THE UNIVERSITY of EDINBURGH
informatics

University of Cyprus
Department of Computer Science

Rule-based Ontologies: From Semantics to Syntax

based on joint work with Marco Console and Phokion G. Kolaitis

Andreas Pieris

School of Informatics, University of Edinburgh

Department of Computer Science, University of Cyprus

Workshop on Formal Logic and Database Theory at Montpellier, February 11, 2025

Ontology Languages: Syntax and Semantics

Informally, an ontology language L has two main ingredients:

- Syntax: typically given by a context-free grammar that spells out the well-formed expressions of L – an ontology O is a finite set of expressions from L
- Semantics: typically given by an inductive definition that spells out what it means for a mathematical structure J to satisfy an expression ϕ of L, denoted $J \vDash \phi$. We further write $J \vDash O$ if, for every expression ϕ in O, $J \vDash \phi$

From Syntax to Semantics

Usually, the formal study of ontology languages proceeds from syntax to semantics:

- An ontology O from a language **L** gives rise to the collection of structures **C** = {J : J ⊨ O}, the models of O – we say that **C** is definable by O

- Main Question: what properties does a collection of structures definable by an ontology from the language **L** has?

From Semantics to Syntax

- **Main Goal:** Characterize definability by **L** in terms of model-theoretic properties

- This line of work was pioneered by Alfred Tarski (1901-1983), who was interested in characterizing notions of "metamathematical origin" in "purely mathematical terms"

Target Theorems: Let **C** be a collection of structures. The following are equivalent:

1. **C** is definable by **L** (i.e., there is an ontology O from **L** such that **C** = {J : J ⊨ O})
2. **C** enjoys certain model-theoretic properties

From Semantics to Syntax

- Main Goal: Characterize definability by L in terms of model-theoretic properties

- This line of work was pioneered by Alfred Tarski (1901-1983), who was interested in characterizing notions of "metamathematical origin" in "purely mathematical terms"

ESSLLI 2022 tutorial – When Semantics Meets Syntax

https://sites.google.com/ucsc.edu/esslli-2022/home

A Rule-based Ontology Language

tuple-generating dependencies (tgds)

$$\forall\bar{x}\forall\bar{y}\ (\phi(\bar{x},\bar{y}) \rightarrow \exists\bar{z}\ \psi(\bar{x},\bar{z}))$$

where $\phi(\bar{x},\bar{y})$ and $\psi(\bar{x},\bar{z})$ are conjunctions of relational atoms (or simply atoms)

$\phi(\bar{x},\bar{y})$ can be empty, $\psi(\bar{x},\bar{z})$ is non-empty

---

equality-generating dependencies (egds)

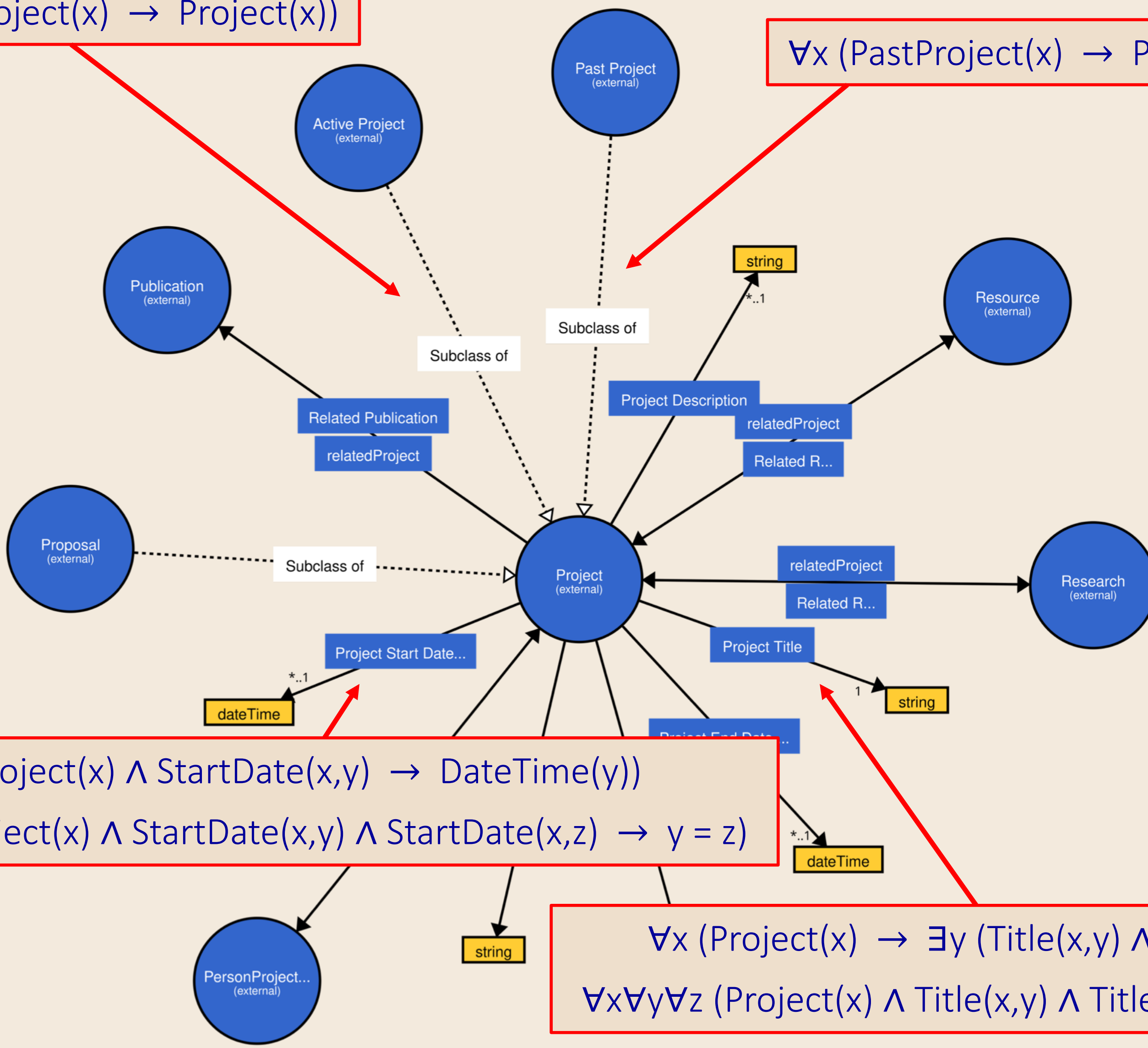
$$\forall\bar{x}\ (\phi(\bar{x}) \rightarrow x_i = x_j)$$

where $\phi(\bar{x})$ is a non-empty conjunction of atoms and x_i,x_j are variables from $\bar{x}$

The Project Ontology

$\forall x\ (ActiveProject(x) \rightarrow Project(x))$

$\forall x\ (PastProject(x) \rightarrow Project(x))$



$\forall x\ (Project(x) \wedge StartDate(x,y) \rightarrow DateTime(y))$

$\forall x \forall y \forall z\ (Project(x) \wedge StartDate(x,y) \wedge StartDate(x,z) \rightarrow y = z)$

$\forall x\ (Project(x) \rightarrow \exists y\ (Title(x,y) \wedge String(y))$

$\forall x \forall y \forall z\ (Project(x) \wedge Title(x,y) \wedge Title(x,z) \rightarrow y = z)$

