


# Query Answering with Transitive & Linear-Ordered Data

Antoine Amarilli<sup>1</sup>, Michael Benedikt<sup>2</sup>,  
Pierre Bourhis<sup>3</sup>, Michael Vanden Boom<sup>2</sup>

1: LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay; Paris, France  
2: University of Oxford; Oxford, United Kingdom  
3: CNRS CRISTAL, Université Lille 1, INRIA Lille; Lille, France

## Query Answering (or query entailment)

 **Initial set of facts**  $\mathcal{F}$  (or A-box)

 **Constraints**  $\Sigma$ : logical rules (or T-box)

 **Boolean query**  $Q$ : CQ or UCQ

**QA problem:** does  $\mathcal{F} \wedge \Sigma$  entail  $Q$ ?

Equivalently:

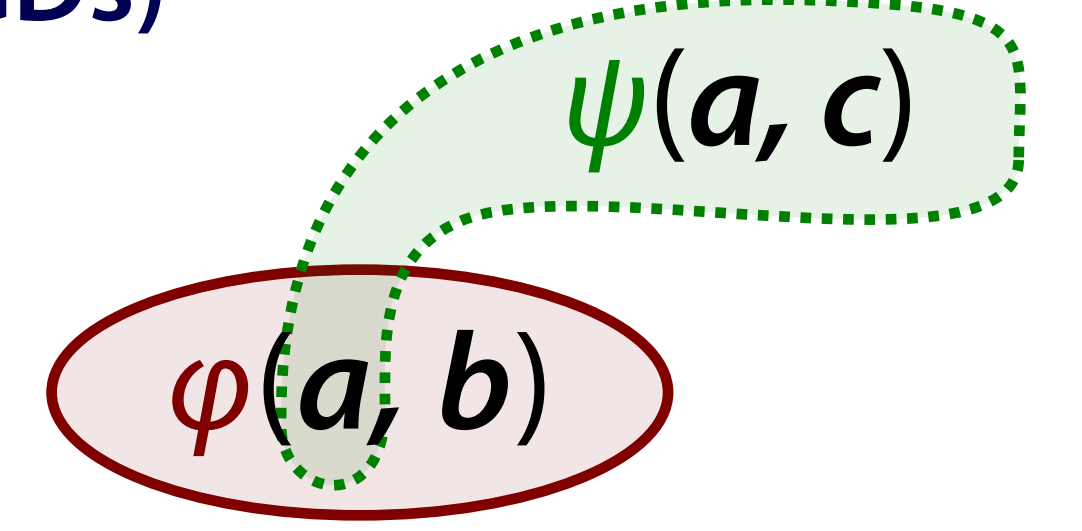
- is  $Q$  certain given  $\mathcal{F}$  and  $\Sigma$ ?
- is  $\mathcal{F} \wedge \Sigma \wedge \neg Q$  unsatisfiable?

**Our goal:** identify and study constraint languages for which QA is decidable, even when some relations are restricted to be **transitive** or to be **linear orders**.

## Constraint languages

**Tuple-generating dependencies (TGDs)**  
(or existential rules)

$\forall x y \varphi(x, y) \rightarrow \exists z \psi(x, z)$   
where  $\varphi$  (body) and  $\psi$  (head) are conjunctions of atoms



**Frontier-guarded TGDs (FGTGDs)**

$\forall x y \varphi(x, y) \wedge G(x) \rightarrow \exists z \psi(x, z)$

Atom using variables  $x$  is called a **guard** for  $x$

**Guarded Negation Fragment (GNF)**

rules built up from atoms using

- disjunction
- guarded negation
- existential quantification

$G(x) \wedge \neg \psi(x)$

## Our approach

Fix relational signature  $\sigma := \sigma_B \sqcup \sigma_D$  where

- $\sigma_D$ : **distinguished** binary relations
- $\sigma_B$ : **base** relations

We consider query answering with three different special interpretations for the distinguished relations:

- **QAttr**: each  $R \in \sigma_D$  is transitively closed
- **QAtc**: each  $R^+ \in \sigma_D$  is the transitive closure of  $R \in \sigma_B$
- **QAlin**: each  $R \in \sigma_D$  is a linear order

We introduce **base-frontier-guarded** and **base-covered** constraint languages that disallow the use of distinguished relations as guards.

