Query Answering with Transitive and Linear-Ordered Data

Antoine Amarilli\textsuperscript{1}, Michael Benedikt\textsuperscript{2}, Pierre Bourhis\textsuperscript{3} and Michael Vanden Boom\textsuperscript{2}

\textsuperscript{1}LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay
\textsuperscript{2}University of Oxford
\textsuperscript{3}CNRS CRIStAL, Université Lille 1, INRIA Lille

IJCAI 2016
New York, USA
Query answering problem (QA)

Given: finite set of initial facts $\mathcal{F}_0$, constraints $\Sigma$, boolean query $Q$ (UCQ).

The query answering problem $\text{QA}(\mathcal{F}_0, \Sigma, Q)$ asks: does $\mathcal{F}_0 \land \Sigma$ entail $Q$?
Query answering problem (QA)

Given: finite set of initial facts $\mathcal{F}_0$, constraints $\Sigma$, boolean query $Q$ (UCQ).

The query answering problem $\text{QA}(\mathcal{F}_0, \Sigma, Q)$ asks: does $\mathcal{F}_0 \land \Sigma$ entail $Q$?

Equivalently:

- is $Q$ certain given $\mathcal{F}_0$ and $\Sigma$?
- is $\mathcal{F}_0 \land \Sigma \land \neg Q$ unsatisfiable?
- for all sets of facts $\mathcal{F} \supseteq \mathcal{F}_0$ satisfying $\Sigma$, does $\mathcal{F}$ satisfy $Q$?
Query answering problem (QA)

Given: finite set of initial facts $\mathcal{F}_0$, constraints $\Sigma$, boolean query $Q$ (UCQ).

The **query answering problem** QA($\mathcal{F}_0$, $\Sigma$, $Q$) asks: does $\mathcal{F}_0 \land \Sigma$ entail $Q$?

Equivalently:

- is $Q$ certain given $\mathcal{F}_0$ and $\Sigma$?
- is $\mathcal{F}_0 \land \Sigma \land \neg Q$ unsatisfiable?
- for all sets of facts $\mathcal{F} \supseteq \mathcal{F}_0$ satisfying $\Sigma$, does $\mathcal{F}$ satisfy $Q$?

**Example**

$\mathcal{F}_0$ : $S(a, b), R(b, a)$

$\Sigma$ : $\forall xy \ (S(x, y) \rightarrow R(x, y))$

$\quad \forall x \ (R(x, x) \rightarrow \exists y \ T(y))$

$Q : \exists x \ T(x)$

$Q$ is not certain in general...
Query answering problem (QA)

Given: finite set of initial facts $\mathcal{F}_0$, constraints $\Sigma$, boolean query $Q$ (UCQ).

The **query answering problem** $\text{QA}(\mathcal{F}_0, \Sigma, Q)$ asks: does $\mathcal{F}_0 \land \Sigma$ entail $Q$?

Equivalently:

is $Q$ certain given $\mathcal{F}_0$ and $\Sigma$?

is $\mathcal{F}_0 \land \Sigma \land \neg Q$ unsatisfiable?

for all sets of facts $\mathcal{F} \supseteq \mathcal{F}_0$ satisfying $\Sigma$, does $\mathcal{F}$ satisfy $Q$?

**Example**

\[
\begin{align*}
\mathcal{F}_0 & : \quad S(a, b), R(b, a) \\
\Sigma & : \quad \forall xy \ (S(x, y) \rightarrow R(x, y)) \\
& \quad \forall x \ (R(x, x) \rightarrow \exists y \ T(y)) \\
Q & : \quad \exists x \ T(x)
\end{align*}
\]

$Q$ is not certain in general... but it is certain when $R$ is a transitive relation.
Transitivity in description logics

Many DLs support transitive relations.

QA is **decidable** for

- \( ZIQ, ZOQ, ZOI \) [Calvanese et al., 2009]
- Horn-\( SROIQ \) [Ortiz et al., 2011]
- \( \text{regular-}E\mathcal{L}^{++} \) [Krötzsch and Rudolph, 2007]

(sometimes with restrictions on interaction between transitivity & other features).