The Project Ontology

a finite set of tuple- and equality-generating dependencies

$$\forall x\ (PastProject(x) \rightarrow Project(x))$$

$$\forall x\ (ActiveProject(x) \rightarrow Project(x))$$

$$\forall x\ (Project(x) \rightarrow \exists y\ (Title(x,y) \wedge String(y)))$$

$$\forall x \forall y \forall z\ (Project(x) \wedge Title(x,y) \wedge Title(x,z) \rightarrow y = z)$$

$$\forall x\ (Project(x) \wedge EndDate(x,y) \rightarrow DateTime(y))$$

$$\forall x \forall y \forall z\ (Project(x) \wedge EndDate(x,y) \wedge EndDate(x,z) \rightarrow y = z)$$

$$\vdots$$

Relational Structures

- A schema $\mathbf{S} = \{R_1,\ldots,R_n\}$ is a finite set of relation symbols with associated (positive) arity – we write $\text{arity}(R_i)$ for the arity of the relation symbol R_i

- A relational structure $\mathbf{J}$ over $\mathbf{S}$, or simply $\mathbf{S}$-structure, is a tuple $(\text{dom}(\mathbf{J}),R_1^{\mathbf{J}},\ldots,R_n^{\mathbf{J}})$, where $\text{dom}(\mathbf{J})$ is a (finite or infinite) domain and $R_1^{\mathbf{J}},\ldots,R_n^{\mathbf{J}}$ are relations over $\text{dom}(\mathbf{J})$, i.e., $R_i^{\mathbf{J}} \subseteq \text{dom}(\mathbf{J})^{\text{arity}(R_i)}$

- Given $\mathbf{S}$-structures $\mathbf{J}$ and $\mathbf{K}$, a homomorphism from $\mathbf{J}$ to $\mathbf{K}$ is a function $h : \text{dom}(\mathbf{J}) \rightarrow \text{dom}(\mathbf{K})$ such that, $(a_1,\ldots,a_m) \in R_i^{\mathbf{J}}$ implies $(h(a_1),\ldots,h(a_m)) \in R_i^{\mathbf{K}}$

- A conjunction of atoms is a structure – $R(x,y,z) \wedge R(z,z,y) \wedge P(x,w) \wedge T(x,z,w)$ is the structure $\mathbf{J} = (\{x,y,z,w\}, P^{\mathbf{J}}, R^{\mathbf{J}}, T^{\mathbf{J}})$, where $P^{\mathbf{J}} = \{(x,w)\}$, $R^{\mathbf{J}} = \{(x,y,z), (z,z,y)\}$, and $T^{\mathbf{J}} = \{(x,z,w)\}$

  $\Rightarrow$ we can talk about homomorphisms from conjunctions of atoms to structures

Semantics of Rule-based Ontologies

A structure **J** satisfies a tgd $\sigma = \forall\bar{x}\forall\bar{y}\,(\phi(\bar{x},\bar{y}) \rightarrow \exists\bar{z}\,\psi(\bar{x},\bar{z}))$, denoted $\mathbf{J} \models \sigma$, if, whenever there is a homomorphism h from $\phi(\bar{x},\bar{y})$ to **J**, then there exists a homomorphism from $\psi(\bar{x},\bar{z})$ to **J** that agrees with h on the variables of $\bar{x}$

A structure **J** satisfies an egd $\eta = \forall\bar{x}\,(\phi(\bar{x}) \rightarrow x_i = x_j)$, denoted $\mathbf{J} \models \eta$, if, whenever there is a homomorphism h from $\phi(\bar{x})$ to **J**, then $h(x_i) = h(x_j)$

From Semantics to Syntax

$$\begin{array}{c}\text{tgds + egds}\\ \uparrow\\ \text{tgds}\end{array}$$

Target Theorems: Let **C** be a collection of structures. The following are equivalent:

1. **C** is definable by tgds (+ egds)
2. **C** enjoys certain model-theoretic properties

An Early Result for Full TGDs + EGDs

Acta Informatica 23, 231–244 (1986)

Acta Informatica

On the Expressive Power of Data Dependencies

Johann A. Makowsky[1,3] and Moshe Y. Vardi[2,4]

[1] Dept. of Computer Science, Technion – Israel Institute of Technology, Haifa, Israel
[2] IBM Almaden Research Center, San Jose, CA95120, USA

Summary. The class of data dependencies is a class of first-order sentences that seem to be suitable to express semantic constraints for relational databases. We deal with the question of which classes of databases are *axiomatizable* by data dependencies. (A class Γ of databases is said to be axiomatizable by sentences of a certain kind if there exists a set of sentences of that kind such that Γ is the class of all models of that set.) Our results characterize, by algebraic closure conditions, classes of databases that are axiomatizable by dependencies of different kinds. Our technique is model-theoretic, and the characterization easily entails all previously known results on axiomatizability by dependencies.

An Early Result for Full TGDs + EGDs

A tgd is called full it it has no existentially quantified variables

$$\forall \bar{x} \forall \bar{y} \, (\phi(\bar{x},\bar{y}) \rightarrow \psi(\bar{x}))$$

Theorem ([Makowsky and Vardi, 1986]): Let **C** be a collection of structures. The following are equivalent:

1. **C** is definable by full tgds + egds
2. **C** is closed under isomorphisms, 1-critical, domain independent, modular, closed under substructures, and closed under direct products

Closure Under Isomorphisms

closed under isomorphisms, 1-critical, domain independent, closed under substructures, modular, closed under direct products

Let **J** and **K** be two structures:

- An isomorphism from **J** to **K** is an 1-1 homomorphism h from **J** to **K** such that the inverse of h is a homomorphism from **K** to **J**
- We say that **J** and **K** are isomorphic, denoted **J** $\simeq$ **K**, if there is an isomorphism from **J** to **K**

A collection **C** of structures is closed under isomorphisms if the following holds:

if **J** $\in$ **C** and **K** is a structure such that **J** $\simeq$ **K**, then **K** $\in$ **C**

1-criticality

closed under isomorphisms, 1-critical, domain independent, closed under substructures, modular, closed under direct products

A structure **J** = (dom(**J**),R_1^J,…,R_n^J) is 1-critical if the following holds:

- dom(**J**) consists of a single element, let say \$
- R_i^J = dom(**J**)$^{\text{arity}(R_i)}$ = {(\$,…,\$)}, for each i ∈ {1,…,n}

A collection **C** of structures is 1-critical if it contains an 1-critical structure

Domain Independence

closed under isomorphisms, 1-critical, domain independent, closed under substructures, modular, closed under direct products

Informally, a collection **C** of structures is domain independent if, for every two structures that differ only on their domain, either both are in **C** or none is in **C**

A collection **C** of structures is domain independent if the following holds:
if $\mathbf{J} = (\mathrm{dom}(\mathbf{J}), R_1^{\mathbf{J}}, \ldots, R_n^{\mathbf{J}}) \in \mathbf{C}$ and $\mathbf{K} = (\mathrm{dom}(\mathbf{K}), R_1^{\mathbf{K}}, \ldots, R_n^{\mathbf{K}})$ is a structure such $R_i^{\mathbf{J}} = R_i^{\mathbf{K}}$,
for each $i \in \{1,\ldots,n\}$, then $\mathbf{K} \in \mathbf{C}$

Closure Under Substructures

closed under isomorphisms, 1-critical, domain independent, closed under substructures,
modular, closed under direct products

