}(\mathbf{S}) = \max_{R \in \mathbf{S}}\{\mathrm{ar}(R)\}$, while each such linear tgd is of size

$$O\left(\mathrm{ar}(\mathbf{S}) \cdot |\mathbf{S}| \cdot (n+m)^{\mathrm{ar}(\mathbf{S})}\right).$$

We also need to understand the complexity of deciding whether a set of guarded tgds entails a linear tgd (needed in the construction of Σ'), as well as the complexity of deciding whether a set of linear tgds entails a guarded tgd (needed for checking whether $\Sigma' \models \Sigma$). It is easy to see that, given a set of tgds Σ and a single tgd σ of the form $\phi(\bar{x}, \bar{y}) \rightarrow \exists \bar{z}\, \psi(\bar{x}, \bar{z})$, $\Sigma \models \sigma$ iff Σ and the database D_ϕ, obtain by "freezing" $\phi(\bar{x}, \bar{y})$, entails the Boolean conjunctive query q_ϕ obtained from $\exists \bar{z}\, \psi(\bar{x}, \bar{z})$ after "freezing" $\bar{x}$; see, e.g., [13]. Therefore, the complexity of the implication problems in question can be easily inherited from existing results on conjunctive query answering under guarded and linear tgds [5, 6]:

- Given a set $\Sigma_G \in \mathsf{GTGD}$ and a linear tgd σ_L, both over $\mathbf{S}$, we can decide whether $\Sigma_G \models \sigma_L$ in double exponential time in $\mathrm{ar}(\mathbf{S})$, and single exponential time in the size of Σ_G and σ_L.
- Given a set $\Sigma_L \in \mathsf{LTGD}$ and a guarded tgd σ_G, both over $\mathbf{S}$, we can decide whether $\Sigma_L \models \sigma_G$ in exponential time in $\mathrm{ar}(\mathbf{S})$ and the size of σ_G, and in polynomial time in the size of Σ_L.

Putting everything together, we get that the algorithm G-to-L runs in double exponential time in general, and in single exponential time in the case of schemas of bounded arity, as needed.

From Frontier-Guarded to Guarded. Rewrite(FGTGD, GTGD) has been studied in [17], and it was shown to be 2EXPTIME-complete, but without providing an effective procedure that builds an equivalent set of guarded tgds, if one exists. We can provide such a procedure by using the results of this work, in particular, the Guardedization Lemma. We further show that the 2EXPTIME-hardness of Rewrite(FGTGD, GTGD) holds even in the case of bounded arity.

THEOREM 9.2. *The following hold:*

(1) Rewrite(FGTGD, GTGD) *is* 2EXPTIME*-complete even for schemas of bounded arity.*
(2) Given a set $\Sigma \in \mathsf{FGTGD}$*, a set* $\Sigma' \in \mathsf{GTGD}$ *such that* $\Sigma \equiv \Sigma'$*, if one exists, can be computed in triple exponential time, and in double exponential time for schemas of bounded arity.*

The 2EXPTIME-hardness claimed in item (1), even for schemas of bounded arity, is shown via a reduction from conjunctive query answering under frontier-guarded tgds [3]; the proof is deferred

Input: A set $\Sigma \in \mathsf{FGTGD}_{n,m}$ for $n, m \geq 0$ over $\mathbf{S}$
Output: A set $\Sigma' \in \mathsf{GTGD}$ such that $\Sigma \equiv \Sigma'$, if one exists; otherwise, $\bot$

$\Sigma' := \{\sigma \mid \sigma \text{ is over } \mathbf{S}, \{\sigma\} \in \mathsf{GTGD}_{n,m} \text{ and } \Sigma \models \sigma\}$
if $\Sigma' \neq \emptyset$ and $\Sigma' \models \Sigma$ **then**
  return Σ'
else
  return $\bot$

Algorithm 2: FG-to-G

to the Appendix. Consider now a set $\Sigma \in \mathsf{FGTGD}$ over a schema $\mathbf{S}$, where each tgd of Σ has at most $n \geq 0$ universally and at most $m \geq 0$ existentially quantified variables; it is clear that $\Sigma \in \mathsf{FGTGD}_{n,m}$. By the Guardedization Lemma, the following are equivalent:

- There exists $\Sigma' \in \mathsf{GTGD}$ over $\mathbf{S}$ such that $\Sigma \equiv \Sigma'$.
- There exists $\Sigma' \in \mathsf{GTGD}_{n,m}$ over $\mathbf{S}$ such that $\Sigma \equiv \Sigma'$.

This leads to the simple procedure depicted in Algorithm 2 that constructs an equivalent set of guarded tgds, if one exists. It first collects in Σ' all the guarded tgds over $\mathbf{S}$ with at most n universally, and at most m existentially quantified variables, that are entailed by the input set of tgds Σ, and then checks whether Σ' is non-empty and entails Σ. We proceed to analyze its running time.

We observe that the total number of guarded tgds over $\mathbf{S}$ with at most n universally quantified variables, and at most m existentially quantified variables, is bounded by

$$\underbrace{2^{|\mathbf{S}| \cdot n^{\mathrm{ar}(\mathbf{S})}}}_{\geq \text{ \# of guadred bodies}} \cdot \underbrace{2^{|\mathbf{S}| \cdot (n+m)^{\mathrm{ar}(\mathbf{S})}}}_{\geq \text{ \# of heads}}$$

while each such guarded tgd is of size

$$O\left(\mathrm{ar}(\mathbf{S}) \cdot |\mathbf{S}| \cdot (n+m)^{\mathrm{ar}(\mathbf{S})}\right).$$

Concerning the implication checks, we can easily inherit from existing results on conjunctive query answering under frontier-guarded and guarded tgds [3, 5] the following complexities:

- Given $\Sigma_F \in \mathsf{FGTGD}$ and a guarded tgd σ_G, both over $\mathbf{S}$, we can decide whether $\Sigma_F \models \sigma_G$ in double exponential time.
- Given $\Sigma_G \in \mathsf{GTGD}$ and a frontier-guarded tgd σ_F, both over $\mathbf{S}$, we can decide whether $\Sigma_G \models \sigma_F$ in double exponential time in $\mathrm{ar}(\mathbf{S})$, and single exponential time in the size of Σ_G.

Putting everything together, we get that FG-to-G runs in triple exponential time in general, and in double exponential time in the case of schemas of bounded arity, and Theorem 9.2 follows.

10 CONCLUDING REMARKS

In this work, we have established model-theoretic characterizations of TGD-ontologies, including characterizations of ontologies specified by central classes of tgds, such as full, linear, guarded, and frontier-guarded tgds. Our characterizations use the well-known properties of criticality and closure under direct products, as well as a novel locality property for TGD-ontologies. We further used this locality property to decide whether an ontology expressed by frontier-guarded (respectively, guarded) tgds can be rewritten as an equivalent one expressed by tgds in the weaker class of guarded (respectively, linear) tgds, and effectively construct such an equivalent ontology if one exists. Although our results (model-theoretic characterizations and rewritability) focus on unrestricted (finite or infinite) instances, it can be shown that they also hold if we concentrate on finite instances. As a next step, we would like to perform a similar analysis that goes beyond TGD-ontologies. In particualr, we are planning to consider ontologies specified by tgds, egds, and denial constraints. Moreover, in the context of rewritability, it is interesting to investigate the optimality of the size of the equivalent linear or guarded sets of tgds that we build (whenever they exist).

REFERENCES

[1] Marcelo Arenas, Pablo Barceló, Leonid Libkin, and Filip Murlak. 2014. *Foundations of Data Exchange.* Cambridge University Press.
[2] Franz Baader, Ian Horrocks, and Ulrike Sattler. 2009. Description Logics. In *Handbook on Ontologies.* 21–43.
[3] Jean-François Baget, Marie-Laure Mugnier, Sebastian Rudolph, and Michaël Thomazo. 2011. Walking the Complexity Lines for Generalized Guarded Existential Rules. In *IJCAI.* 712–717.
[4] Meghyn Bienvenu and Magdalena Ortiz. 2015. Ontology-Mediated Query Answering with Data-Tractable Description Logics. In *Reasoning Web.* 218–307.
[5] Andrea Calì, Georg Gottlob, and Michael Kifer. 2013. Taming the Infinite Chase: Query Answering under Expressive Relational Constraints. *J. Artif. Intell. Res.* 48 (2013), 115–174.
[6] Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. 2012. A general Datalog-based framework for tractable query answering over ontologies. *J. Web Sem.* 14 (2012), 57–83.
[7] Andrea Calì, Georg Gottlob, Thomas Lukasiewicz, Bruno Marnette, and Andreas Pieris. 2010. Datalog+/-: A Family of Logical Knowledge Representation and Query Languages for New Applications. In *LICS.* 228–242.
[8] Chen C. Chang and H. Jerome Keisler. 1992. *Model theory, Third Edition.* North-Holland.
[9] Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. 2005. Data exchange: semantics and query answering. *Theor. Comput. Sci.* 336, 1 (2005), 89–124.
[10] Ronald Fagin and Moshe Y. Vardi. 1986. The Theory of Data Dependencies - A Survey. In *Symposia in Applied Mathematics.* 19–71.
[11] Maurizio Lenzerini. 2002. Data Integration: A Theoretical Perspective. In *PODS.* 233–246.
[12] Carsten Lutz, Robert Piro, and Frank Wolter. 2011. Description Logic TBoxes: Model-Theoretic Characterizations and Rewritability. In *IJCAI.* 983–988.
[13] David Maier, Alberto O. Mendelzon, and Yehoshua Sagiv. 1979. Testing Implications of Data Dependencies. *ACM Trans. Database Syst.* 4, 4 (1979), 455–469.
[14] Johann A. Makowsky and Moshe Y. Vardi. 1986. On the Expressive Power of Data Dependencies. *Acta Inf.* 23, 3 (1986), 231–244.
[15] Marie-Laure Mugnier and Michaël Thomazo. 2014. An Introduction to Ontology-Based Query Answering with Existential Rules. In *Reasoning Web.* 245–278.
[16] Balder ten Cate and Phokion G. Kolaitis. 2009. Structural characterizations of schema-mapping languages. In *ICDT.* 63–72.
[17] Heng Zhang, Yan Zhang, and Guifei Jiang. 2020. Model-theoretic Characterizations of Existential Rule Languages. In *IJCAI.*

A PROOFS FROM SECTION 4

Proof of Claim 4.6

Recall that $\Phi^I_{K,m}(\bar{x})$ is obtained from the m-diagram of K relative to I by renaming each constant $c \in \text{dom}(K)$ to a new variable x_c; let ρ be the renaming function, i.e., $\rho(c) = x_c$ for each $c \in \text{dom}(K)$. Therefore, $\Phi^I_{K,m}(\bar{x})$ is a formula of the form

$$\bigwedge_{\alpha \in \text{facts}(K)} \rho(\alpha) \wedge \bigwedge_{\substack{c,d \in \text{dom}(K),\\ c \neq d}} \neg(\rho(c) = \rho(d)) \wedge \bigwedge_{\substack{\gamma(\bar{y}) \in C_{K,\ell},\\ I \not\models \exists \bar{y}\, \gamma(\bar{y})}} \neg \exists \bar{y}\, \rho(\gamma(\bar{y})).$$

It is clear that $\neg \exists \bar{x}\, \Phi^I_{K,m}(\bar{x})$ is equivalent to the sentence

$$\delta \;=\; \forall \bar{x}(\phi(\bar{x}) \;\rightarrow\; \psi(\bar{x})),$$

where

$$\phi(\bar{x}) \;=\; \bigwedge_{\alpha \in \text{facts}(K)} \rho(\alpha)$$

$$\psi(\bar{x}) \;=\; \bigvee_{\substack{c,d \in \text{dom}(K),\\ c \neq d}} \rho(c) = \rho(d) \vee \bigvee_{\substack{\gamma(\bar{y}) \in C_{K,\ell},\\ I \models \exists \bar{y}\, \gamma(\bar{y})}} \exists \bar{y}\, \rho(\gamma(\bar{y})).$$

It remains to show that δ is an edd from $\mathsf{E}_{n,m}$. To this end, we need to show that (i) $\psi(\bar{x})$ is non-empty, (ii) each variable in $\psi(\bar{x})$ is either existentially quantified, or appears in $\phi(\bar{x})$, and (iii) δ mentions at most n universally and m existentially quantified variables:

(i) Observe that $\Phi^I_{K,m}(\bar{x})$ has at least one negative conjunct. Assume, by contradiction, that this is not the case. Then, a 1-critical instance in O, which exists since O is 1-critical, is trivially a model of $\exists \bar{x}\, \Phi^I_{K,m}(\bar{x})$. But this contradicts Claim 4.5, which states that, for each $J \in O$, $J \not\models \exists \bar{x}\, \Phi^I_{K,m}(\bar{x})$. Therefore, we conclude that $\psi(\bar{x})$ is non-empty.