QA is **undecidable** for

- \( \text{ALCOIF}^* \) [Ortiz et al., 2010]
- \( ZOIQ \) [Ortiz, 2010]

QA is **open** for

- \( SROIQ \) and \( SHOIQ \) [Ortiz and Šimkus, 2012]
QA with tuple generating dependencies (a.k.a. existential rules)

\[ \forall xy \ (\varphi(x, y) \rightarrow \exists z \ \psi(y, z)) \]

body \( \varphi \) and head \( \psi \) are a conjunction of atoms
QA with tuple generating dependencies (a.k.a. existential rules)

\[ \text{TGD: } \forall xy (\varphi(x, y) \rightarrow \exists z \psi(y, z)) \n\]
body \( \varphi \) and head \( \psi \) are a conjunction of atoms

**Frontier-guarded TGD (FGTGD):**
\( \varphi \) includes atom using all of the frontier variables \( y \)

\[ \forall x y_1 y_2 (S(x, y_1) \land S(x, y_2) \land R(y_1, y_2) \rightarrow \exists z (S(y_2, z) \land T(y_1))) \]
QA with tuple generating dependencies (a.k.a. existential rules)

**TGD:** $\forall xy \ (\varphi(x,y) \rightarrow \exists z \ \psi(y,z))$

body $\varphi$ and head $\psi$ are a conjunction of atoms

**Frontier-guarded TGD (FGTGD):**
$\varphi$ includes atom using all of the frontier variables $y$

$\forall x y_1 y_2 \ (S(x, y_1) \land S(x, y_2) \land R(y_1, y_2) \rightarrow \exists z \ (S(y_2, z) \land T(y_1)))$

QA is **decidable** with FGTGD constraints and UCQ. [Baget et al., 2011]
QA with tuple generating dependencies (a.k.a. existential rules)

\[ \forall xy \left( \varphi(x, y) \rightarrow \exists z \psi(y, z) \right) \]

body \( \varphi \) and head \( \psi \) are a conjunction of atoms

Frontier-guarded TGD (FGTGD):
\( \varphi \) includes atom using all of the frontier variables \( y \)
\[ \forall x \ y_1 \ y_2 \left( S(x, y_1) \land S(x, y_2) \land R(y_1, y_2) \rightarrow \exists z \left( S(y_2, z) \land T(y_1) \right) \right) \]

QA is \text{decidable} \ with FGTGD constraints and UCQ. [Baget et al., 2011]

FGTGDs cannot express transitivity, and QA is \text{undecidable} \ with FGTGDs when some relations are required to be transitive. [Gottlob et al., 2013]
QA with tuple generating dependencies (a.k.a. existential rules)

**TGD:** \( \forall xy (\varphi(x, y) \rightarrow \exists z \psi(y, z)) \)

body \( \varphi \) and head \( \psi \) are a conjunction of atoms

**Frontier-guarded TGD (FGTGD):**
\( \varphi \) includes atom using all of the frontier variables \( y \)
\( \forall x y_1 y_2 (S(x, y_1) \land S(x, y_2) \land R(y_1, y_2) \rightarrow \exists z (S(y_2, z) \land T(y_1))) \)

QA is **decidable** with FGTGD constraints and UCQ. [Baget et al., 2011]

FGTGDs cannot express transitivity, and QA is **undecidable** with FGTGDs when some relations are required to be transitive. [Gottlob et al., 2013]

How can we recover **decidability** for QA with transitive relations?
- restrict to (subclass of) linear TGDs [Baget et al., 2015];
- disallow the transitive relations as guards (our approach).
Our approach

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

$\sigma_D$: **distinguished binary relations** with special interpretations (e.g., transitively closed) 

$\sigma_B$: **base relations**

We introduce constraint languages that disallow $\sigma_D$-relations as guards:

**Base FGTGD:** FGTGD where guard for frontier variables is from $\sigma_B$. 

$$\forall x y_1 y_2 \left( R(x, y_1) \land R(x, y_2) \land S(y_1, y_2) \rightarrow \exists z \left( R(y_2, z) \land T(y_1) \right) \right)$$
Our approach

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

$\sigma_D$: distinguished binary relations with special interpretations (e.g., transitively closed)

$\sigma_B$: base relations

We introduce constraint languages that disallow $\sigma_D$-relations as guards:

Base FGTGD: FGTGD where guard for frontier variables is from $\sigma_B$.