**BaseFGTGDs**

FGTGDs where guards for frontier variables are from  $\sigma_B$   
e.g.  $\forall x y_1 y_2 R(x, y_1) \wedge R(x, y_2) \wedge S(y_1, y_2) \rightarrow \exists z R(y_2, z) \wedge T(y_1)$   
where  $\sigma_D = \{R\}$  and  $\sigma_B = \{S, T\}$

**BaseCovFGTGDs**

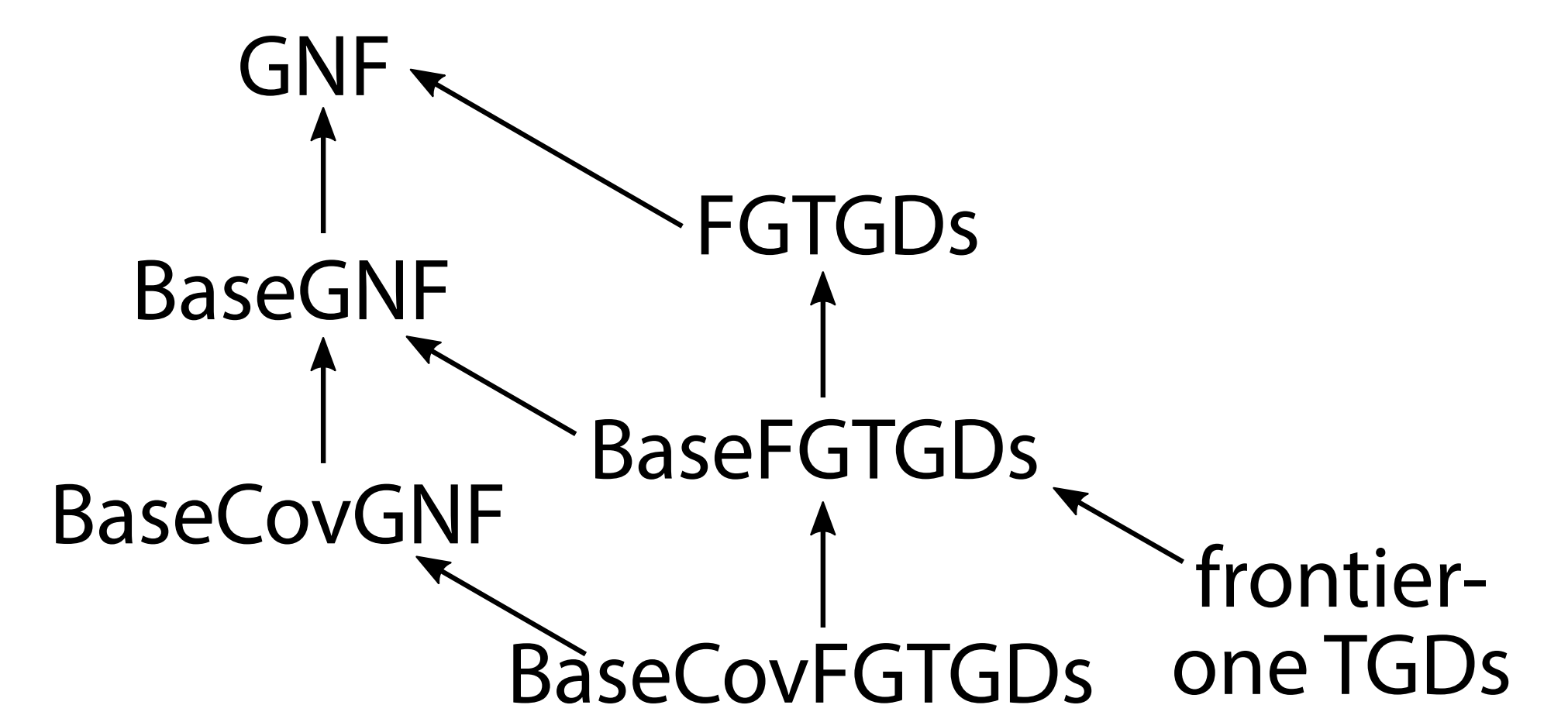
BaseFGTGDs where for every  $\sigma_D$  atom in the body using variables  $v$ , there is a  $\sigma_B$  atom in the body guarding  $v$