(ii) Observe that $\text{dom}(K) = \text{adom}(K)$ since $K \leq I$ and $\text{dom}(I) = \text{adom}(I)$; recall that the latter relies on the domain independence of O, which is guaranteed by Lemma 3.8. This implies that each variable in $\psi(\bar{x})$ is either existentially quantified, or appears in $\phi(\bar{x})$, as needed.

(iii) Finally, it is easy to verify that δ has at most n universally and m existentially quantified variables since $|\text{dom}(K)| = |\text{adom}(K)| \leq n$, while, by construction, $\Phi^I_{K,m}(\bar{x})$ mentions at most m existentially quantified variables.

This completes the proof of Claim 4.6.

Proof of Claim 4.8

Since, for each $j \in [k]$, $I_j \not\models \sigma_j$, we conclude that there exists a function $h_j : \bar{x} \rightarrow \text{dom}(I_j)$ such that $h_j(\phi(\bar{x})) \subseteq \text{facts}(I_j)$. We define the function $h : \bar{x} \rightarrow \text{dom}(J)$ such that, for each variable $x \in \bar{x}$.

$$h(x) \;=\; (h_1(x), \ldots, h_k(x)).$$

By the definition of direct products, we conclude that $h(\phi(\bar{x})) \subseteq \text{facts}(J)$. It remains to show that, for each $j \in [k]$, if $\psi_j(\bar{x}_j)$ is the equality expression $y = z$, then $h(y) \neq h(z)$, and if $\psi_j(\bar{x}_j)$ is not an equality expression, then there is no extension h' of h such that $h'(\psi_j(\bar{x}_j)) \subseteq \text{facts}(J)$. Fix an arbitrary $j \in [k]$. We proceed by case analysis on the type of the formula $\psi_j(\bar{x}_j)$:

(1) Assume first that $\psi_j(\bar{x}_j)$ is the equality expression $y = z$. By contradiction, assume that $h(y) = h(z)$. By the definition of h, we get that $h_j(y) = h_j(z)$, which contradicts the fact that $I_j \models \exists \bar{x}(\phi(\bar{x}) \wedge \neg(y = z))$.

(2) Assume that $\psi_j(\bar{x}_j)$ is a formula of the form $\exists \bar{y}_j\, \chi_j(\bar{x}_j, \bar{y}_j)$, and assume, by contradiction, that there exists an extension h' of h such that $h'(\chi_j(\bar{x}_j, \bar{y}_j)) \subseteq \text{facts}(J)$. For a constant $c \in \text{dom}(J)$, we write $c[i]$ for the i-th component of c, which is actually a constant from $\text{dom}(I_i)$. We define the function $h'_j : \bar{x} \cup \bar{y}_j \rightarrow \text{dom}(I_j)$ such that $h'_j(v) = h'(v)[j]$ for each variable $v \in \bar{x} \cup \bar{y}_j$. Since h' is an extension of h, and $h_j(v) = h(v)[j]$, for each variable $v \in \bar{x}$, we conclude that h'_j is an extension of h_j. By hypothesis, $h'(\alpha) \in \text{facts}(J)$ for each conjunct α of $\chi_j(\bar{x}_j, \bar{y}_j)$. Assume that $h'(\alpha) = R(\bar{c}_i, \ldots, \bar{c}_m)$. By the definition of the direct product, we conclude that $R(\bar{c}_1[j], \ldots, c_m[j]) \in \text{facts}(I_j)$. Consequently, $h'_j(\psi_j(\bar{x}_j)) \subseteq \text{facts}(I_j)$, which in turn implies that $I_j \models \sigma_j$. But this is a contradiction, and the claim follows.

B PROOFS FROM SECTION 5

Proof of Theorem 5.6 – Direction (1) ⇒ (2)

By hypothesis, O is an FTGD-ontology, which means that there exists a set $\Sigma \in \text{FTGD}$ over S such that $I \in O$ iff $I \models \Sigma$. The fact that O is 1-critical follows by Lemma 3.2, while the fact that is domain independent follows from Lemma 3.8. We proceed to show that O enjoys the other three properties, i.e., (1) is n-modular for some $n > 0$, (2) is closed under intersections, and (3) is closed under non-oblivious duplicating extensions:

(1) Let $n \geq 0$ be the maximum number of variables in the body of a tgd of Σ. Consider an S-instance $I \notin O$. There exists a full tgd $\sigma \in \Sigma$ of the form $\phi(\bar{x}, \bar{y}) \rightarrow \psi(\bar{x})$ such that $I \not\models \sigma$, which can be witnessed via a function $h : \bar{x} \cup \bar{y} \rightarrow \text{dom}(I)$. Let C be the set of constants $h(\bar{x}) \cup h(\bar{y})$. Consider the S-instance I_σ such that $\text{dom}(I_\sigma) = C$, and, for each $R \in \mathsf{S}$, $R^{I_\sigma} = R^I_{|C}$. Clearly, $I_\sigma \leq I$ and $|\text{dom}(I_\sigma)| \leq n$. It is also easy to see that $I_\sigma \not\models \sigma$, and thus, $I_\sigma \notin O$. Therefore, O is n-modular.

(2) Let $I, J \in O$ and assume that, for some full tgd $\sigma \in \Sigma$ of the form $\phi(\bar{x}, \bar{y}) \rightarrow \psi(\bar{x})$, there exists a function $h : \bar{x} \cup \bar{y} \rightarrow \text{dom}(I \cap J)$ such that $h(\phi(\bar{x}, \bar{y})) \subseteq \text{facts}(I \cap J)$. By definition, $h(\phi(\bar{x}, \bar{y})) \subseteq \text{facts}(I)$ and $h(\phi(\bar{x}, \bar{y})) \subseteq \text{facts}(J)$. Since $I \models \Sigma$ and $J \models \Sigma$, we get that $h(\psi(\bar{x}))$ is subeset of both $\text{facts}(I)$ and $\text{facts}(J)$. Therefore, $h(\psi(\bar{x})) \subseteq \text{facts}(I \cap J)$, which implies that $I \cap J \models \sigma$, and thus, $I \cap J \in O$.

(3) Consider an instance $I \in O$, and let J be a non-oblivious duplicating extension of I. We need to show that $J \in O$, or, equivalently, $J \models \Sigma$. By definition, there are constants $c \in \text{dom}(I)$ and $d \notin \text{dom}(I)$ such that $d \in \text{dom}(J) \setminus \text{dom}(I)$, and the function $h_c : \text{dom}(I) \cup \{d\} \rightarrow \text{dom}(I)$ that is the identity on $\text{dom}(I)$ and $h(d) = c$, is such that $h_c(\text{facts}(J)) \subseteq \text{facts}(I)$. Assume there is a full tgd $\sigma \in \Sigma$ of the form $\phi(\bar{x}, \bar{y}) \rightarrow \psi(\bar{x})$ and a function $h : \bar{x} \cup \bar{y} \rightarrow \text{dom}(J)$ such that $h(\phi(\bar{x}, \bar{y})) \subseteq \text{facts}(J)$. Therefore, the function $\mu = h_c \circ h$ is such that $\mu(\phi(\bar{x}, \bar{y})) \subseteq \text{facts}(I)$. Since $I \models \sigma$, $\mu(\psi(\bar{x})) \subseteq \text{facts}(I)$. From the definition of non-oblivious duplicating extensions, we can conclude that $\mu(\psi(\bar{x})) \subseteq \text{facts}(J)$, and the claim follows.

This completes the proof.

Proof of Theorem 5.6 – Direction (2) ⇒ (1)

We first show the following auxiliary claim:

Claim B.1. *O is closed under subinstances.*

Proof. Since O is 1-critical, there are 1-critical instances $I, J \in O$ such that $\text{dom}(I) \neq \text{dom}(J)$. Since O is closed under intersections, we have that the empty instance is in O. Consider now two instances I, J such that $I \in O$, and $J \preceq I$. We proceed to show that $J \in O$. If $\text{dom}(J) = \emptyset$, then J is the empty instance, and therefore $J \in O$. Suppose now that $\text{dom}(J) \neq \emptyset$, and let K be a critical instance such that $\text{dom}(K) = \text{dom}(J)$. Since O is critical, we can conclude that $K \in O$. From closure under intersections, we can conclude that $K \cap I \in O$. It remains to show that $K \cap I = J$. Consider an n-ary relation $R \in \mathbf{S}$. Since K is critical, it has a fact $R(\bar{c})$ for each $\bar{c} \in \text{dom}(K)^n$. This implies that $\text{facts}(K \cap I)$ contains every fact $R(\bar{c})$ of $\text{facts}(I)$ such that $\bar{c} \in \text{dom}(J)^n$, and therefore $K \cap I = J$. ☐

Before we proceed further, we need to recall some auxiliary notions. A *disjunctive dependency* (dd) over a schema $\mathbf{S}$ is an edd without existentially quantified variables, and every disjunct in the right-hand side of the implication is either an equality expression or an atom over $\mathbf{S}$. We also need the standard notion of the diagram of an instance. Consider an $\mathbf{S}$-instance I, and let A be the set of atomic formulas that can be formed using predicates from $\mathbf{S}$ and constants from $\text{dom}(I)$. The *diagram of* I, denoted Δ_I, is

$$\bigwedge_{\alpha \in I} \alpha \ \wedge \ \bigwedge_{\alpha \in A \setminus I} \neg \alpha \ \wedge \ \bigwedge_{c,d \in \text{dom}(I) \text{ and } c \neq d} \neg(c = d).$$

We can now proceed with the rest of the proof. Assuming that O is over the schema $\mathbf{S}$, let $\Sigma^{\vee}$ be the set of all dds over $\mathbf{S}$ with at most n variables that are satisfied by every $I \in O$, i.e., $\delta \in \Sigma^{\vee}$ iff $I \models \delta$ for each instance $I \in O$. It is clear that $\Sigma^{\vee}$ is finite (up to logical equivalence). We proceed to show the following technical lemma.