$\forall x_1 y_1 y_2 \left( R(x, y_1) \land R(x, y_2) \land S(y_1, y_2) \rightarrow \exists z \left( R(y_2, z) \land T(y_1) \right) \right)$

Base-covered FGTGD: Base FGTGD where for every $\sigma_D$-atom in the body, there is a $\sigma_B$-atom in the body using its variables.

$\forall x_1 y_1 y_2 \left( C(x, y_1) \land R(x, y_1) \land C(x, y_2) \land R(x, y_2) \land S(y_1, y_2) \rightarrow \exists z \left( R(y_2, z) \land T(y_1) \right) \right)$
Our contribution

We consider three different special interpretations for relations in $\sigma_D$:

- **QAt** 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
Our contribution

We consider three different special interpretations for relations in $\sigma_D$:

- **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

**Theorem**

QAttr and QAtc are **decidable** with base FGTGDs and UCQ.

QAlin is **decidable** with base-covered FGTGDs and base-covered UCQ.
Our contribution

We consider three different special interpretations for relations in $\sigma_D$:

- **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

**Theorem**

QAtTr and QAtc are **decidable** with base FGTGDs and UCQ.

QAlin is **decidable** with base-covered FGTGDs and base-covered UCQ.

We also analyze combined complexity and data complexity, and show that slight changes in the restrictions lead to undecidability.
Theorem

\[ \text{QAt}r(\mathcal{F}_0, \Sigma, Q) \text{ is decidable in } 2\text{EXPTIME combined complexity and PTIME data complexity for base-covered FGTGDs } \Sigma \text{ and base-covered UCQ } Q. \]

Proof idea:

Reduce in PTIME to traditional QA problem \( \text{QA}(\mathcal{F}_0, \Sigma', Q) \) with FGTGDs \( \Sigma' \).
Transitive relations

Theorem

\( \text{QAt}(\mathcal{F}_0, \Sigma, Q) \) is \textbf{decidable} in 2EXPTIME combined complexity and PTIME data complexity for base-covered FGTGDs \( \Sigma \) and base-covered UCQ \( Q \).

Proof idea:

Reduce in PTIME to traditional QA problem \( \text{QA}(\mathcal{F}_0, \Sigma', Q) \) with FGTGDs \( \Sigma' \).

\textbf{Bad news}: we cannot axiomatize transitivity using FGTGDs.
Transitive relations

Theorem

$\text{QAt}(\mathcal{F}_0, \Sigma, Q)$ is decidable in 2EXPTIME combined complexity and PTIME data complexity for base-covered FGTGDs $\Sigma$ and base-covered UCQ $Q$.

Proof idea:

Reduce in PTIME to traditional QA problem $\text{QA}(\mathcal{F}_0, \Sigma', Q)$ with FGTGDs $\Sigma'$.

**Bad news:** we cannot axiomatize transitivity using FGTGDs.

**Good news:** we can approximate transitivity using FGTGD constraints $\Sigma' \supseteq \Sigma$.

If $\mathcal{F}_0 \land \Sigma' \land \neg Q$ is satisfiable, then it has a tree-like witness (a set of facts with a tree decomposition of some bounded tree-width).

**Key technical result:** This tree-like witness can be extended to a set of facts satisfying $\mathcal{F}_0 \land \Sigma \land \neg Q$ where $R \in \sigma_D$ is transitively closed.
Transitive relations

Theorem

\[ \text{QA}_\text{tr}(\mathcal{F}_0, \Sigma, Q) \text{ is decidable} \text{ in 2EXPTIME combined complexity and PTIME data complexity for base-covered FGTGDs } \Sigma \text{ and base-covered UCQ } Q. \]

Proof idea:

Reduce in PTIME to traditional QA problem QA(\mathcal{F}_0, \Sigma', Q) with FGTGDs \Sigma'.

Bad news: we cannot axiomatize transitivity using FGTGDs.

Good news: we can approximate transitivity using FGTGD constraints \( \Sigma' \supseteq \Sigma \).

If \( \mathcal{F}_0 \land \Sigma' \land \neg Q \) is satisfiable, then it has a tree-like witness (a set of facts with a tree decomposition of some bounded tree-width).