J = (dom(**J**),$R_1^{\mathbf{J}}$,…,$R_n^{\mathbf{J}}$) is a substructure of **K** = (dom(**K**),$R_1^{\mathbf{K}}$,…,$R_n^{\mathbf{K}}$), denoted **J** $\sqsubseteq$ **K**, if:

- dom(**J**) $\subseteq$ dom(**K**)
- $R_i^{\mathbf{J}} = R_i^{\mathbf{K}} \cap \mathrm{dom}(\mathbf{J})^{\mathrm{arity}(R_i)}$, for each $i \in \{1,\ldots,n\}$

A collection **C** of structures is closed under substructures if the following holds:

if **J** $\in$ **C** and **K** is a structure such that **K** $\sqsubseteq$ **J**, then **K** $\in$ **C**

Modularity

closed under isomorphisms, 1-critical, domain independent, closed under substructures,
<u>modular</u>, closed under direct products

Informally, a collection **C** of structures is modular if there is a “small witness” structure with a bounded number of elements of why a structure does not belong to **C**

A collection **C** of **S**-structures is n-modular, for $n \geq 0$, if, for every **S**-structure $\mathbf{K} \notin \mathbf{C}$, there is an **S**-structure $\mathbf{J} \sqsubseteq \mathbf{K}$ with $|\mathrm{dom}(\mathbf{J})| \leq n$ such that $\mathbf{J} \notin \mathbf{C}$; we further say that **C** is modular if it is n-modular for some $n \geq 0$

Closure Under Direct Products

closed under isomorphisms, 1-critical, domain independent, closed under substructures, modular, <u>closed under direct products</u>

Let **J** = (dom(**J**),R_1^J,...,R_n^J) and **K** = (dom(**K**),R_1^K,...,R_n^K) be two structures. The direct product of **J** and **K** is the structure **J** $\otimes$ **K** = (dom(**J**) × dom(**K**),$R_1^{J\otimes K}$,...,$R_n^{J\otimes K}$), where, for i ∈ {1,...,n},

$R_i^{J\otimes K} = \{((a_1,b_1),\ldots,(a_{arity(R_i)},b_{arity(R_i)})) : (a_1,\ldots,a_{arity(R_i)}) \in R_i^J \text{ and } (b_1,\ldots,b_{arity(R_i)}) \in R_i^K\}$.

A collection **C** of structures is closed under direct products if the following holds:

if J ∈ C and K ∈ C, then J ⊗ K ∈ C

An Early Result for Full TGDs + EGDs

Theorem ([Makowsky and Vardi, 1986]): Let **C** be a collection of structures. The following are equivalent:

1. **C** is definable by full tgds + egds
2. **C** is closed under isomorphisms, 1-critical, domain independent, modular, closed under substructures, and closed under direct products

$$\text{tgds} \xrightarrow{???} \text{tgds + egds}$$

Definability by TGDs + EGDs

When is a collection **C** of structures definable by tgds + egds?

i.e., there exists an ontology O from tgds + egds such that **C** = {J : J ⊨ O}

Theorem ([Console, Kolaitis, and P., PODS 2021]): Let **C** be a collection of structures.

The following are equivalent:

1. C is definable by tgds + egds
2. C is closed under isomorphisms, 1-critical, local, and closed under direct products

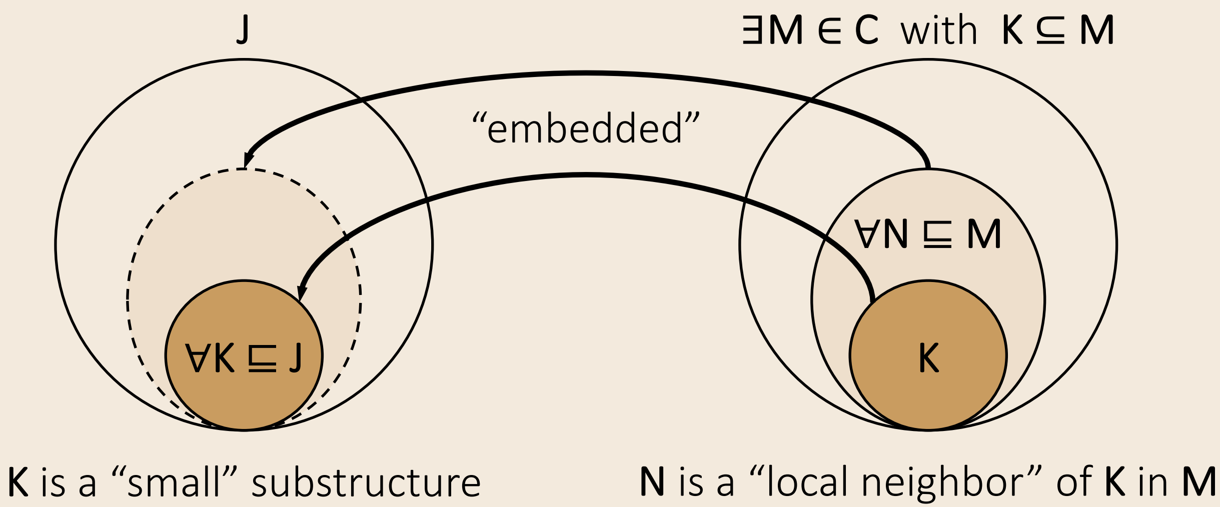
Locality (First Informally)

- A collection **C** of structures is local if, for every structure **J** such that **C** is "locally embeddable" in it, we have that **J** ∈ **C**
- **C** is "locally embeddable" in **J** if the following hold:

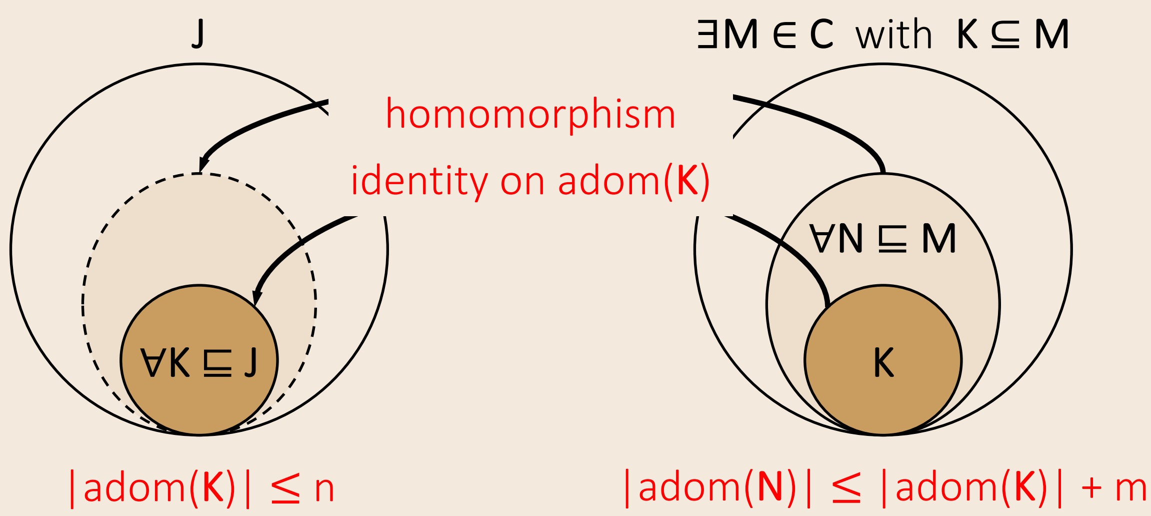
  for every "small" substructure **K** of **J**

  there exists a structure **M** ∈ **C** that contains **K** such that

  every "local neighbor" **N** of **K** in **M** can be "embedded" in **J** while preserving **K**