Lemma B.2. *For each $\mathbf{S}$-instance I, $I \in O$ iff $I \models \Sigma^{\vee}$.*

Proof. The (⇒) direction holds by construction. Consider an $\mathbf{S}$-instance $I \notin O$. We show that $I \not\models \Sigma^{\vee}$. Since O is n-modular for some $n > 0$, there exists an $\mathbf{S}$-instance $I_n \notin O$, with $|\text{dom}(I_n)| \leq n$, such that $I_n \preceq I$. Since O is domain independent, we can assume that $\text{dom}(I_n) = \text{adom}(I_n)$. We proceed to show that $I_n \not\models \Sigma^{\vee}$, which in turn implies that $I \not\models \Sigma^{\vee}$ since universal first-order sentences are preserved under subinstances, as needed. Let Δ_{I_n} be the diagram of I_n, and let $\Phi_{I_n}(\bar{x})$ be the formula obtained from Δ_{I_n} by replacing each $c \in \text{dom}(I_n)$ with a new variable $x_c \in \mathbf{V}$. We show that:

Claim B.3. *For every $J \in O$, $J \models \neg \exists \bar{x}\, \Phi_{I_n}(\bar{x})$.*

Proof. By contradiction, assume that there exists $J \in O$ such that $J \models \exists \bar{x}\, \Phi_{I_n}(\bar{x})$. We can show that there is an $\mathbf{S}$-instance K such that $K \preceq J$ and $I_n \simeq K$. Since, by Claim B.1, O is closed under subinstances, we get that $K \in O$, and thus, $I_n \in O$ (by definition, O is closed under isomorphisms), which is a contradiction. The rest of the proof is devoted to showing the existence of the instance K.

Let ρ be the renaming function used to obtain the formula $\Phi_{I_n}(\bar{x})$ from the diagram of I_n, i.e., the bijection that maps each constant $c \in \text{dom}(I_n)$ to a distinct variable $x_c \in \mathbf{V}$. The fact that $J \models \exists \bar{x}\, \Phi_{I_n}(\bar{x})$ can be witnessed via a function $h : \bar{x} \rightarrow \text{dom}(J)$. Let $C = \{h(x) \mid x \in \bar{x}\} \subseteq \text{dom}(J)$, and let K be the $\mathbf{S}$-instance such that $\text{dom}(K) = C$, and, for each $R \in \mathbf{S}$, $R^K = R^J_{|C}$. Clearly, $K \preceq J$. Moreover, since in $\Phi_{I_n}(\bar{x})$ there is a conjunct $\neg(x_i = x_j)$ for each pair of distinct variables x_i, x_j of $\bar{x}$, which in turn implies that $h(x_i) \neq h(x_j)$, it is not difficult to verify that $\mu = h \circ \rho$ is an 1-1 function such that $\mu(\text{facts}(I_n)) \subseteq K$, and μ^{-1} is such that $\mu^{-1}(\text{facts}(K)) \subseteq J$. Therefore, $I_n \simeq K$, and the claim follows. ☐

We now proceed to show the following claim:

Claim B.4. *There exists a dd δ such that $\delta \equiv \neg \exists \bar{x}\, \Phi_{I_n}(\bar{x})$.*

Proof. Since O is closed under subinstances, we get that $\text{facts}(I_n) \neq \emptyset$; this holds since the empty instance is a subinstance of every $\mathbf{S}$-instance. This implies that Δ_{I_n} contains at least one positive relational atom α. Now, since O contains an 1-critical $\mathbf{S}$-instance, and is closed under isomorphisms, we conclude that I_n is not an 1-critical instance; otherwise, $I_n \in O$, which is not the case. This implies that either $|\text{dom}(I_n)| \geq 2$, or $|\text{dom}(I_n)| = 1$ and there is a relation $R \in \mathbf{S}$ such that R^{I_n} is empty; otherwise, I_n would be 1-critical, which is not the case. Therefore, Δ_{I_n} contains either the negation of an equality atom β, or the negation of a relational atom γ. Consequently, $\neg \exists \bar{x}\, \Phi_{I_n}(\bar{x})$ can be equivalently rewritten as a dd. The latter relies on the fact that, since $\text{dom}(I_n) = \text{adom}(I_n)$ (recall that this exploits domain independence), in Δ_{I_n} each $c \in \text{dom}(I_n)$ occurring in a negative atom occurs also in a positive atom. ☐

By Claims B.3 and B.4, there exists a dd $\delta \in \Sigma^{\vee}$ such that $\delta \equiv \neg \exists \bar{x}\, \Phi_{I_n}(\bar{x})$, and therefore, $I_n \not\models \Sigma^{\vee}$. Since $I_n \preceq I$, we get that $I \not\models \Sigma^{\vee}$. This completes the proof of Lemma B.2 ☐

Having Lemma B.2 in place, to show that O is an FTGD-ontology, it remains to establish the following lemma:

Lemma B.5. *There exists a set $\Sigma \in$ FTGD such that $\Sigma \equiv \Sigma^{\vee}$.*

Proof. We define Σ as the set of tgds

$$\left\{\delta \in \Sigma^{\vee} \mid \delta \text{ is a full tgd}\right\},$$

i.e., Σ is the subset of $\Sigma^{\vee}$ consisting of full TGDs. We show that $\Sigma \models \Sigma^{\vee}$; the other direction holds trivially. By contradiction, assume that $\Sigma \not\models \Sigma^{\vee}$. Thus, there exists a dd $\delta \in \Sigma^{\vee}$ of the form

$$\forall \bar{x} \left(\phi(\bar{x}) \rightarrow \bigvee_{i=1}^{k} \alpha_i(\bar{x}_i)\right)$$

such that $\Sigma \not\models \delta$. Let $i_1, \ldots, i_\ell$ be distinct integers from $[k]$ such that the following holds: $\alpha_j(\bar{x}_j)$, where $j \in [k]$, is a relational atom (i.e., is not an equality expression) iff $j \in \{i_1, \ldots, i_\ell\}$. In simple words, $\{i_1, \ldots, i_\ell\}$ collects all the indices from $[k]$ such that $\alpha_{i_j}(\bar{x}_{i_j})$, for each $j \in [\ell]$, is a relational atom (and not an equality expression). We proceed to show the following auxiliary claim:

Claim B.6. *For each $j \in \{i_1, \ldots, i_\ell\}$ there exists $I_j \in O$ and an injective function $h : \bar{x} \rightarrow \text{dom}(I_j)$ such that $h(\alpha_j(\bar{x}_j)) \notin \text{facts}(I_j)$.*

Proof. Since $\Sigma \not\models \delta$, for each $j \in \{i_1, \ldots, i_\ell\}$, $\sigma_j = \forall \bar{x}(\phi(\bar{x}) \rightarrow \alpha_j(\bar{x}_j))$ is not in Σ. Hence, for each $j \in \{i_1, \ldots, i_\ell\}$, there exists $K_j \in O$ such that $K_j \models \exists \bar{x}(\phi(\bar{x}) \wedge \neg \alpha_j(\bar{x}_j))$. Thus, there exists a function $h : \bar{x} \rightarrow \text{dom}(K_j)$, for each $j \in \{i_1, \ldots, i_\ell\}$, such that $h(\alpha_j(\bar{x}_j)) \notin \text{facts}(K_j)$. Suppose that h maps two distinct variables x, y to the same constant $c \in \text{dom}(K_j)$, and let K'_j be a non-oblivious duplicating extension of K_j witnessed by c such that $\mathsf{dom}(K'_j) = \mathsf{dom}(K_j) \cup \{d\}$. Moreover, let $h' : \bar{x} \to \mathsf{dom}(K'_j)$ be the function such that $h'(z) = h(z)$, for each variable $z \neq y$, and $h'(y) = d$. We proceed to show the following claim:

Claim B.7. *$h'(\phi(\bar{x})) \subseteq \mathsf{facts}(K'_j)$ and $h'(\alpha_j(\bar{x}_j)) \notin \mathsf{facts}(K'_j)$.*

Proof. We first observe that there exists a function μ that is the identity on $\mathsf{dom}(K_j)$ and $\mu(d) = c$ such that $\mu(h'(\phi(\bar{x}))) \subseteq h(\phi(\bar{x}))$. From the fact that $h(\phi(\bar{x})) \subseteq \mathsf{facts}(K_j)$ and K'_j being a non-oblivious duplicating extension of K_j, we can conclude that $h'(\phi(\bar{x})) \subseteq \mathsf{facts}(K'_j)$. Suppose now, towards a contradiction, that $h'(\alpha_j(\bar{x}_j)) \in \mathsf{facts}(K'_j)$. This implies that $\mu(h'(\alpha_j(\bar{x}_j))) \in \mathsf{facts}(K_j)$ since K'_j is a non-oblivious duplicating extension of K_j. From the definition of μ and h', we get that $\mu(h'(\alpha_j(\bar{x}_j))) = h(\alpha_j(\bar{x}_j))$, contradicting the hypothesis that $h(\alpha_j(\bar{x}_j)) \notin \mathsf{facts}(K_j)$. □

With the above result in place, the desired I_j can be constructed starting from K_j and repeatedly duplicating constants. □

For each $j \in \{i_1, \dots, i_\ell\}$, let I_j be the instance provided by Claim B.6. By definition, there exists an injective function $h_j : \bar{x} \to \mathsf{dom}(I_j)$ such that $h_j(\phi(\bar{x})) \subseteq \mathsf{facts}(I_j)$ and $h_j(\alpha_j(\bar{x}_j)) \notin \mathsf{facts}(I_j)$. Since O is closed under isomorphisms, we can assume that, for every $i, j \in \{i_1, \dots, i_\ell\}$ and every variable x occurring in $\bar{x}$, $h_i(x) = h_j(x)$. With h being one of the functions h_j, $h(\phi(\bar{x})) \subseteq \mathsf{facts}(I_j)$, for each $j \in \{i_1, \dots, i_\ell\}$. Finally, let I be the intersection of each I_j, i.e., $I = \bigcap_j I_j$, with $j \in \{i_1, \dots, i_\ell\}$. Since $h(\phi(\bar{x})) \subseteq \mathsf{facts}(I_j)$, for each $j \in \{i_1, \dots, i_\ell\}$, we have that $h(\phi(\bar{x})) \subseteq \mathsf{facts}(I)$. To conclude the proof, we show that $h(\alpha_j(\bar{x}_j)) \notin \mathsf{facts}(I)$, for each $\alpha_j(\bar{x}_j)$ in $\mathsf{head}(\delta)$, contradicting the fact that $\delta \in \Sigma^{\vee}$. If $\alpha_j(\bar{x}_j)$ is an equality expression of the form $x = y$, by definition, $h(x) \neq h(y)$ sine h is an injective function. Assume now that $\alpha_j(\bar{x}_j)$ is a relational atom. Therefore, $j \in \{i_1, \dots, i_\ell\}$ and $h(\alpha_j(\bar{x}_j)) \notin \mathsf{facts}(I_j)$. Since $I = \bigcap_j I_j$, with $j \in \{i_1, \dots, i_\ell\}$, we can conclude $h(\alpha_j(\bar{x}_j)) \notin \mathsf{facts}(I_j)$. □

The claim follows from Lemma B.2 and Lemma B.5.

C PROOFS FROM SECTION 6

C.1 Proof of Lemma 6.3

We assume that O is over the schema $\mathsf{S} = \{R_1, \dots, R_\ell\}$. Since O is a $\mathsf{TGD}_{n,m}$-ontology, there exists a set $\Sigma \in \mathsf{TGD}_{n,m}$ over S such that, for every S-instance I, $I \in O$ iff $I \models \Sigma$. We proceed to show the directions (1) ⇒ (3) and (3) ⇒ (2); (2) ⇒ (1) holds trivially.