Key technical result: This tree-like witness can be extended to a set of facts satisfying \( \mathcal{F}_0 \land \Sigma \land \neg Q \) where \( R \in \sigma_D \) is transitively closed.

(Similar approach for linear order: approximate transitivity and totality.)
## Conclusion

<table>
<thead>
<tr>
<th></th>
<th>QAtr data</th>
<th>QAtc data</th>
<th>QAlin data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BaseFGTGDs</strong></td>
<td>in coNP 2EXP-c</td>
<td>coNP-c 2EXP-c</td>
<td>undecidable</td>
</tr>
<tr>
<td><strong>BaseCovFGTGDs</strong></td>
<td>P-c 2EXP-c</td>
<td>coNP-c 2EXP-c</td>
<td>coNP-c 2EXP-c</td>
</tr>
</tbody>
</table>
## Conclusion

<table>
<thead>
<tr>
<th></th>
<th>QAtr</th>
<th>QAtc</th>
<th>QAlin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>data</td>
<td>combined</td>
<td>data</td>
</tr>
<tr>
<td>BaseFGTGDs</td>
<td>in coNP</td>
<td>2EXP-c</td>
<td>coNP-c</td>
</tr>
<tr>
<td>BaseCovFGTGDs</td>
<td>P-c</td>
<td>2EXP-c</td>
<td>coNP-c</td>
</tr>
</tbody>
</table>

**Also in paper:**

- Generalization to “guarded” logics that include disjunction and some negation (rather than just TGDs);

- Lower bounds for QA,tc and QAlin even with inclusion dependencies (reduction from QA with *disjunctive* inclusion dependencies, using distinguished relations to emulate disjunction).
## Conclusion

<table>
<thead>
<tr>
<th></th>
<th>QA\text{tr}</th>
<th>QA\text{tc}</th>
<th>QA\text{lin}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>data</td>
<td>combined</td>
<td>data</td>
</tr>
<tr>
<td>BaseFGTGDs</td>
<td>in co\text{NP}</td>
<td>2EXP-\text{c}</td>
<td>co\text{NP}-\text{c}</td>
</tr>
<tr>
<td>BaseCovFGTGDs</td>
<td>P-\text{c}</td>
<td>2EXP-\text{c}</td>
<td>co\text{NP}-\text{c}</td>
</tr>
</tbody>
</table>

Also in paper:

- Generalization to “guarded” logics that include disjunction and some negation (rather than just TGDs);

- Lower bounds for QA\text{tc} and QA\text{lin} even with inclusion dependencies (reduction from QA with disjunctive inclusion dependencies, using distinguished relations to emulate disjunction).

Open questions

Is query answering decidable . . . for other special interpretations? when we restrict only to finite sets of facts?
For FGTGD constraints $\Sigma$ and a UCQ $Q$:
if $\mathcal{F}_0 \land \Sigma \land \neg Q$ is satisfiable, then there is a witness $\mathcal{F}$ that has a **tree decomposition** of some bounded tree-width.

A tree decomposition of tree-width $k - 1$
for a set of facts $\mathcal{F} \supseteq \mathcal{F}_0$ is a tree $t$ with each node labelled by a set $S \subseteq \mathcal{F}$ s.t.

- the root is labelled with $\mathcal{F}_0$;
- every fact appears in some node in $t$;
- each non-root node mentions at most $k$ elements;
- for each element, the set of nodes with this element is connected in $t$. 
Tree decompositions

For FGTGD constraints $\Sigma$ and a UCQ $Q$:
if $\mathcal{F}_0 \land \Sigma \land \neg Q$ is satisfiable, then there is a witness $\mathcal{F}$ that has a tree decomposition of some bounded tree-width.

A tree decomposition of tree-width $k - 1$ for a set of facts $\mathcal{F} \supseteq \mathcal{F}_0$ is a tree $t$ with each node labelled by a set $S \subseteq \mathcal{F}$ s.t.

- the root is labelled with $\mathcal{F}_0$;
- every fact appears in some node in $t$;
- each non-root node mentions at most $k$ elements;
- for each element, the set of nodes with this element is connected in $t$. 