Query Answering with Transitive and Linear-Ordered Data

Antoine Amarilli[1], Michael Benedikt[2],
Pierre Bourhis[3] and **Michael Vanden Boom**[2]

[1]LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay
[2]University of Oxford
[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 Σ, boolean query Q (UCQ).

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

Query answering problem (QA)

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

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

Equivalently:

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

Query answering problem (QA)

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

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

Equivalently:

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

Example

$$\begin{aligned}
\mathcal{F}_0 :&\quad S(\mathrm{a},\mathrm{b}), R(\mathrm{b},\mathrm{a}) \\
\Sigma :&\quad \forall xy\,\big(S(x,y) \to R(x,y)\big) \\
&\quad \forall x\,\big(R(x,x) \to \exists y\, T(y)\big) \\
Q :&\quad \exists x\, T(x)
\end{aligned}$$

Q is not certain in general...

Query answering problem (QA)

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

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

Equivalently:

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

Example

$$\begin{aligned}
\mathcal{F}_0 :&\quad S(\mathrm{a}, \mathrm{b}), R(\mathrm{b}, \mathrm{a})\\
\Sigma :&\quad \forall xy\, \big(S(x,y) \rightarrow R(x,y)\big)\\
&\quad \forall x\, \big(R(x,x) \rightarrow \exists y\, T(y)\big)\\
Q :&\quad \exists x\, T(x)
\end{aligned}$$

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

- $\mathcal{ZIQ}$, $\mathcal{ZOQ}$, $\mathcal{ZOI}$ [Calvanese et al., 2009]
- Horn-$\mathcal{SROIQ}$ [Ortiz et al., 2011]
- regular-$\mathcal{EL}^{++}$ [Krötzsch and Rudolph, 2007]

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

QA is undecidable for

- $\mathcal{ALCOIF}^*$ [Ortiz et al., 2010]
- $\mathcal{ZOIQ}$ [Ortiz, 2010]

QA is open for

- $\mathcal{SROIQ}$ and $\mathcal{SHOIQ}$ [Ortiz and Šimkus, 2012]

QA with tuple generating dependencies (a.k.a. existential rules)

TGD: $\forall \boldsymbol{x}\boldsymbol{y} \left(\varphi(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \exists \boldsymbol{z}\, \psi(\boldsymbol{y}, \boldsymbol{z})\right)$

body φ and head ψ are a conjunction of atoms

QA with tuple generating dependencies (a.k.a. existential rules)

TGD: $\forall \boldsymbol{x}\boldsymbol{y}\left(\varphi(\boldsymbol{x},\boldsymbol{y}) \rightarrow \exists \boldsymbol{z}\,\psi(\boldsymbol{y},\boldsymbol{z})\right)$

body φ and head ψ are a conjunction of atoms

Frontier-guarded TGD (FGTGD):
φ includes atom using all of the frontier variables $\boldsymbol{y}$

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

QA with tuple generating dependencies (a.k.a. existential rules)

TGD: $\forall \boldsymbol{x}\boldsymbol{y} \left(\varphi(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \exists \boldsymbol{z}\, \psi(\boldsymbol{y}, \boldsymbol{z})\right)$

body φ and head ψ are a conjunction of atoms

Frontier-guarded TGD (FGTGD):
φ includes atom using all of the frontier variables $\boldsymbol{y}$

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

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

QA with tuple generating dependencies (a.k.a. existential rules)

TGD: $\forall \boldsymbol{x}\boldsymbol{y} \left(\varphi(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \exists \boldsymbol{z}\, \psi(\boldsymbol{y}, \boldsymbol{z}) \right)$

body φ and head ψ are a conjunction of atoms

Frontier-guarded TGD (FGTGD):
φ includes atom using all of the frontier variables $\boldsymbol{y}$

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

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]

QA with tuple generating dependencies (a.k.a. existential rules)

TGD: $\forall \boldsymbol{x}\boldsymbol{y}\, \big(\varphi(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \exists \boldsymbol{z}\, \psi(\boldsymbol{y}, \boldsymbol{z})\big)$

body φ and head ψ are a conjunction of atoms

Frontier-guarded TGD (FGTGD):
φ includes atom using all of the frontier variables $\boldsymbol{y}$

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

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 $\sigma := \sigma_{\mathcal{B}} \sqcup \sigma_{\mathcal{D}}$ where

$\sigma_{\mathcal{D}}$: **distinguished binary relations** with special interpretations (e.g., transitively closed)

$\sigma_{\mathcal{B}}$: **base relations**

We introduce constraint languages that disallow $\sigma_{\mathcal{D}}$-relations as guards:

Base FGTGD: FGTGD where guard for frontier variables is from $\sigma_{\mathcal{B}}$.