Direction (1) ⇒ (3)

Consider an S-instance I. We need to show the following: if O is linearly (n,m)-locally embeddable in I, then $I \in O$ (or, equivalently, $I \models \Sigma$), which in turn implies that O is linearly (n,m)-local, as needed. Consider a tgd $\sigma \in \Sigma$ of the form $\phi(\bar{x},\bar{y}) \to \exists \bar{z}\, \psi(\bar{x},\bar{z})$ such that there is a function $h : \bar{x} \cup \bar{y} \to \mathsf{dom}(I)$ with $h(\phi(\bar{x},\bar{y})) \subseteq \mathsf{facts}(I)$. We need to show that there exists an extension h' of h such that $h'(\psi(\bar{x},\bar{z})) \subseteq \mathsf{facts}(I)$. Observe that at most n distinct variables occur in $\bar{x} \cup \bar{y}$, which implies that each fact $\alpha \in h(\phi(\bar{x},\bar{y}))$ contains at most n distinct constants from $\mathsf{adom}(I)$. Since, by hypothesis, O is linearly (n,m)-locally embeddable in I, for each $K \subseteq h(\phi(\bar{x},\bar{y}))$ with $|\mathsf{facts}(K)| \leq 1$, there exists $J_K \in O$ such that $K \subseteq J_K$ and, for each J' in the m-neighbourhood of K in J_K, there exists a function $f_K : \mathsf{adom}(J') \to \mathsf{adom}(I)$, which is the identity on $\mathsf{adom}(K)$, such that $f(\mathsf{facts}(J')) \subseteq \mathsf{facts}(I)$. Since $h(\phi(\bar{x},\bar{y}))$ consists of finitely many atoms, and O is closed under isomorphisms, we can assume that $\mathsf{dom}(J_K) \cap \mathsf{dom}(J_L) = \mathsf{dom}(K) \cap \mathsf{dom}(L)$, for each $K, L \subseteq h(\phi(\bar{x},\bar{y}))$ with $|\mathsf{facts}(K)| \leq 1$ and $|\mathsf{facts}(L)| \leq 1$. In other words, we can assume that, for each pair $K, L \subseteq h(\phi(\bar{x},\bar{y}))$, with $|\mathsf{facts}(K)| \leq 1$ and $|\mathsf{facts}(L)| \leq 1$, the constants of $\mathsf{dom}(J_K)$ not occurring in K do not occur in $\mathsf{dom}(J_L)$ either. Let J be an S-instance such that

$$\mathsf{facts}(J) \;=\; \bigcup_{\substack{K \subseteq h(\phi(\bar{x},\bar{y})),\\ |\mathsf{facts}(K)| \leq 1}} \mathsf{facts}(J_K).$$

Since O is an LTGD-ontology, we have that O is closed under unions, which in turn implies that $J \in O$, and thus, $J \models \sigma$. From the definition of J, we get that $h(\phi(\bar{x},\bar{y})) \subseteq J$. Since $J \models \sigma$, there exists an extension g of h such that $g(\psi(\bar{x},\bar{z})) \subseteq J$. Let $L = (\mathsf{dom}(L), R_1^L, \dots, R_\ell^L)$ with $\mathsf{dom}(L) = \mathsf{dom}(g(\phi(\bar{x},\bar{y}))) \cup \mathsf{dom}(g(\psi(\bar{x},\bar{z})))$, and for $i \in [\ell]$, $R_i^L = R^J_{i\,|\mathsf{dom}(L)}$. Since, σ mentions at most m existentially quantified variables, we have that $g(\psi(\bar{x},\bar{z}))$ mentions at most m constants not occurring in $h(\phi(\bar{x},\bar{y}))$. This implies that $\mathsf{adom}(L)$ contains at most m constants not occurring in $h(\phi(\bar{x},\bar{y}))$. We now define an extension h' of h such that $h'(\psi(\bar{x},\bar{z})) \subseteq \mathsf{facts}(I)$, as needed. To this end, for each $K \subseteq h(\phi(\bar{x},\bar{y}))$, with $|\mathsf{facts}(K)| \leq 1$, let L_K be the S-instance such that $\mathsf{dom}(L_K) = \mathsf{adom}(L_K)$ and $\mathsf{facts}(L_K) = \mathsf{facts}(L) \cap \mathsf{facts}(J_K)$. Clearly, $\mathsf{adom}(L_K)$ contains at most m constants not occurring in $\mathsf{adom}(K)$, which in turn implies that L_K is in the m-neighbourhood of K in J_K. Thus, there is a function $f_K : \mathsf{adom}(L_K) \to \mathsf{adom}(I)$, which is the identity on $\mathsf{adom}(K)$, such that $f_K(\mathsf{facts}(L_K)) \subseteq \mathsf{facts}(I)$. Let f be the binary relation

$$\{(c,c) \mid c \text{ occurs in } h(\phi(\bar{x},\bar{y}))\} \cup \{(c, f_K(c)) \mid c \in \mathsf{adom}(L_K)\}.$$

It is not difficult to show that f is a total function from $\mathsf{adom}(L)$ to $\mathsf{adom}(I)$ such that $f(\mathsf{facts}(L)) \subseteq \mathsf{facts}(I)$. To conclude our proof, consider the mapping $h' = f \circ g$. For each variable $x \in \bar{x}$, $h'(x)$ is equal to $f(g(x)) = g(x) = h(x)$, and hence, h' is an extension of h. Moreover, for each atom α occurring in $\psi(\bar{x},\bar{z})$, $g(\alpha) \in \mathsf{facts}(L)$, and therefore, $h'(\alpha) \in \mathsf{facts}(I)$. Consequently, h' is an extension of h such that $h'(\psi(\bar{x},\bar{z})) \subseteq \mathsf{facts}(I)$, and the claim follows.

Direction (3) ⇒ (2)

We start by defining the set of linear tgds

$$\Sigma_L \;=\; \{\sigma \in \mathsf{LTGD}_{n,m} \mid \Sigma \models \sigma\}.$$

Our goal is to establish that $\Sigma \equiv \Sigma_L$, which implies that O is an $\mathsf{LTGD}_{n,m}$-ontology. The fact that $\Sigma \models \Sigma_L$ holds trivially. We proceed to show the non-trivial direction. Consider an S-instance I such that $I \not\models \Sigma$, and assume, by contradiction, that $I \models \Sigma_L$. Recall that $\Phi^I_{K,m}(\bar{x})$, for an S-instance $K \subseteq I$, is obtained from the m-diagram of K relative to I by renaming each constant $c \in \mathsf{dom}(K)$ to a new variable x_c. We show the following auxiliary claim.

Claim C.1. *There exists an* S*-instance $K \subseteq I$, with $\mathsf{dom}(K) = \mathsf{adom}(K)$, $|\mathsf{adom}(K)| \leq n$, and $|\mathsf{facts}(K)| \leq 1$, such that, for each $J \in O$, it holds that $J \models \neg \exists \bar{x}\, \Phi^I_{K,m}(\bar{x})$.*

Proof. Towards a contradiction, assume that, for every $K \subseteq I$ with $\mathsf{dom}(K) = \mathsf{adom}(K)$, $|\mathsf{adom}(K)| \leq n$, and $|\mathsf{facts}(K)| \leq 1$, there exists $J \in O$ such that $J \models \exists \bar{x}\, \Phi^I_{K,m}(\bar{x})$. We proceed to show that in this case O is linearly (n,m)-locally embeddable in I, which in turn implies that $I \in O$ since, by hypothesis, O is linearly (n,m)-local. But this contradicts the fact that $I \not\models \Sigma$ (and thus, $I \notin O$).

Consider an arbitrary S-instance $K \subseteq I$ with $|\text{adom}(K)| \leq n$ and $|\text{facts}(K)| \leq 1$, and assume that $J \in O$ is the instance such that $J \models \exists \bar{x}\, \Phi^I_{K,m}(\bar{x})$. We first observe that $J \models \exists \bar{x}\, \Phi^I_{K,m}(\bar{x})$ implies the existence of an instance $J_K \subseteq J$ such that $K \simeq J_K$. We can therefore assume, w.l.o.g., that $K \subseteq J$. To show that O is linearly (n,m)-locally embeddable in I, it suffices to show that for every S-instance J' in the m-neighbourhood of K in J, there exists a function $h : \text{adom}(J') \to \text{adom}(I)$, which is the identity on $\text{adom}(K)$, such that $h(\text{facts}(J')) \subseteq \text{facts}(I)$. This is because, for every $K' \subseteq I$ with $\text{facts}(K') = \text{facts}(K)$ and $K' \subseteq J$, the m-neighbourhood of K in J coincides with the m-neighbourhood of K' in J. By contradiction, assume that there is J' in the m-neighbourhood of K in J for which there is no function $h : \text{adom}(J') \to \text{adom}(I)$ that is the identity on $\text{adom}(K)$ with $h(\text{facts}(J')) \subseteq \text{facts}(I)$. Let L be the S-instance defined as the difference between J' and K, i.e., L is such that $\text{facts}(L) = \text{facts}(J') \setminus \text{facts}(K)$, while $\text{dom}(L)$ consists of all the constants occurring in $\text{facts}(J') \setminus \text{facts}(K)$, i.e., $\text{dom}(L) = \text{adom}(L)$. Clearly, there is no function $h : \text{adom}(L) \to \text{adom}(I)$ that is the identity on $\text{adom}(K)$ such that $h(\text{facts}(L)) \subseteq I$. Observe that $|\text{adom}(L) \setminus \text{adom}(K)| \leq m$; we assume that $\text{adom}(L) \setminus \text{adom}(K) = \{d_1, \ldots, d_{m'}\}$ for $m' \leq m$. Let $\gamma(\bar{y})$ be the formula obtained from $\bigwedge_{\alpha \in \text{facts}(L)} \alpha$ after renaming each constant d_i to the variable $\star_i$; clearly, $\bar{y} = \star_1, \ldots, \star_{m'}$. Since there is no function $h : \text{adom}(L) \to \text{adom}(I)$ that is the identity on $\text{adom}(K)$ such that $h(\text{facts}(L)) \subseteq I$, we can conclude that $I \not\models \exists \bar{y}\, \gamma(\bar{y})$. Observe now that, by construction, $\neg\exists \bar{y}\, \gamma(\bar{y})$ is a conjunct of $\Delta^I_{K,m}$. The formula $\neg\exists \bar{z}\, \gamma(\bar{z})$ obtained from $\neg\exists \bar{y}\, \gamma(\bar{y})$ after renaming each constant $c \in \text{adom}(K)$ to the variable x_c is a conjunct of $\Phi^I_{K,m}(\bar{x})$. Since $L \subseteq J$, we conclude that $J \models \exists \bar{z}\, \gamma(\bar{z})$, which in turn implies that $J \not\models \exists \bar{x}\, \Phi^I_{K,m}(\bar{x})$. But this contradicts the fact that $J \models \exists \bar{x}\, \Phi^I_{K,m}(\bar{x})$, and the claim follows. □

Recall that $\mathsf{E}_{n,m}$ is the set that collects all the edds over S of the form $\forall \bar{x}(\phi(\bar{x}) \to \bigvee_{i=1}^k \psi_i(\bar{x}_i))$ such that $\bar{x}$ consists of at most n distinct variables, and, for each $i \in [k]$, $\psi_i(\bar{x}_i)$ mentions at most m existentially quantified variables. Let K be the S-instance provided by Claim C.1. From Claim 4.6, we know that $\neg\Phi^I_{K,m}(\bar{x})$ is equivalent to an edd $\delta \in \mathsf{E}_{n,m}$ of the form

$$\forall \bar{x} \left(\phi(\bar{x}) \to \bigvee_{i}^{k} \psi_i(\bar{x}_i) \right).$$