Locality (First Informally)



C is (n,m)-locally embeddable in J

Locality (Formally)

Let **C** be a collection of structures and **J** a structure. For n,m ≥ 0, **C** is (n,m)-locally embeddable in **J** if the following hold:

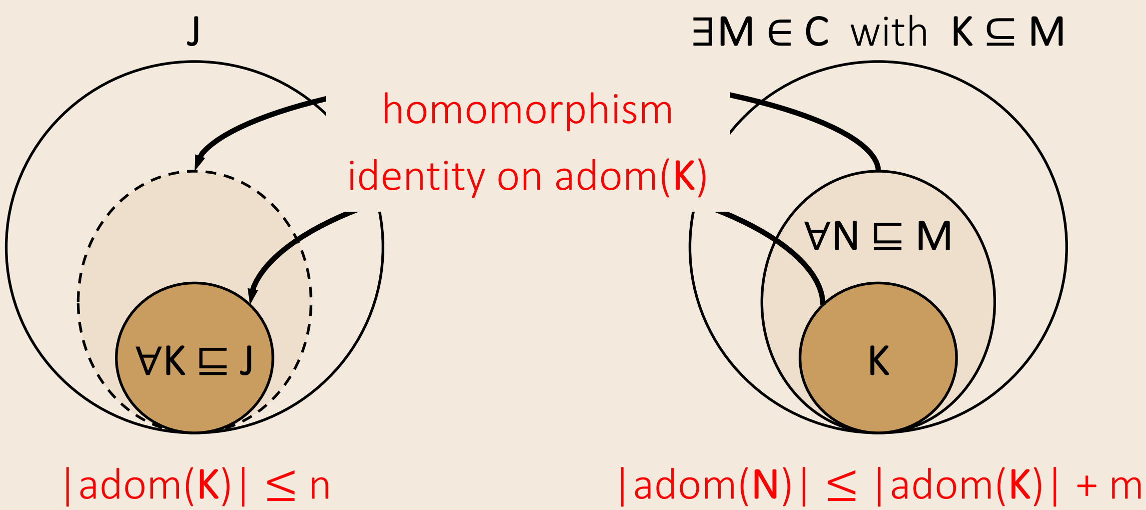
<u>for every</u> substructure **K** of **J** with |adom(**K**)| ≤ n

<u>there exists</u> **M** ∈ **C** that contains **K** such that

<u>for every</u> **N** ∈ {**U** : adom(**K**) ⊆ adom(**U**), **U** ⊑ **M** and |adom(**U**)| ≤ |adom(**K**)| + m}

(*the m-neighbourhood of **K** in **M***)

<u>there exists</u> a homomorphism from **N** to **J** that is the identity on adom(**K**)



Locality (Formally)

Let **C** be a collection of structures and **J** a structure. For n,m $\geq$ 0, **C** is (n,m)-locally embeddable in **J** if the following hold:

for every substructure **K** of **J** with |adom(**K**)| $\leq$ n

there exists **M** ∈ **C** that contains **K** such that

for every **N** ∈ {**U** : adom(**K**) ⊆ adom(**U**), **U** ⊑ **M** and |adom(**U**)| $\leq$ |adom(**K**)| + m}

(*the m-neighbourhood of **K** in **M***)

there exists a homomorphism from **N** to **J** that is the identity on adom(**K**)

A collection **C** of **S**-structures is (n,m)-local, for n,m $\geq$ 0, if, for every **S**-structure **J** such that **C** is (n,m)-locally embeddable in it, it holds that **J** ∈ **C**;
we further say that **C** is local if it is (n,m)-local for some integers n,m $\geq$ 0

Definability by TGDs + EGDs

When is a collection **C** of structures definable by tgds + egds?

i.e., there exists an ontology O from tgds + egds such that **C** = {J : J ⊨ O}

Theorem ([Console, Kolaitis, and P., PODS 2021]): Let **C** be a collection of structures.

The following are equivalent:

1. C is definable by tgds + egds
2. C is closed under isomorphisms, 1-critical, local, and closed under direct products

Definability by TGDs + EGDs

tgds[n,m]: tgds with at most n ∀-variables and m ∃-variables

egds[n]: egds with at most n ∀-variables

Theorem: Let **C** be a collection of structures and n,m $\geq$ 0. The following are equivalent:

1. **C** is definable by tgds[n,m] + egds[n]
2. **C** is closed under isomorphisms, 1-critical, (n,m)-local, and closed under direct products

(1) ⇒ (2): not very difficult

Definability by TGDs Implies Locality

$\mathbf{C}$ is definable by tgds[n,m] $\Rightarrow$ $\mathbf{C}$ is (n,m)-local

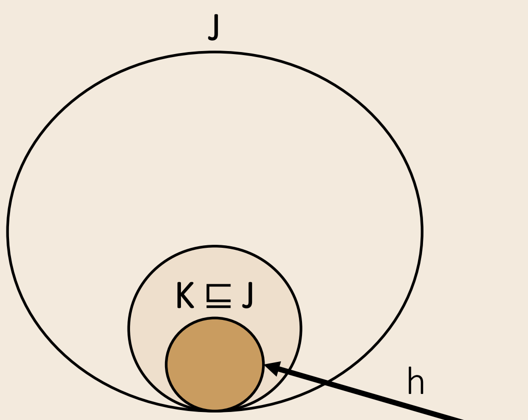
- By hypothesis, there exists a finite set Σ of tgds[n,m] such that $\mathbf{C} = \{\mathbf{J} : \mathbf{J} \models \Sigma\}$
- Consider a structure $\mathbf{J}$ and assume that $\mathbf{C}$ is (n,m)-locally embeddable in $\mathbf{J}$. We need to show that $\mathbf{J} \in \mathbf{C}$, i.e., $\mathbf{J} \models \Sigma$
- Consider a tgd $\sigma \in \Sigma$ of the form $\forall\bar{x}\forall\bar{y}\ (\phi(\bar{x},\bar{y}) \ \rightarrow\ \exists\bar{z}\ \psi(\bar{x},\bar{z}))$, and assume that there is a homomorphism h from $\phi(\bar{x},\bar{y})$ to $\mathbf{J}$. We need to show that there exists a homomorphism from $\psi(\bar{x},\bar{z})$ to $\mathbf{J}$ that agrees with h on the variables of $\bar{x}$
- Let $\mathbf{K} = (\mathrm{dom}(\mathbf{K}),R_1^{\mathbf{K}},\ldots,R_k^{\mathbf{K}})$, where $\mathrm{dom}(\mathbf{K})$ is the set of terms $h(\bar{x}) \cup h(\bar{y})$, i.e., the set of terms occurring in the image of $\phi(\bar{x},\bar{y})$ via h, and, for each $i \in \{1,\ldots,k\}$, $R_i^{\mathbf{K}} = R_i^{\mathbf{J}} \cap \mathrm{dom}(\mathbf{K})^{\mathrm{arity}(R_i)}$; it is clear that $\mathbf{K} \sqsubseteq \mathbf{J}$ with $|\mathrm{adom}(\mathbf{K})| \leq n$