e.g.  $\forall x y_1 y_2 C(x, y_1) \wedge R(x, y_1) \wedge C(x, y_2) \wedge R(x, y_2) \wedge S(y_1, y_2) \rightarrow \exists z R(y_2, z) \wedge T(y_1)$   
where  $\sigma_D = \{R\}$  and  $\sigma_B = \{S, T, C\}$

**BaseGNF**

GNF where guards for negation are from  $\sigma_B$

**BaseCovGNF**  
Generalization of BaseCovFGTGDs



## Main results

**QAttr** & **QAtc** are decidable for **BaseGNF** (undecidable for FGTGDs).

**QAlin** is decidable for **BaseCovGNF** (undecidable for BaseFGTGDs).

## Complexity

	QAttr		QAtc		QAlin	
	data	combined	data	combined	data	combined
<b>BaseGNF</b>	coNP-c	2EXP-c	coNP-c	2EXP-c	undecidable	
<b>BaseCovGNF</b>	coNP-c	2EXP-c	coNP-c	2EXP-c	coNP-c	2EXP-c
<b>BaseFGTGDs</b>	in coNP	2EXP-c	coNP-c	2EXP-c	undecidable	
<b>BaseCovFGTGDs</b>	P-c	2EXP-c	coNP-c	2EXP-c	coNP-c	2EXP-c

## Proof ideas

**Key property:**

For  $\varphi$  in GNF, if  $\varphi$  is satisfiable then it has a **tree-like** witness: a set of facts satisfying  $\varphi$  that has a **tree decomposition** of bounded tree-width.

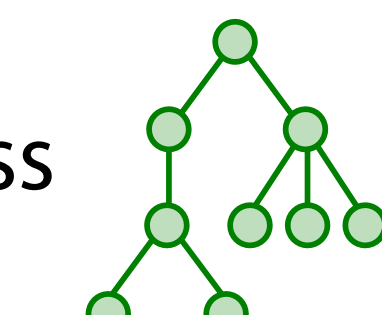
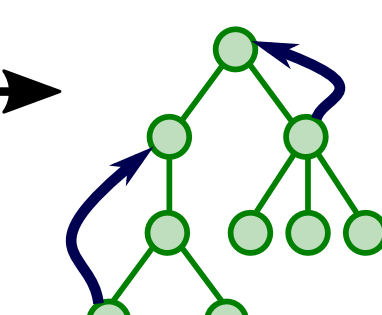
For **QAttr** & **QAtc** with base-frontier-guarded constraints  $\Sigma$ : reduce to tree automaton emptiness test.  $L(\mathcal{A}) = \emptyset?$

For BaseGNF, there are tree-like witnesses even when each distinguished relation is required to be the transitive closure of some base relation.

Hence it suffices to construct a tree automaton  $\mathcal{A}$  that runs on encodings of tree-like sets of facts and checks  $\mathcal{F} \wedge \Sigma \wedge \neg Q$ .

For **QAttr** & **QAlin** with base-covered constraints  $\Sigma$ : reduce to traditional QA with GNF  $\Sigma'$ .

Cannot axiomatize transitivity or totality using GNF, but can approximate using  $\Sigma'$  in GNF. Key technical result shows that a tree-like approximate witness can be extended to an actual witness respecting special interpretations for  $\sigma_D$  relations.

Tree-like witness for  $\mathcal{F} \wedge \Sigma' \wedge \neg Q$    $\rightarrow$   Actual witness for  $\mathcal{F} \wedge \Sigma \wedge \neg Q$