Since $|\text{facts}(K)| \leq 1$, we further know that $\phi(\bar{x})$ consists of a single atom. Observe that $I \models \exists \bar{x}\, \Phi^I_{K,m}(\bar{x})$, and thus, $I \not\models \delta$. Since $I \models \Sigma_L$, we can conclude that $\Sigma_L \not\models \delta$. Let $\{i_1, \ldots, i_\ell\}$ be the subset of $[k]$ such that $\psi_i(\bar{x})$ is a conjunction of atoms (i.e., are not equality expressions). In other words, every $\psi_j(\bar{x}_j)$ with $j \notin \{i_1, \ldots, i_\ell\}$ is an equality expression. Moreover, let δ_i denote the linear tgd

$$\forall \bar{x}\, (\phi(\bar{x}) \to \psi_i(\bar{x}_i)).$$

Since $\Sigma_L \not\models \delta$, we get that $\delta_i \notin \Sigma_L$, for each $i \in \{i_1, \ldots, i_\ell\}$. This implies that, for each $i \in \{i_1, \ldots, i_\ell\}$, $\Sigma \not\models \delta_i$. Let I_δ be the finite instance such that $\text{dom}(I_\delta) = \text{adom}(I_\delta)$ and $\text{facts}(I_\delta)$ is obtained by "freezing" $\phi(\bar{x})$, i.e., by replacing each variable in $\bar{x}$ with a distinct constant. We also write $\text{chase}(D_\delta, \Sigma)$ for the possibly infinite instance obtained by chasing $\text{facts}(I_\delta)$ using the tgds of Σ; we assume the reader is familiar with the chase procedure. By exploiting the fact that the chase builds universal instances, i.e., $\text{chase}(I_\delta, \Sigma)$ can be homomorphically mapped into every model M of Σ with $\text{facts}(I_\delta) \subseteq \text{facts}(M)$, we can show the following auxiliary claim:

Claim C.2. *For each* $i \in \{i_1, \ldots, i_\ell\}$, $\text{chase}(I_\delta, \Sigma) \not\models \delta_i$.

We can now conclude the proof. By Claim C.2, we have that $\text{chase}(I_\delta, \Sigma)$ does not satisfy δ_i, for each $i \in \{i_1, \ldots, i_\ell\}$. Furthermore, by construction, I_δ violates every equality expression in δ. This implies that $\text{chase}(D_\delta, \Sigma) \not\models \delta$. However, by construction, $\text{chase}(I_\delta, \Sigma) \models \Sigma$, and therefore, $\text{chase}(I_\delta, \Sigma) \in O$. But this contradicts Claim C.1, which states that, for every $J \in O$, $J \models \delta$.

D PROOFS FROM SECTION 7

We assume that O is over the schema $\mathsf{S} = \{R_1, \ldots, R_\ell\}$. Since O is a $\mathsf{TGD}_{n,m}$-ontology, there exists a set $\Sigma \in \mathsf{TGD}_{n,m}$ over S such that, for every S-instance I, $I \in O$ iff $I \models \Sigma$. We proceed to show the directions (1) ⇒ (3) and (3) ⇒ (2); (2) ⇒ (1) holds trivially.

Direction (1) ⇒ (3)

Consider an S-instance I. We need to show the following: if O is guardedly (n,m)-locally embeddable in I, then $I \in O$ (or, equivalently, $I \models \Sigma$), which in turn implies that O is guardedly (n,m)-local, as needed. Consider a tgd $\sigma \in \Sigma$ of the form $\phi(\bar{x}, \bar{y}) \to \exists \bar{z}\, \psi(\bar{x}, \bar{z})$ such that there is a function $h : \bar{x} \cup \bar{y} \to \text{dom}(I)$ with $h(\phi(\bar{x}, \bar{y})) \subseteq \text{facts}(I)$. We need to show that there exists an extension h' of h such that $h'(\psi(\bar{x}, \bar{z})) \subseteq \text{facts}(I)$. Observe that at most n distinct variables occur in $\bar{x} \cup \bar{y}$, which implies that each fact $\alpha \in h(\phi(\bar{x}, \bar{y}))$ contains at most n distinct constants from $\text{adom}(I)$. Assume that $h(\phi(\bar{x}, \bar{y}))$ contains $d \geq 0$ distinct facts $\{\alpha_1, \ldots, \alpha_d\}$, then there exist S-instances $K_1, \ldots, K_d$ such that, for each $i \in [d]$, K_i is guarded, $K_i \preceq I$, and $h(\phi(\bar{x}, \bar{y})) \subseteq \bigcup_{i=1}^k \text{facts}(K_i)$. Since, by hypothesis, O is guardedly (n,m)-locally embeddable in I, for each $i \in [d]$, there exists $J_i \in O$ such that $K_i \subseteq J_i$ and, for each J' in the m-neighbourhood of K_i in J_i, there exists a function $f : \text{adom}(J') \to \text{adom}(I)$, which is the identity on $\text{adom}(K_i)$, such that $f(\text{facts}(J')) \subseteq \text{facts}(I)$. Since $h(\phi(\bar{x}, \bar{y}))$ consists of finitely many atoms, and O is closed under isomorphisms, we can assume that $\text{dom}(J_i) \cap \text{dom}(J_j) = \text{dom}(K_i) \cap \text{dom}(K_j)$, for each $i, j \in [d]$. In other words, we can assume that, for each pair $i, j \in [d]$, the constants of $\text{dom}(J_i)$ not occurring in K_i do not occur in $\text{dom}(J_j)$ either. Let J be an S-instance such that

$$\text{facts}(J) \;=\; \bigcup_{i=1}^{d} \text{facts}(J_i).$$

We now prove an auxiliary claim:

Claim D.1. *It holds that* $J \in O$.

Proof. Since, by hypothesis, O is a GTGD-ontology, there exists $\Sigma_G \in \text{GTGD}$ such that $I \in O$ iff $I \models \Sigma_G$. Let $\delta \in \Sigma_G$, and γ be the guard of $\text{body}(\delta)$. Consider a function λ such that $\lambda(\text{body}(\delta)) \subseteq \text{facts}(J)$. We will prove that $J \models \delta$, in turn proving $J \in O$. Since $\lambda(\gamma) \in J$, there exists $i \in [d]$ such that $\lambda(\gamma) \in J_i$. We proceed to show that $\lambda(\text{body}(\delta)) \subseteq \text{facts}(J_i)$. In turn, this will prove that $J \models \delta$, since $J_i \in O$ and then $J_i \models \delta$, which, in turn, implies that there exists an extension $\lambda' \supseteq \lambda$ such that $\lambda'(\text{head}(\delta)) \subseteq \text{facts}(J_i) \subseteq \text{facts}(J)$.

Assume that the constants occurring in $\lambda(\gamma)$ are the set $G \cup F$ such that $G \subseteq \text{adom}(K_i)$, and $F \subseteq \text{adom}(J_i) \setminus \text{adom}(K_i)$. In other words, G collects all the constants in $\lambda(\gamma)$ that occur in $\text{adom}(K_i)$, and F collects all the constants in $\lambda(\gamma)$ that occur in $\text{adom}(J_i)$ but not in $\text{adom}(K_i)$. By construction, no constant of F occurs in J_j with $i \neq j$, and hence, every atom in $\lambda(\text{body}(\delta))$ that mention constants form F occurs in J_i by definition. Consider an atom $\alpha \in \lambda(\text{body}(\delta))$ that only mentions constants from G. In turn, this means that α only mentions constants occurring in the guard atom of K_i, and therefore $\alpha \in K_i \subseteq J_i$. Since atoms occurring in $\lambda(\text{body}(\delta))$ only mentions constants from $G \cup F$, these observations imply that $\lambda(\text{body}(\delta)) \subseteq \text{facts}(J_i)$. □

From the definition of J, we get that $h(\phi(\bar{x}, \bar{y})) \subseteq J$. Since $J \models \sigma$, there exists an extension g of h such that $g(\psi(\bar{x}, \bar{z})) \subseteq J$. Let $L = (\text{dom}(L), R_1^L, \ldots, R_\ell^L)$ with $\text{dom}(L) = \text{dom}(g(\phi(\bar{x}, \bar{y}))) \cup \text{dom}(g(\psi(\bar{x}, \bar{z})))$, and for $i \in [\ell]$, $R_i^L = R_i^J{}_{|\text{dom}(L)}$. Since, σ mentions at most m existentially quantified variables, we have that $g(\psi(\bar{x}, \bar{z}))$ mentions at most m constants not occurring in $h(\phi(\bar{x}, \bar{y}))$. This implies that $\text{adom}(L)$ contains at most m constants not occurring in $h(\phi(\bar{x}, \bar{y}))$. We now define an extension h' of h such that $h'(\psi(\bar{x}, \bar{z})) \subseteq \text{facts}(I)$, as needed. To this end, for each $i \in [d]$, let L_i be the S-instance such that $\text{dom}(L_i) = \text{adom}(L_i)$ and $\text{facts}(L_i) = \text{facts}(L) \cap \text{facts}(J_i)$. Clearly, $\text{adom}(L_i)$ contains at most m constants not occurring in $\text{adom}(K_i)$, which in turn implies that L_i is in the m-neighbourhood of K_i in J_I. Thus, there is a function $f_i : \text{adom}(L_i) \to \text{adom}(I)$, which is the identity on $\text{adom}(K_i)$, such that $f_i(\text{facts}(L_i)) \subseteq \text{facts}(I)$. Let f be the binary relation

$$\{(c, c) \mid c \text{ occurs in } h(\phi(\bar{x}, \bar{y}))\} \cup \{(c, f_i(c)) \mid c \in \text{adom}(L_i)\}.$$

It is not difficult to show that f is a total function from $\text{adom}(L)$ to $\text{adom}(I)$ such that $f(\text{facts}(L)) \subseteq \text{facts}(I)$. To conclude our proof, consider the mapping $h' = f \circ g$. For each variable $x \in \bar{x}$, $h'(x)$ is equal to $f(g(x)) = g(x) = h(x)$, and hence, h' is an extension of h. Moreover, for each atom α occurring in $\psi(\bar{x}, \bar{z})$, $g(\alpha) \in \text{facts}(L)$, and therefore, $h'(\alpha) \in \text{facts}(I)$. Consequently, h' is an extension of h such that $h'(\psi(\bar{x}, \bar{z})) \subseteq \text{facts}(I)$, and the claim follows.

Direction (3) ⇒ (2)

We start by defining the set of guarded tgds

$$\Sigma_G \; = \; \{\sigma \in \text{GTGD}_{n,m} \mid \Sigma \models \sigma\}.$$

Our goal is to establish that $\Sigma \equiv \Sigma_G$, which implies that O is an $\text{GTGD}_{n,m}$-ontology. The fact that $\Sigma \models \Sigma_G$ holds trivially. We proceed to show the non-trivial direction. Consider an S-instance I such that $I \not\models \Sigma$, and assume, by contradiction, that $I \models \Sigma_G$. Since O is domain independent, we can assume that $\text{dom}(I) = \text{adom}(I)$. Recall that $\Phi^I_{K,m}(\bar{x})$, for an S-instance $K \subseteq I$, is obtained from the m-diagram of K relative to I by renaming each constant $c \in \text{dom}(K)$ to a new variable x_c. We show the following auxiliary claim.