Definability by TGDs Implies Locality

$\mathbf{C}$ is definable by tgds[n,m] $\Rightarrow$ $\mathbf{C}$ is (n,m)-local

(continued)

- By hypothesis, there exists a finite set Σ of tgds[n,m] such that $\mathbf{C} = \{\mathbf{J} : \mathbf{J} \models \Sigma\}$
- Consider a structure $\mathbf{J}$ and assume that $\mathbf{C}$ is (n,m)-locally embeddable in $\mathbf{J}$. We need to show that $\mathbf{J} \in \mathbf{C}$, i.e., $\mathbf{J} \models \Sigma$

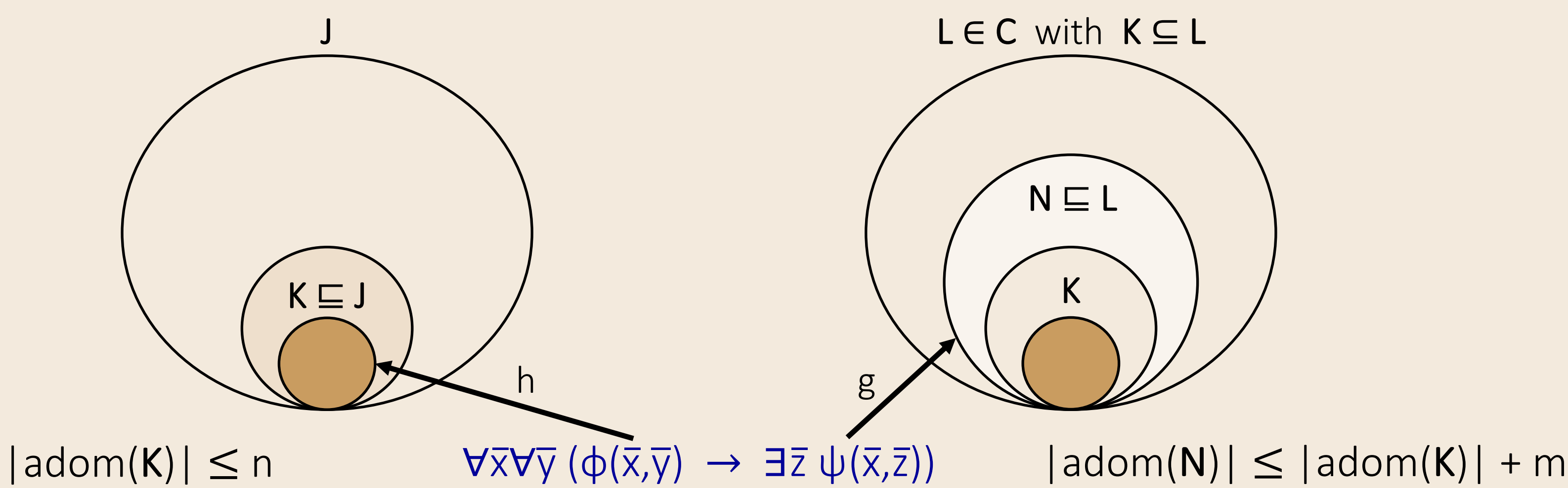


$|\mathrm{adom}(\mathbf{K})| \leq n$ $\quad\quad$ $\forall \bar{x} \forall \bar{y}\ (\phi(\bar{x},\bar{y}) \rightarrow \exists \bar{z}\ \psi(\bar{x},\bar{z}))$

Definability by TGDs Implies Locality

(continued)

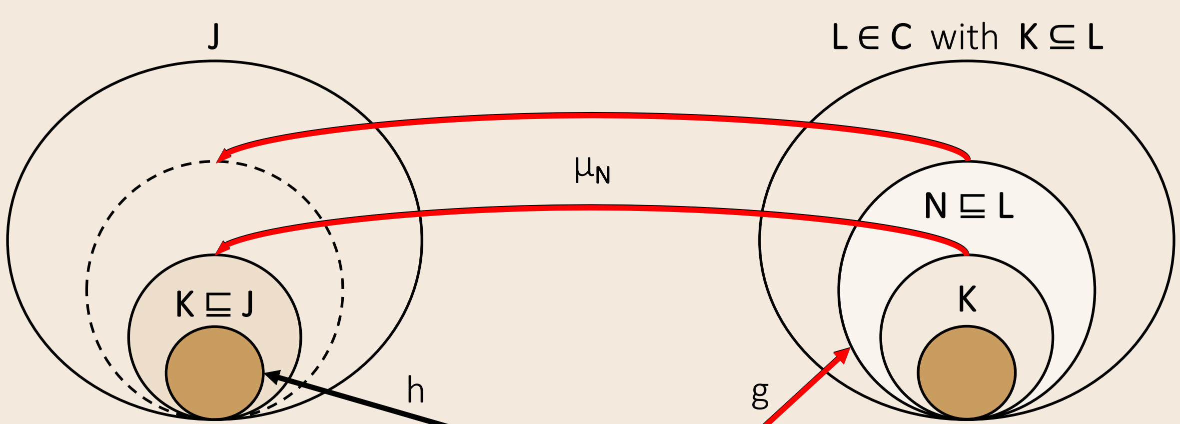
- Since $\mathbf{C}$ is (n,m)-locally embeddable in $\mathbf{J}$, we get that there is a structure $\mathbf{L} \in \mathbf{C}$ such that $\mathbf{K} \subseteq \mathbf{L}$, and, for every $\mathbf{M}$ in the m-neighbourhood of $\mathbf{K}$ in $\mathbf{L}$, there exists a homomorphism $\mu_{\mathbf{M}}$ from $\mathbf{M}$ to $\mathbf{J}$ that is the identity on adom($\mathbf{K}$)
- It is clear that h is a homomorphism from $\phi(\bar{x},\bar{y})$ to $\mathbf{L}$. Since $\mathbf{L} \models \Sigma$, there exists a homomorphism g from $\psi(\bar{x},\bar{z})$ to $\mathbf{L}$ that agrees with h on the variables of $\bar{x}$
- Let $\mathbf{N} = (\mathrm{dom}(\mathbf{N}), R_1^{\mathbf{N}}, \ldots, R_k^{\mathbf{N}})$, where dom($\mathbf{N}$) is the set of terms $g(\bar{x}) \cup g(\bar{z})$, and, for $i \in \{1,\ldots,k\}$, $R_i^{\mathbf{N}} = R_i^{\mathbf{L}} \cap \mathrm{dom}(\mathbf{N})^{\mathrm{arity}(R_i)}$; clearly, $\mathbf{N} \sqsubseteq \mathbf{L}$ with $|\mathrm{adom}(\mathbf{N})| \leq |\mathrm{adom}(\mathbf{K})| + m$



Definability by TGDs Implies Locality

(continued)

- Observe that **N** is in the in the m-neighbourhood of **K** in **L**
- Thus, there exists a homomorphism $\mu_{\mathbf{N}}$ from **N** to **J** that is the identity on adom(**K**)
- Consider the function $\lambda = \mu_{\mathbf{N}} \circ g$
- Since g agrees with h on the variables of $\bar{x}$, and $\mu_{\mathbf{N}}$ is the identity on $h(\bar{x})$, we get that λ agrees with h on the variables of $\bar{x}$
- Hence, λ is a homomorphism from $\psi(\bar{x},\bar{z})$ to **J** that agrees with h on the variables of $\bar{x}$, and therefore, $\mathbf{J} \models \Sigma$ (i.e., $\mathbf{J} \in \mathbf{C}$), as needed



$|\text{adom}(\mathbf{K})| \leq n$ $\quad\forall\bar{x}\forall\bar{y}\,(\phi(\bar{x},\bar{y}) \rightarrow \exists\bar{z}\,\psi(\bar{x},\bar{z}))\quad$ $|\text{adom}(\mathbf{N})| \leq |\text{adom}(\mathbf{K})| + m$

Definability by TGDs + EGDs

tgds[n,m]: tgds with at most n ∀-variables and m ∃-variables

egds[n]: egds with at most n ∀-variables

Theorem: Let **C** be a collection of structures and n,m $\geq$ 0. The following are equivalent:

1. **C** is definable by tgds[n,m] + egds[n]
2. **C** is closed under isomorphisms, 1-critical, (n,m)-local, and closed under direct products

(1) ⇒ (2): not very difficult

(2) ⇒ (1): this direction requires some work

Existential Disjunctive Dependencies

$$\forall \bar{x}\, (\phi(\bar{x}) \rightarrow \bigvee_{i=1}^{k} \psi_i(\bar{x}_i))$$

where $\phi(\bar{x})$ is a (possibly empty) conjunction of atoms, and

for each $i \in \{1,\dots,k\}$, $x_i \subseteq x$, and $\psi_i(\bar{x}_i)$ is either an equality of the form $y = z$,

or a formula $\exists \bar{y}_i\, \chi_i(\bar{x}_i,\bar{y}_i)$ with $\bar{x}_i \cap \bar{y}_i = \emptyset$ and $\chi_i(\bar{x}_i,\bar{y}_i)$ being a non-empty conjunction of atoms

---

A structure **J** satisfies an edd $\delta = \forall \bar{x}\, (\phi(\bar{x}) \rightarrow \bigvee_{i=1}^{k} \psi_i(\bar{x}_i))$, denoted $\mathbf{J} \models \delta$, if, whenever there is

a homomorphism h from $\phi(\bar{x})$ to **J**, then there exists $i \in \{1,\dots,k\}$ such that

if $\psi_i(\bar{x}_i)$ is an equality $y = z$, then $h(y) = h(z)$; otherwise,

if $\psi_i(\bar{x}_i)$ is $\exists \bar{y}_i\, \chi_i(\bar{x}_i,\bar{y}_i)$, then there is a homomorphism from $\chi_i(\bar{x}_i,\bar{y}_i)$ to **J** that agrees with h on $\bar{x}_i$

Definability by TGDs + EGDs

C is closed under isomorphisms, 1-critical, (n,m)-local, and closed under direct products

⇒ **C** is definable by tgds[n,m] + egds[n]

The proof is carried out in two main steps; assume that **C** consists of **S**-structures:

1. We construct a finite set Σ[V] of edds over **S** with at most n ∀-variables and m ∃-variables such that **C** = {**J** : **J** ⊨ Σ[V]}. In particular,

$$\Sigma[V] = \{ \delta \in \text{edds}[n,m] : \text{for each } \mathbf{J} \in \mathbf{C}, \text{ it holds that } \mathbf{J} \models \delta \}$$

we exploit closure under isomorphisms, 1-criticality, and (n,m)-locality

together with Robinson's method of diagrams

Definability by TGDs + EGDs

C is closed under isomorphisms, 1-critical, (n,m)-local, and closed under direct products

⇒ **C** is definable by tgds[n,m] + egds[n]

The proof is carried out in two main steps; assume that **C** consists of **S**-structures:

2. We show that there exists a finite set $\Sigma[\exists,=]$ of tgds[n,m] and egds[n] over **S** such that $\Sigma[\vee]$ and $\Sigma[\exists,=]$ are logically equivalent. In particular, we show that

$$\Sigma[\exists,=] \;=\; \{\delta \in \Sigma[\vee] : \delta \text{ is a tgd or an egd}\}$$

we exploit closure under direct products

together with McKinsey's method of elimination of disjunctions

Definability by TGDs + EGDs

When is a collection **C** of structures definable by tgds + egds?

i.e., there exists an ontology O from tgds + egds such that **C** = {J : J ⊨ O}

Theorem ([Console, Kolaitis, and P., PODS 2021]): Let **C** be a collection of structures.

The following are equivalent:

1. C is definable by tgds + egds
2. C is closed under isomorphisms, 1-critical, local, and closed under direct products

Definability by Full TGDs + EGDs

Theorem: Let **C** be a collection of structures. The following are equivalent:

1. **C** is definable by full tgds + egds
2. **C** is closed under isomorphisms, 1-critical, (n,0)-local for some n > 0, and closed under direct products

domain independent, modular, closed under substructures

⇓

Theorem ([Makowsky and Vardi, 1986]): Let **C** be a collection of structures. The following are equivalent:

1. **C** is definable by full tgds + egds
2. **C** is closed under isomorphisms, 1-critical, domain independent, modular, closed under substructures, and closed under direct products

So Far

tgds + egds

closure under isomorphisms, 1-criticality, locality, and closure under direct products

↑

tgds

???

Definability by TGDs

tgds + egds

closure under isomorphisms, 1-criticality, locality, and closure under direct products

↑

tgds

???

Adopt an additional property together with closure under isomorphisms, 1-criticality, locality, and closure under direct products, which allows us to show that $\Sigma[\exists,=]$ is logically equivalent to $\{\delta \in \Sigma[\exists,=] : \delta \text{ is a tgd}\}$

Definability by TGDs

tgds + egds

closure under isomorphisms, 1-criticality, locality, and closure under direct products

↑

tgds

+ criticality

Adopt an additional property together with closure under isomorphisms, 1-criticality, locality, and closure under direct products, which allows us to show that $\Sigma[\exists,=]$ is logically equivalent to $\{\delta \in \Sigma[\exists,=] : \delta \text{ is a tgd}\}$

Criticality

A structure **J** = (dom(**J**),$R_1^{\mathbf{J}}$,…,$R_n^{\mathbf{J}}$) is k-critical, for some integer k > 0, if:

- dom(**J**) consists of k distinct elements
- $R_i^{\mathbf{J}} = \mathrm{dom}(\mathbf{J})^{\mathrm{arity}(R_i)}$, for each i ∈ {1,…,n}

A collection **C** of structures is critical if it contains a k-critical structure for each k > 0

Definability by TGDs (+ EGDs)

Theorem ([Console, Kolaitis, and P., PODS 2021]**):** Let **C** be a collection of structures.

The following are equivalent:

1. **C** is definable by tgds + egds
2. **C** is closed under isomorphisms, 1-critical, local, and closed under direct products

Theorem ([Console, Kolaitis, and P., PODS 2021]**):** Let **C** be a collection of structures.

The following are equivalent:

1. **C** is definable by tgds
2. **C** is closed under isomorphisms, critical, local, and closed under direct products

Definability by Subclasses of TGDs

tgds

closure under isomorphisms, criticality, locality, and closure under direct products

full tgds

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

???

guarded tgds

$\forall \bar{x} \forall \bar{y}\, (R(\bar{x},\bar{y}) \wedge \phi(\bar{x},\bar{y}) \;\rightarrow\; \exists \bar{z}\, \psi(\bar{x},\bar{z}))$

???

linear tgds

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

???