Claim D.2. *There is a guarded S-instance $K \leq I$ with $|\text{adom}(K)| \leq n$ such that, for each $J \in O$, it holds that $J \models \neg\exists\bar{x}\, \Phi^I_{K,m}(\bar{x})$.*

Proof. Towards a contradiction, assume that, for every guarded instance $K \leq I$ with $|\text{adom}(K)| \leq n$, there exists $J \in O$ such that $J \models \exists\bar{x}\, \Phi^I_{K,m}(\bar{x})$.We proceed to show that in this case O is guardedly (n, m)-locally embeddable in I, which in turn implies that $I \in O$ since, by hypothesis, O is guardedly (n, m)-local. But this contradicts the fact that $I \not\models \Sigma$ (and thus, $I \notin O$).

Consider an arbitrary guarded S-instance $K \leq I$ with $|\text{adom}(K)| \leq n$ and assume that $J \in O$ is the instance such that $J \models \exists\bar{x}\, \Phi^I_{K,m}(\bar{x})$. We first observe that $J \models \exists\bar{x}\, \Phi^I_{K,m}(\bar{x})$ implies the existence of an instance $J_K \subseteq J$ such that $K \simeq J_K$. We can therefore assume, w.l.o.g., that $K \subseteq J$. To show that O is guardedly (n, m)-locally embeddable in I, it suffices to show that for every S-instance J' in the m-neighbourhood of K in J, there exists a function $h : \text{adom}(J') \to \text{adom}(I)$, which is the identity on $\text{adom}(K)$, such that $h(\text{facts}(J')) \subseteq \text{facts}(I)$. By contradiction, assume that there is J' in the m-neighbourhood of K in J for which there is no function $h : \text{adom}(J') \to \text{adom}(I)$ that is the identity on $\text{adom}(K)$ with $h(\text{facts}(J')) \subseteq \text{facts}(I)$. Let L be the S-instance defined as the difference between J' and K, i.e., L is such that $\text{facts}(L) = \text{facts}(J') \setminus \text{facts}(K)$, while $\text{dom}(L)$ consists of all the constants occurring in $\text{facts}(J') \setminus \text{facts}(K)$, i.e., $\text{dom}(L) = \text{adom}(L)$. Clearly, there is no function $h : \text{adom}(L) \to \text{adom}(I)$ that is the identity on $\text{adom}(K)$ such that $h(\text{facts}(L)) \subseteq I$. Observe that $|\text{adom}(L) \setminus \text{adom}(K)| \leq m$; we assume that $\text{adom}(L) \setminus \text{adom}(K) = \{d_1, \ldots, d_{m'}\}$ for $m' \leq m$. Let $\gamma(\bar{y})$ be the formula obtained from $\bigwedge_{\alpha \in \text{facts}(L)} \alpha$ after renaming each constant d_i to the variable $\star_i$; clearly, $\bar{y} = \star_1, \ldots, \star_{m'}$. Since there is no function $h : \text{adom}(L) \to \text{adom}(I)$ that is the identity on $\text{adom}(K)$ such that $h(\text{facts}(L)) \subseteq I$, we can conclude that $I \not\models \exists\bar{y}\, \gamma(\bar{y})$. Observe now that, by construction, $\neg\exists\bar{y}\, \gamma(\bar{y})$ is a conjunct of $\Delta^I_{K,m}$. The formula $\neg\exists\bar{z}\, \gamma(\bar{z})$ obtained from $\neg\exists\bar{y}\, \gamma(\bar{y})$ after renaming each constant $c \in \text{adom}(K)$ to the variable x_c is a conjunct of $\Phi^I_{K,m}(\bar{x})$. Since $L \subseteq J$, we conclude that $J \models \exists\bar{z}\, \gamma(\bar{z})$, which in turn implies that $J \not\models \exists\bar{x}\, \Phi^I_{K,m}(\bar{x})$. But this contradicts the fact that $J \models \exists\bar{x}\, \Phi^I_{K,m}(\bar{x})$, and the claim follows. □

Recall that $\mathsf{E}_{n,m}$ is the set that collects all the edds over S of the form $\forall\bar{x}(\phi(\bar{x}) \to \bigvee_{i=1}^k \psi_i(\bar{x}_i))$ such that $\bar{x}$ consists of at most n distinct variables, and, for each $i \in [k]$, $\psi_i(\bar{x}_i)$ mentions at most m existentially quantified variables. Let K be the S-instance provided by Claim D.2. From Claim 4.6, we know that $\neg\Phi^I_{K,m}(\bar{x})$ is equivalent to an edd $\delta \in \mathsf{E}_{n,m}$ of the form

$$\forall\bar{x}\left(\phi(\bar{x}) \to \bigvee_{i}^{k} \psi_i(\bar{x}_i)\right).$$

Since K is guarded, we further know that $\phi(\bar{x})$ is guarded as well. Observe that $I \models \exists\bar{x}\, \Phi^I_{K,m}(\bar{x})$, and thus, $I \not\models \delta$. Since $I \models \Sigma_G$, we can conclude that $\Sigma_G \not\models \delta$. Let $\{i_1, \ldots, i_\ell\}$ be the subset of $[k]$ such that $\psi_i(\bar{x})$ is a conjunction of atoms (i.e., are not equality expressions). In other words, every $\psi_j(\bar{x}_j)$ with $j \notin \{i_1, \ldots, i_\ell\}$ is an equality expression. Moreover, let δ_i denote the guarded tgd

$$\forall\bar{x}\,(\phi(\bar{x}) \to \psi_i(\bar{x}_i)).$$

Since $\Sigma_G \not\models \delta$, we get that $\delta_i \notin \Sigma_G$, for each $i \in \{i_1, \ldots, i_\ell\}$. This implies that, for each $i \in \{i_1, \ldots, i_\ell\}$, $\Sigma \not\models \delta_i$. Let I_δ be the finite instance such that $\text{dom}(I_\delta) = \text{adom}(I_\delta)$ and $\text{facts}(I_\delta)$ is obtained by "freezing" $\phi(\bar{x})$, i.e., by replacing each variable in $\bar{x}$ with a distinct constant. We also write $\text{chase}(D_\delta, \Sigma)$ for the possibly infinite instance obtained by chasing $\text{facts}(I_\delta)$ using the tgds of Σ; we assume the reader is familiar with the chase procedure. By exploiting the fact that the chase builds universal instances, i.e., $\text{chase}(I_\delta, \Sigma)$

can be homomorphically mapped into every model M of Σ with $\mathsf{facts}(I_\delta) \subseteq \mathsf{facts}(M)$, we can show the following auxiliary claim:

Claim D.3. *For each $i \in \{i_1, \ldots, i_\ell\}$, $\mathsf{chase}(I_\delta, \Sigma) \not\models \delta_i$.*

We can now conclude the proof. By Claim D.3, we have that $\mathsf{chase}(I_\delta, \Sigma)$ does not satisfy δ_i, for each $i \in \{i_1, \ldots, i_\ell\}$. Furthermore, I_δ violates every equality expression in δ since, by construction, each variables in $\mathsf{body}(\delta)$ is replaced with a distinct constant. This implies that $\mathsf{chase}(D_\delta, \Sigma) \not\models \delta$. However, by construction, $\mathsf{chase}(I_\delta, \Sigma) \models \Sigma$, and therefore, $\mathsf{chase}(I_\delta, \Sigma) \in O$. But this contradicts Claim D.2, which states that, for every $J \in O$, $J \models \delta$.

E PROOFS FROM SECTION 8

E.1 Proof of Lemma 8.3

Direction (1) ⇒ (2)

Consider an $\mathsf{FGTGD}_{n,m}$-ontology O over a schema $\mathbf{S}$. By definition, there exists a set $\Sigma \in \mathsf{FGTGD}_{n,m}$ such that, for every $\mathbf{S}$-instance I, $I \in O$ iff $I \models \Sigma$. Consider an $\mathbf{S}$-instance I, and assume that O is fr-guardedly (n,m)-locally embeddable in I. We proceed to show that $I \in O$, or, equivalently, $I \models \Sigma$. Consider a tgd $\sigma \in \Sigma$ of the form $\phi(\bar{x}, \bar{y}) \rightarrow \exists \bar{z}\, \psi(\bar{x}, \bar{z})$, where $\phi(\bar{x}, \bar{y})$ contains an atom $R(\bar{w})$ with $\bar{x} \subseteq \bar{w}$, i.e., $R(\bar{w})$ is considered as the frontier-guard of σ. Assume that there exists a function $h : \bar{x} \cup \bar{y} \rightarrow \mathsf{dom}(I)$ such that $h(\phi(\bar{x}, \bar{y})) \subseteq \mathsf{facts}(I)$. We show that there exists an extension λ of h such that $\lambda(\psi(\bar{x}, \bar{z})) \subseteq \mathsf{facts}(I)$. Let $K = (\mathsf{dom}(K), R_1^K, \ldots, R_\ell^K)$ where $\mathsf{dom}(K)$ is the set of constants occurring in $h(\phi(\bar{x}, \bar{y}))$, and, for each $i \in [\ell]$, $R_i^K = R^I_{i\,|K}$, and let F be the set of constants occurring in $h(\bar{x})$. Since $h(R(\bar{w})) \in K$, it is clear that K is an F-guarded subinstance of I. Moreover, $|\mathsf{adom}(K)| \leq n$ since $\phi(\bar{x}, \bar{y})$ mentions at most n distinct variables. Since, by hypothesis, O is fr-guardedly (n,m)-locally embeddable in I, we conclude that there exists $J_K \in O$ such that $K \subseteq J_K$, and, for every J' in the m-neighbourhood of F in J_K, there is a function $\mu_{J'} : \mathsf{adom}(J') \rightarrow \mathsf{adom}(I)$, which is the identity on F, such that $\mu_{J'}(\mathsf{facts}(J')) \subseteq \mathsf{facts}(I)$. It is clear that $h(\phi(\bar{x}, \bar{y})) \subseteq \mathsf{facts}(J_K)$. Since $J_K \in O$, or, equivalently, $J_K \models \Sigma$, there exists an extension g of h such that $g(\psi(\bar{x}, \bar{z})) \subseteq \mathsf{facts}(J_K)$. Let $L = (\mathsf{dom}(L), R_1^L, \ldots, R_\ell^L)$ where $\mathsf{dom}(L)$ are the constants occurring in $g(\psi(\bar{x}, \bar{z}))$, and, for each $i \in [\ell]$, $R_i^L = R_i^{J_K}{}_{|\mathsf{dom}(L)}$. It is clear that L is in the m-neighbourhood of F in J_K since $\bar{z}$ has at most m variables. Therefore, there is a function $\mu_L : \mathsf{adom}(L) \rightarrow \mathsf{adom}(I)$, which is the identity on F, such that $\mu_L(\mathsf{facts}(L)) \subseteq \mathsf{facts}(I)$. Consider the function $\lambda : \bar{x} \cup \bar{y} \cup \bar{z} \rightarrow \mathsf{adom}(I)$ defined as follows: $\lambda(v) = h(v)$, for each $v \in \bar{x} \cup \bar{y}$, $\lambda(z) = \mu_L(g(z))$, otherwise. Clearly, λ is an extension of h. Moreover, observe that g is an extension of h, and μ_L is the identity over $F = h(\bar{x})$. Therefore, we have that $\lambda(\psi(\bar{x}, \bar{y})) = \mu_L(g(\psi(\bar{x}, \bar{y})))$. Since $g(\psi(\bar{x}, \bar{z})) \subseteq \mathsf{facts}(L)$, we get that $\lambda(\psi(\bar{x}, \bar{z})) \subseteq \mathsf{facts}(I)$, which in turn implies that $I \models \sigma$.

Direction (2) ⇒ (1)

We assume that O is over the schema $\mathbf{S} = \{R_1, \ldots, R_\ell\}$. Since O is a $\mathsf{TGD}_{n,m}$-ontology, there is $\Sigma \in \mathsf{TGD}_{n,m}$ over $\mathbf{S}$ such that, for every $\mathbf{S}$-instance I, $I \in O$ iff $I \models \Sigma$. We define the set of tgds

$$\Sigma_F \;=\; \{\sigma \in \mathsf{FGTGD}_{n,m} \mid \Sigma \models \sigma\}.$$

Our goal is to establish that $\Sigma \equiv \Sigma_F$, which implies that O is an $\mathsf{FGTGD}_{n,m}$-ontology. The fact that $\Sigma \models \Sigma_F$ holds trivially. We proceed to show the non-trivial direction. Consider an $\mathbf{S}$-instance I such that $I \not\models \Sigma$, and assume, by contradiction, that $I \models \Sigma_F$. Since O is domain independent, we can assume that $\mathsf{dom}(I) = \mathsf{adom}(I)$. Assume now a finite $F \subseteq \mathsf{dom}(I)$, an F-guarded subinstance $K \preceq I$, and let $K_F = (\mathsf{dom}(L), R_1^{K_F}, \ldots, R_\ell^{K_F})$ where $\mathsf{dom}(K_F) = F$, and, for each $i \in [\ell]$, $R_i^{K_F} = R^K_{i\,|F}$. We define the formula $\Delta^I_{K,m,F}$ as

$$\bigwedge_{\alpha \in K} \alpha \;\wedge \bigwedge_{\substack{c,d \in \mathsf{dom}(K),\\ c \neq d}} \neg(c = d) \;\wedge\; \Delta^I_{K_F,m}$$

where $\Delta^I_{K_F,m}$ is the m-diagram of K_F relative to I. Intuitively, $\Delta^I_{K,m,F}$ is obtained from the m-diagram of K relative to I ($\Delta^I_{K,m}$) by removing all the conjuncts $\neg\exists\bar{y}\gamma(\bar{y})$ that mention constants outside F. Finally, we build the formula $\Phi^I_{K,m,F}(\bar{x})$ from $\Delta^I_{K,m,F}$ by replacing each constant element $c \in \mathsf{dom}(K)$ with a new variable $x_c \in \mathbf{V} \setminus \{\star_1, \ldots, \star_\ell\}$. We proceed to show the following claim.

Claim E.1. *There exists a finite $F \subseteq \mathsf{dom}(I)$ and an F-guarded subinstance $K \preceq I$ with $|\mathsf{adom}(K)| \leq n$ such that, for each $J \in O$, it holds that $J \models \neg\exists\bar{x}\, \Phi^I_{K,m,F}(\bar{x})$.*

Proof. Towards a contradiction, assume that, for every finite $F \subseteq \mathsf{dom}(I)$ and F-guarded subinstance $K \preceq I$ with $|\mathsf{adom}(K)| \leq n$, there exists $J \in O$ such that $J \models \exists\bar{x}\, \Phi^I_{K,m,F}(\bar{x})$. We proceed to show that in this case O is fr-guardedly (n,m)-locally embeddable in I, which in turn implies that $I \in O$ since, by hypothesis, O is guardedly (n,m)-local. But this contradicts the fact that $I \not\models \Sigma$ (and thus, $I \notin O$). Consider an arbitrary finite $F \subseteq \mathsf{dom}(I)$ and F-guarded subinstance $K \preceq I$ with $|\mathsf{adom}(K)| \leq n$ and assume that $J \in O$ is the instance such that $J \models \exists\bar{x}\, \Phi^I_{K,m,F}(\bar{x})$. We first observe that $J \models \exists\bar{x}\, \Phi^I_{K,m,F}(\bar{x})$ implies the existence of an instance $J_K \subseteq J$ such that $K \simeq J_K$. We can therefore assume, w.l.o.g., that $K \subseteq J$. To show that O is fr-guardedly (n,m)-locally embeddable in I, it suffices to show that for every $\mathbf{S}$-instance J' in the m-neighbourhood of F in J, there exists a function $h : \mathsf{adom}(J') \rightarrow \mathsf{adom}(I)$, which is the identity on F, such that $h(\mathsf{facts}(J')) \subseteq \mathsf{facts}(I)$. By contradiction, assume that there is J' in the m-neighbourhood of F in J for which there is no function $h : \mathsf{adom}(J') \rightarrow \mathsf{adom}(I)$ that is the identity on F with $h(\mathsf{facts}(J')) \subseteq \mathsf{facts}(I)$. Let L be the $\mathbf{S}$-instance defined as the difference between J' and K, i.e., L is such that $\mathsf{facts}(L) = \mathsf{facts}(J') \setminus \mathsf{facts}(K)$, while $\mathsf{dom}(L)$ consists of all the constants occurring in $\mathsf{facts}(J') \setminus \mathsf{facts}(K)$, i.e., $\mathsf{dom}(L) = \mathsf{adom}(L)$. Clearly, there is no function $h : \mathsf{adom}(L) \rightarrow \mathsf{adom}(I)$ that is the identity on $\mathsf{adom}(K)$ such that $h(\mathsf{facts}(L)) \subseteq I$. Observe that $|\mathsf{adom}(L) \setminus F| \leq m$; we assume that $\mathsf{adom}(L) \setminus F = \{d_1, \ldots, d_{m'}\}$ for $m' \leq m$. Let $\gamma(\bar{y})$ be the formula obtained from $\bigwedge_{\alpha \in \mathsf{facts}(L)} \alpha$ after renaming each constant d_i to the variable $\star_i$; clearly, $\bar{y} = \star_1, \ldots, \star_{m'}$. Since there is no function $h : \mathsf{adom}(L) \rightarrow \mathsf{adom}(I)$ that is the identity on F such that $h(\mathsf{facts}(L)) \subseteq I$, we can conclude that $I \not\models \exists\bar{y}\, \gamma(\bar{y})$. Observe now that, by construction, $\neg\exists\bar{y}\, \gamma(\bar{y})$ is a conjunct of $\Delta^I_{K,m,F}$. The formula $\neg\exists\bar{z}\, \gamma(\bar{z})$ obtained from $\neg\exists\bar{y}\, \gamma(\bar{y})$ after renaming each constant $c \in F$ to the variable x_c is a conjunct of $\Phi^I_{K,m,F}(\bar{x})$. Since $L \subseteq J$, we conclude that $J \models \exists\bar{z}\, \gamma(\bar{z})$, which in turn implies that $J \not\models \exists\bar{x}\, \Phi^I_{K,m,F}(\bar{x})$. But this contradicts the fact that $J \models \exists\bar{x}\, \Phi^I_{K,m,F}(\bar{x})$, and the claim follows. □

Recall that $\mathsf{E}_{n,m}$ is the set that collects all the edds over $\mathbf{S}$ of the form $\forall \bar{x}(\phi(\bar{x}) \rightarrow \bigvee_{i=1}^{k} \psi_i(\bar{x}_i))$ such that $\bar{x}$ consists of at most n distinct variables, and, for each $i \in [k]$, $\psi_i(\bar{x}_i)$ mentions at most m existentially quantified variables. Let K be the $\mathbf{S}$-instance provided by Claim E.1. From Claim 4.6, we know that $\neg\Phi^{I}_{K,m,F}(\bar{x})$ is equivalent to an edd $\delta \in \mathsf{E}_{n,m}$ of the form

$$\forall \bar{x} \left(\phi(\bar{x}) \rightarrow \bigvee_{i}^{k} \psi_i(\bar{x}_i) \right).$$

Let $\{i_1, \ldots, i_\ell\}$ be the subset of $[k]$ such that $\psi_i(\bar{x})$ is a conjunction of atoms (i.e., are not equality expressions). In other words, every $\psi_j(\bar{x}_j)$ with $j \notin \{i_1, \ldots, i_\ell\}$ is an equality expression. Since K is F-guarded, we further know that there exists an atom $\gamma(\bar{w})$ in $\phi(\bar{x})$ such that, for each $i \in \{i_1, \ldots, i_\ell\}$, $\bar{w}$ contains all the variables occurring in $\bar{x}_i$. Observe that $I \models \exists \bar{x}\, \Phi^{I}_{K,m_F}(\bar{x})$, and thus, $I \not\models \delta$. Since $I \models \Sigma_F$, we can conclude that $\Sigma_F \not\models \delta$. Moreover, for each $i \in \{i_1, \ldots, i_\ell\}$, let δ_i denote the frontier-guarded tgd

$$\forall \bar{x}\, (\phi(\bar{x}) \rightarrow \psi_i(\bar{x}_i))\,.$$

Since $\Sigma_F \not\models \delta$, we get that $\delta_i \notin \Sigma_F$, for each $i \in \{i_1, \ldots, i_\ell\}$. This implies that, for each $i \in \{i_1, \ldots, i_\ell\}$, $\Sigma \not\models \delta_i$. Let I_δ be the finite instance such that $\mathsf{dom}(I_\delta) = \mathsf{adom}(I_\delta)$ and $\mathsf{facts}(I_\delta)$ is obtained by "freezing" $\phi(\bar{x})$, i.e., by replacing each variable in $\bar{x}$ with a distinct constant. We also write $\mathsf{chase}(D_\delta, \Sigma)$ for the possibly infinite instance obtained by chasing $\mathsf{facts}(I_\delta)$ using the tgds of Σ; we assume the reader is familiar with the chase procedure. By exploiting the fact that the chase builds universal instances, i.e., $\mathsf{chase}(I_\delta, \Sigma)$ can be homomorphically mapped into every model M of Σ with $\mathsf{facts}(I_\delta) \subseteq \mathsf{facts}(M)$, we can show the following auxiliary claim:

Claim E.2. *For each $i \in \{i_1, \ldots, i_\ell\}$, $\mathsf{chase}(I_\delta, \Sigma) \not\models \delta_i$.*

We can now conclude the proof. By Claim E.2, we have that $\mathsf{chase}(I_\delta, \Sigma)$ does not satisfy δ_i, for each $i \in \{i_1, \ldots, i_\ell\}$. Furthermore, I_δ violates every equality expression in δ since, by construction, each variables in $\mathsf{body}(\delta)$ is replaced with a distinct constant. This implies that $\mathsf{chase}(D_\delta, \Sigma) \not\models \delta$. However, by construction, $\mathsf{chase}(I_\delta, \Sigma) \models \Sigma$, and therefore, $\mathsf{chase}(I_\delta, \Sigma) \in O$. But this contradicts Claim E.1, which states that, for every $J \in O$, $J \models \delta$.

F PROOFS FROM SECTION 9

Proof of Theorem 9.1 – Lower Bounds of Item (1)

We provide a reduction from the problem of atomic query answering under guarded tgds, which is 2ExpTime-hard in general, and ExpTime-hard for predicates of bounded arity [5]. In particular, given a set $\Sigma \in \mathsf{GTGD}$ over a schema $\mathbf{S}$, and an atomic query of the form $\exists \bar{x}\, Q(\bar{x})$, where $Q \in \mathbf{S}$, we are going to devise in polynomial time a set $\Sigma' \in \mathsf{GTGD}$ such that the following are equivalent:

1. $\Sigma \models \exists \bar{x}\, Q(\bar{x})$.
2. There exists a set $\Sigma_L \in \mathsf{LTGD}$ such that $\Sigma' \equiv \Sigma_L$.