Definability by Subclasses of TGDs

tgds

closure under isomorphisms, criticality, locality, and closure under direct products

refine locality

full tgds

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

(n,0)-locality for some n > 0

guarded tgds

$\forall \bar{x} \forall \bar{y} \, (R(\bar{x},\bar{y}) \wedge \phi(\bar{x},\bar{y}) \;\rightarrow\; \exists \bar{z} \, \psi(\bar{x},\bar{z}))$

guarded locality

linear tgds

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

linear locality

Linear Structures

A structure $\mathbf{J} = (dom(\mathbf{J}),R_1^{\mathbf{J}},\ldots,R_n^{\mathbf{J}})$ is linear if:

- $|R_i^{\mathbf{J}}| = 0$, for each $i \in \{1,\ldots,n\}$, or
- there exists $i \in \{1,\ldots,n\}$ such that $|R_i^{\mathbf{J}}| = 1$ and $|R_j^{\mathbf{J}}| = 0$, for each $j \in \{1,\ldots,n\} \setminus \{i\}$

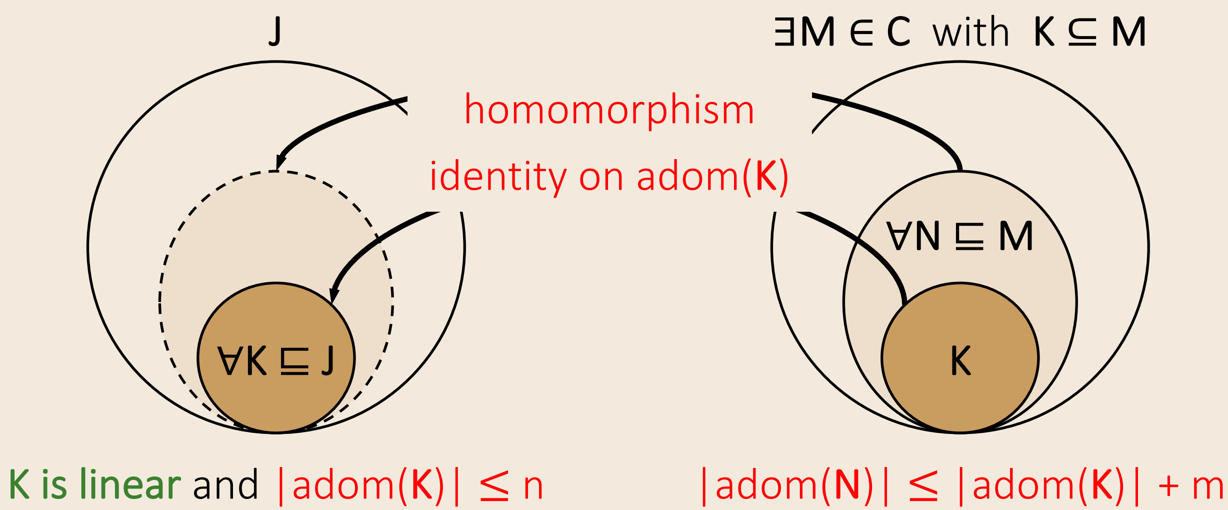
Consider a structure $\mathbf{J} = (\{a,b\}, R^J, P^J)$, where R is binary and P is ternary

linear	non-linear
$R^J = \emptyset$ and $P^J = \emptyset$	$R^J = \{(a,a), (a,b)\}$ and $P^J = \emptyset$
$R^J = \{(a,a)\}$ and $P^J = \emptyset$	$R^J = \{(a,a)\}$ and $P^J = \{(b,a,a)\}$

From Locality to Linear Locality

Let **C** be a collection of structures and **J** a structure. For n,m ≥ 0, **C** is linearly (n,m)-locally embeddable in **J** if the following hold:

> for every linear substructure **K** of **J** with |adom(**K**)| ≤ n
>
> there exists **M** ∈ **C** that contains **K** such that
>
> for every **N** ∈ {**U** : adom(**K**) ⊆ adom(**U**), **U** ⊑ **M** and |adom(**U**)| ≤ |adom(**K**)| + m}
>
> (*the m-neighbourhood of **K** in **M***)
>
> there exists a homomorphism from **N** to **J** that is the identity on adom(**K**)



From Locality to Linear Locality

Let **C** be a collection of structures and **J** a structure. For n,m $\geq$ 0, **C** is linearly (n,m)-locally embeddable in **J** if the following hold:

<u>for every</u> linear substructure **K** of **J** with |adom(**K**)| $\leq$ n

<u>there exists</u> **M** $\in$ **C** that contains **K** such that

<u>for every</u> **N** $\in$ {**U** : adom(**K**) $\subseteq$ adom(**U**), **U** $\sqsubseteq$ **M** and |adom(**U**)| $\leq$ |adom(**K**)| + m}

(*the m-neighbourhood of **K** in **M***)

<u>there exists</u> a homomorphism from **N** to **J** that is the identity on adom(**K**)

A collection **C** of **S**-structures is linear (n,m)-local, for n,m $\geq$ 0, if, for every **S**-structure **J** such that **C** is linearly (n,m)-locally embeddable in it, it holds that **J** $\in$ **C**;
we further say that **C** is linear local if it is linear (n,m)-local for some integers n,m $\geq$ 0

Guarded Structures

A structure **J** = (dom(**J**),R_1^J,…,R_n^J) is guarded if:

- $|R_i^J| = 0$, for each i ∈ {1,…,n}, or
- there is i ∈ {1,…,n} and a tuple $(a_1,\ldots,a_{arity(R_i)}) \in R_i^J$ such that adom(**J**) = $\{a_1,\ldots,a_{arity(R_i)}\}$

Consider a structure **J** = ({a,b,c}, R^J, P^J), where R is binary and P is ternary

guarded	**non-guarded**
R^J = {(a,a), (a,b)} and P^J = Ø	R^J = {(a,a)} and P^J = {(b,b,b)}
R^J = {(a,a)} and P^J = {(b,a,a)}	R^J = {(a,a), (a,b)} and P^J = {(b,b,c)}

From Locality to Guarded Locality

Let **C** be a collection of structures and **J** a structure. For n,m $\geq$ 0, **C** is guardedly (n,m)-locally embeddable in **J** if the following hold:

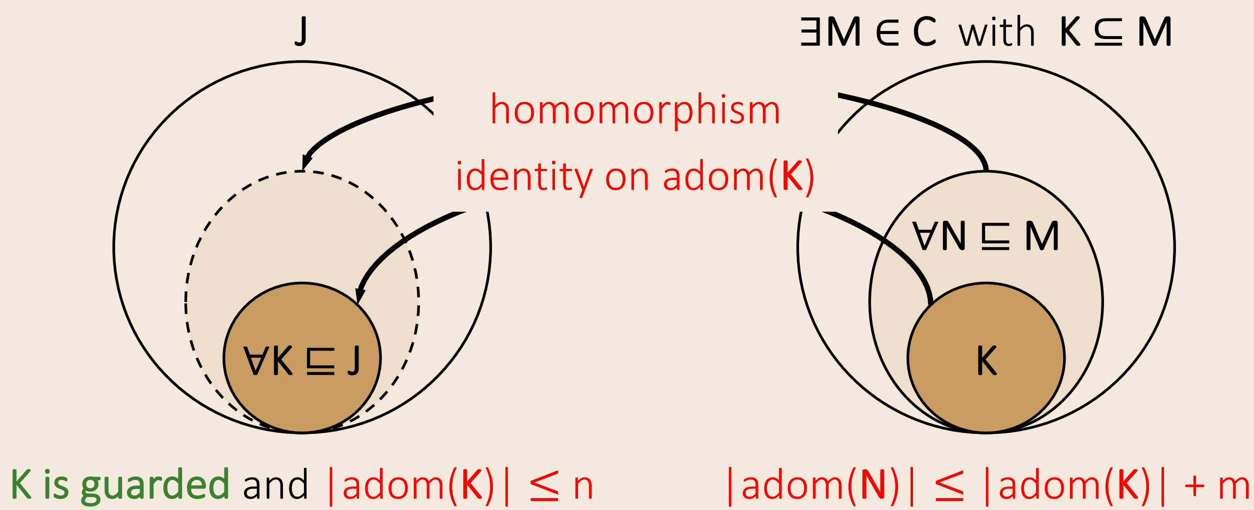
<u>for every</u> guarded substructure **K** of **J** with |adom(**K**)| $\leq$ n

<u>there exists</u> **M** ∈ **C** that contains **K** such that

<u>for every</u> **N** ∈ {**U** : adom(**K**) ⊆ adom(**U**), **U** ⊑ **M** and |adom(**U**)| $\leq$ |adom(**K**)| + m}

(*the m-neighbourhood of* ***K*** *in* ***M***)

<u>there exists</u> a homomorphism from **N** to **J** that is the identity on adom(**K**)



From Locality to Guarded Locality

Let **C** be a collection of structures and **J** a structure. For n,m $\geq$ 0, **C** is guardedly (n,m)-locally embeddable in **J** if the following hold:

for every guarded substructure **K** of **J** with |adom(**K**)| $\leq$ n

there exists **M** $\in$ **C** that contains **K** such that

for every **N** $\in$ {**U** : adom(**K**) $\subseteq$ adom(**U**), **U** $\sqsubseteq$ **M** and |adom(**U**)| $\leq$ |adom(**K**)| + m}

(*the m-neighbourhood of **K** in **M***)

there exists a homomorphism from **N** to **J** that is the identity on adom(**K**)

A collection **C** of **S**-structures is guarded (n,m)-local, for n,m $\geq$ 0, if, for every **S**-structure **J** such that **C** is guardedly (n,m)-locally embeddable in it, it holds that **J** $\in$ **C**;
we further say that **C** is guarded local if it is guarded (n,m)-local for some integers n,m $\geq$ 0

Definability by Subclasses of TGDs

Theorem: Let **C** be a collection of structures. The following are equivalent:

1. **C** is definable by full tgds
2. **C** is closed under isomorphisms, 1-critical, (n,0)-local for some n > 0, and closed under direct products

Theorem: Let **C** be a collection of structures. The following are equivalent:

1. **C** is definable by linear tgds
2. **C** is closed under isomorphisms, critical, linear local, and closed under direct products

Theorem: Let **C** be a collection of structures. The following are equivalent:

1. **C** is definable by guarded tgds
2. **C** is closed under isomorphisms, critical, guarded local, and closed under direct products

Rewritability Theorems

Theorem: Let Σ be finite set of tgds. The following are equivalent:

1. Σ is logically equivalent to a finite set of full tgds
2. Σ is (n,0)-local for some integer $n > 0$ (i.e., $\{J : J \models \Sigma\}$ is (n,0)-local)

Theorem: Let Σ be finite set of tgds. The following are equivalent:

1. Σ is logically equivalent to a finite set of linear tgds
2. Σ is linear local (i.e., $\{J : J \models \Sigma\}$ is linear local)

Theorem: Let Σ be finite set of tgds. The following are equivalent:

1. Σ is logically equivalent to a finite set of guarded tgds
2. Σ is guarded local (i.e., $\{J : J \models \Sigma\}$ is guarded local)

Rewritability Problem

Rewrite(L_1,L_2)

Input: A finite set Σ of tgds from L_1

Question: Is there a finite set Σ' of tgds from L_2 such that $\Sigma \equiv \Sigma'$?

Theorem: Rewrite(guarded tgds, linear tgds) is 2EXPTIME-complete, and EXPTIME-complete for schemas of bounded arity

Linearization Theorem

Theorem: Let Σ be finite set of tgds. The following are equivalent:

1. Σ is logically equivalent to a finite set of linear tgds
2. Σ is linear local (i.e., {J : J ⊨ Σ} is linear local)

Linearization Theorem: Let Σ be a finite set of tgds[n,m]. The following are equivalent:

1. Σ is logically equivalent to a finite set of linear tgds
2. Σ is logically equivalent to a finite set of linear tgds[n,m]
3. Σ is linear (n,m)-local (i.e., {J : J ⊨ Σ} is linear (n,m)-local)

From Guarded to Linear

> **Theorem:** Rewrite(guarded tgds, linear tgds) is 2EXPTIME-complete, and EXPTIME-complete for schemas of bounded arity

- Let Σ be a finite set of guarded tgds[n,m] over a schema **S**
- By the Linearizaton Theorem, if Σ is logically equivalent to a finite set of linear tgds, then it is logically equivalent to a finite set of linear tgds[n,m]
- Let $\Sigma' = \{\sigma : \sigma$ is a linear tgd[n,m] over **S** such that $\Sigma \models \sigma\}$; observe that Σ' is finite (up to variable renaming)
- Clearly, if $\Sigma' \neq \emptyset$ and $\Sigma' \models \Sigma$, then $\Sigma' \equiv \Sigma$; otherwise, there is no finite set of linear tgds that is logically equivalent to Σ
- We can build Σ' and check whether $\Sigma' \models \Sigma$ in 2EXPTIME, and in EXPTIME assuming that the relation names of **S** have bounded arity

Recap

Main Question: When is a collection **C** of structures definable by a language **L**?

tgds + egds

closure under isomorphisms, 1-criticality, locality, and closure under direct products

refine 1-criticality

tgds

criticality

refine locality

full tgds

(n,0)-locality for some n > 0

guarded tgds

guarded locality

refine guarded locality

linear tgds

linear locality

Ongoing Research

disjunctive tuple-generating dependencies

$$\forall \bar{x} \, (\phi(\bar{x}) \ \rightarrow \ \bigvee_{i=1}^{k} \exists \bar{z}_i \ \psi_i(\bar{x}_i, \bar{z}_i))$$

where $\phi(\bar{x})$ is a (possibly empty) conjunction of atoms, and

for each $i \in \{1,\ldots,k\}$, $x_i \subseteq x$ and $\psi_i(\bar{x}_i,\bar{z}_i)$ is a non-empty conjunction of atoms

When is a collection **C** of structures definable by disjunctive tgds (+ egds)?

currently working on this with Marco Calautti and Marco Console

Open Problems

- When is a collection of structures definable by tgds from classes that are based on global conditions (acyclic, weakly-acyclic, sticky, weakly-(frontier-)guarded, etc.)?

- When is a collection of <u>databases</u> **C** definable by <u>ontology-mediated queries</u> (i.e., there is a Boolean omq Q from (tgd,cq) – or subclasses – such that **C** = {**D** : **D** $\models$ Q})?

Thank You!