The Construction of Σ'

The set Σ' is the union of two sets Σ'_1 and Σ'_2 defined as follows. The set Σ'_1 contains, for each tgd $\sigma \in \Sigma$ of the form $G(\bar{x}, \bar{y}), \phi(\bar{x}, \bar{y}) \rightarrow \exists \bar{z}\, \psi(\bar{x}, \bar{z})$, with $G(\bar{x}, \bar{y})$ being the guard atom, the tgd

$$\sigma_{\mathrm{Aux}} \;=\; G(\bar{x}, \bar{y}), \mathrm{Aux} \;\rightarrow\; \exists \bar{z}\, \psi(\bar{x}, \bar{z}),$$

where Aux is an auxiliary 0-ary predicate not occurring in $\mathbf{S}$. The set Σ'_2 consists of the tgds

$$\begin{aligned} \sigma_Q &= Q(\bar{x}) \;\rightarrow\; \mathrm{Aux} \\ \sigma_{R\mathrm{Aux}} &= R(x), \mathrm{Aux} \;\rightarrow\; T(x) \\ \sigma_{RS} &= R(x), S(x) \;\rightarrow\; T(x), \end{aligned}$$

where R, S, T are "fresh" unary predicates not occurring in $\mathbf{S}$. It is clear that Σ' is a set of guarded tgds that can be constructed in polynomial time. Notice also that the new predicates are of bounded arity, which means that if $\mathbf{S}$ consists of predicates of bounded arity, then the same holds for the schema $\mathbf{S} \cup \{\mathrm{Aux}, R, S, T\}$.

Correctness of the Reduction

We proceed to show that the above construction is indeed a reduction, i.e., we need to show that the statements (1) and (2) above are equivalent. We start with the direction (1) $\Rightarrow$ (2).

(1) $\Rightarrow$ (2). Consider the set $\Sigma_L \in \mathsf{LTGD}$ that consists of σ_Q and the following tgds: for each $\sigma_{\mathrm{Aux}} \in \Sigma'_1$ of the form $G(\bar{x}, \bar{y}), \mathrm{Aux} \rightarrow \exists \bar{z}\, \psi(\bar{x}, \bar{z})$, we have the linear tgd

$$G(\bar{x}, \bar{y}) \;\rightarrow\; \exists \bar{z}\, \psi(\bar{x}, \bar{z})$$

and we also have the linear tgd

$$R(x) \;\rightarrow\; T(x).$$

We claim that $\Sigma' \equiv \Sigma_L$.

($\Sigma' \models \Sigma_L$) Consider an instance I such that $I \models \Sigma'$, and assume that there exists a tgd $\sigma \in \Sigma_L$ of the form $\alpha(\bar{x}) \rightarrow \exists \bar{x}\, \psi(\bar{x}, \bar{z})$ and a function $h : \bar{x} \rightarrow \mathsf{dom}(I)$ such that $h(\alpha(\bar{x})) \in \mathsf{facts}(I)$. We need to show that there exists an extension h' of h such that $h(\psi(\bar{x}, \bar{z})) \subseteq \mathsf{facts}(I)$. By hypothesis, $\Sigma \models \exists \bar{x}\, Q(\bar{x})$. Observe also that $I \models \Sigma$, and thus, $I \models \exists \bar{x}\, Q(\bar{x})$. Thus, due to $\sigma_Q \in \Sigma'$, we have that $\mathrm{Aux} \in \mathsf{facts}(I)$. If $\alpha(\bar{x}) = Q(\bar{x})$ or $\alpha(\bar{x}) = R(\bar{x})$, then the desired extension h' of h exists. Assume now that $\alpha(\bar{x})$ is the body of some tgd of the form $G(\bar{x}, \bar{y}) \rightarrow \exists \bar{z}\, \psi(\bar{x}, \bar{z})$, i.e., a linear tgd obtained from some σ_{Aux} of the form $G(\bar{x}, \bar{y}), \mathrm{Aux} \;\rightarrow\; \exists \bar{z}\, \psi(\bar{x}, \bar{z})$ by dropping the relation Aux. Since $\mathrm{Aux} \in \mathsf{facts}(I)$, $h(\mathsf{body}(\sigma_{\mathrm{Aux}})) \subseteq \mathsf{facts}(I)$, and thus, the desired extension h' of h exists.

($\Sigma_L \models \Sigma'$) Consider now an instance I such that $I \models \Sigma_L$, and assume that there exists a tgd $\sigma \in \Sigma'$ of the form $\phi(\bar{x}, \bar{y}) \rightarrow \exists \bar{z}\, \psi(\bar{x}, \bar{z})$, and a function $h : \bar{x} \cup \bar{y} \rightarrow \mathsf{dom}(I)$ such that $h(\phi(\bar{x}, \bar{y})) \subseteq \mathsf{facts}(I)$. We need to show that there exists an extension h' of h such that $h'(\psi(\bar{x}, \bar{z})) \subseteq \mathsf{facts}(I)$. Assume first that $\sigma \in \Sigma'_1$. This implies that there exists in Σ_L a tgd of the form $G(\bar{x}, \bar{y}), \mathrm{Aux} \rightarrow \exists \bar{z}\, \psi(\bar{x}, \bar{z})$ with $G(\bar{x}, \bar{y})$ being the guard atom of σ. Clearly, $h(G(\bar{x}, \bar{y})) \in \mathsf{facts}(I)$, and the desired extension h' of h exists. Assume now that $\sigma \in \Sigma'_2$. If $\sigma = \sigma_Q$, then the existence of h' hold trivially since $\sigma_Q \in \Sigma'$. If σ is either $\sigma_{R\mathrm{Aux}}$ or σ_{RS}, then $h(R(x)) \in \mathsf{facts}(I)$, and thus, $h(T(x)) \in \mathsf{facts}(I)$, due to the tgd $R(x) \rightarrow T(x)$ in Σ_L.

This completes the proof of (1) $\Rightarrow$ (2).

(2) $\Rightarrow$ (1). Assume that $\Sigma \not\models \exists \bar{x}\, Q(\bar{x})$, and let I be a model of Σ such that $I \not\models \exists \bar{x}\, Q(\bar{x})$. We proceed to show that Σ is not closed under union, which in turn implies that Σ is not equivalent to a set of linear tgds. To this end, we define two instances J and J' as follows:

- $\mathsf{dom}(J) = \mathsf{adom}(J)$ and $\mathsf{facts}(J) = \mathsf{facts}(I) \cup \{R(c)\}$, and

- $\mathsf{dom}(J') = \mathsf{adom}(J')$ and $\mathsf{facts}(J') = \mathsf{facts}(I) \cup \{S(c)\}$,

where $c \notin \mathsf{adom}(I)$. It is easy to verify that $J \models \Sigma$ and $J' \models \Sigma$. Now, observe that the instance $J \cup J'$, which has domain $\mathsf{dom}(J) \cup \mathsf{dom}(J')$ and facts $\mathsf{facts}(J) \cup \mathsf{facts}(J')$, is not a model of Σ. This in turn implies that Σ is not closed under union, and the claim follows.

Proof of Theorem 9.2 – Lower Bound of Item (1)

We provide a reduction from the problem of atomic query answering under frontier-guarded tgds, which is 2ExpTime-hard, even in the case of bounded arity predicates [3]. In particular, given a set $\Sigma \in$ FGTGD over a schema $\mathbf{S}$, and an atomic query of the form $\exists \bar{x}\, Q(\bar{x})$, where $Q \in \mathbf{S}$, we are going to devise in polynomial time a set $\Sigma' \in$ FGTGD such that the following are equivalent:

(1) $\Sigma \models \exists \bar{x}\, Q(\bar{x})$.
(2) There exists a set $\Sigma_G \in$ GTGD such that $\Sigma' \equiv \Sigma_G$.

The Construction of Σ'

The construction of Σ' is similar in spirit to the one employed in the proof of the lower bounds of item (1) in Theorem 9.1. The set Σ' is the union of two sets Σ'_1 and Σ'_2 defined as follows. The set Σ'_1 contains, for each tgd $\sigma \in \Sigma$ of the form $G(\bar{x}, \bar{y}), \phi(\bar{x}, \bar{y}) \rightarrow \exists \bar{z}\, \psi(\bar{x}, \bar{z})$, with $G(\bar{x}, \bar{y})$ being the frontier-guard atom, the tgd

$$\sigma_{\mathrm{Aux}} \ = \ G(\bar{x}, \bar{y}), \mathrm{Aux} \ \rightarrow \ \exists \bar{z}\, \psi(\bar{x}, \bar{z}),$$

where Aux is an auxiliary 0-ary predicate not occurring in $\mathbf{S}$. The set Σ'_2 consists of the tgds

$$\begin{array}{rcl} \sigma_Q & = & Q(\bar{x}) \ \rightarrow \ \mathrm{Aux} \\ \sigma_{R\mathrm{Aux}} & = & R(x), \mathrm{Aux} \ \rightarrow \ T(x) \\ \sigma_{RS} & = & R(x), S(y) \ \rightarrow \ T(x), \end{array}$$

where R, S, T are “fresh” unary predicates not occurring in $\mathbf{S}$. It is clear that Σ' is a set of frontier-guarded tgds that can be constructed in polynomial time. Notice also that the new predicates are of bounded arity, which means that if $\mathbf{S}$ consists of predicates of bounded arity, then the same holds for the schema $\mathbf{S} \cup \{\mathrm{Aux}, R, S, T\}$.

Correctness of the Reduction

We proceed to show that the above construction is indeed a reduction, i.e., we need to show that the statements (1) and (2) above are equivalent. In fact, the direction (1) $\Rightarrow$ (2) can be shown using an argument similar to the one given in the proof of Theorem 9.1 for showing an analogous claim.

Let us now argue about the direction (2) $\Rightarrow$ (1). Assume that $\Sigma \not\models \exists \bar{x}\, Q(\bar{x})$, and let I be a model of Σ such that $I \not\models \exists \bar{x}\, Q(\bar{x})$. We proceed to show that Σ is not closed under disjoint union, which in turn allows us to conclude that Σ is not equivalent to a set of guarded tgds. Recall that Σ is closed under disjoint union if, for each pair of models K and K' of Σ such that $X = \mathsf{dom}(K) \cap \mathsf{dom}(K') \neq \emptyset$ implies $K_{|X} = K'_{|X}$, we have that $K \cup K'$ is a model of Σ. To this end, we consider the instances J and J' such that:

- $\mathsf{dom}(J) = \mathsf{adom}(J)$ and $\mathsf{facts}(J) = \mathsf{facts}(I) \cup \{R(c)\}$, and
- $\mathsf{dom}(J') = \mathsf{adom}(J')$ and $\mathsf{facts}(J') = \mathsf{facts}(I) \cup \{S(c)\}$,

where $c \notin \mathsf{adom}(I)$. It is easy to verify that $J \models \Sigma$ and $J' \models \Sigma$. Let K and K' be isomorphic copies of J and J', respectively, such that $\mathsf{dom}(K) \cap \mathsf{dom}(K') = \emptyset$. Observe that the instance $K \cup K'$ is not a model of Σ. This in turn implies that Σ is not closed under disjoint union, and the claim follows.