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

Our approach

Fix relational signature $\sigma := \sigma_{\mathcal{B}} \sqcup \sigma_{\mathcal{D}}$ where

$\sigma_{\mathcal{D}}$: **distinguished binary relations** with special interpretations (e.g., transitively closed)

$\sigma_{\mathcal{B}}$: **base relations**

We introduce constraint languages that disallow $\sigma_{\mathcal{D}}$-relations as guards:

Base FGTGD: FGTGD where guard for frontier variables is from $\sigma_{\mathcal{B}}$.

$$\forall x\, y_1\, y_2 \left(R(x, y_1) \wedge R(x, y_2) \wedge S(y_1, y_2) \to \exists z \left(R(y_2, z) \wedge T(y_1) \right) \right)$$

Base-covered FGTGD: Base FGTGD where for every $\sigma_{\mathcal{D}}$-atom in the body, there is a $\sigma_{\mathcal{B}}$-atom in the body using its variables.

$$\forall x\, y_1\, y_2 \left(C(x, y_1) \wedge R(x, y_1) \wedge C(x, y_2) \wedge R(x, y_2) \wedge S(y_1, y_2) \to \exists z \left(R(y_2, z) \wedge T(y_1) \right) \right)$$

Our contribution

We consider three different special interpretations for relations in $\sigma_{\mathcal{D}}$:

QAtr	each $R \in \sigma_{\mathcal{D}}$ is transitively closed
QAtc	each $R^+ \in \sigma_{\mathcal{D}}$ is the transitive closure of $R \in \sigma_{\mathcal{B}}$
QAlin	each $R \in \sigma_{\mathcal{D}}$ is a linear order

Our contribution

We consider three different special interpretations for relations in $\sigma_{\mathfrak{D}}$:

QAtr	each $R \in \sigma_{\mathfrak{D}}$ is transitively closed
QAtc	each $R^{+} \in \sigma_{\mathfrak{D}}$ is the transitive closure of $R \in \sigma_{\mathfrak{B}}$
QAlin	each $R \in \sigma_{\mathfrak{D}}$ is a linear order

Theorem

QAtr 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 $\sigma_{\mathfrak{D}}$:

QAtr	each $R \in \sigma_{\mathfrak{D}}$ is transitively closed
QAtc	each $R^+ \in \sigma_{\mathfrak{D}}$ is the transitive closure of $R \in \sigma_{\mathfrak{B}}$
QAlin	each $R \in \sigma_{\mathfrak{D}}$ is a linear order

Theorem

QAtr 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.

Transitive relations

Theorem

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

Proof idea:

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

Transitive relations

Theorem

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

Proof idea:

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

Bad news: we cannot axiomatize transitivity using FGTGDs.

Transitive relations

Theorem

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

Proof idea:

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

Bad news: we cannot axiomatize transitivity using FGTGDs.

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

If $\mathcal{F}_0 \wedge \Sigma' \wedge \neg Q$ 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 $\mathcal{F}_0 \wedge \Sigma \wedge \neg Q$ where $R \in \sigma_{\mathcal{D}}$ is transitively closed.

Transitive relations

Theorem

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

Proof idea:

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

Bad news: we cannot axiomatize transitivity using FGTGDs.

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

If $\mathcal{F}_0 \wedge \Sigma' \wedge \neg Q$ 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 $\mathcal{F}_0 \wedge \Sigma \wedge \neg Q$ where $R \in \sigma_{\mathfrak{D}}$ is transitively closed.

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

Conclusion

	QAtr		QAtc		QAlin	
	data	**combined**	**data**	**combined**	**data**	**combined**
BaseFGTGDs	in coNP	2EXP-c	coNP-c	2EXP-c	undecidable	
BaseCovFGTGDs	P-c	2EXP-c	coNP-c	2EXP-c	coNP-c	2EXP-c

Conclusion

	QAtr		QAtc		QAlin	
	data	combined	data	combined	data	combined
BaseFGTGDs	in coNP	2EXP-c	coNP-c	2EXP-c	undecidable	
BaseCovFGTGDs	P-c	2EXP-c	coNP-c	2EXP-c	coNP-c	2EXP-c

Also in paper:

- generalization to “guarded” logics that include disjunction and some negation (rather than just TGDs);
- lower bounds for QAtc and QAlin even with inclusion dependencies (reduction from QA with *disjunctive* inclusion dependencies, using distinguished relations to emulate disjunction).

Conclusion

	QAtr		QAtc		QAlin	
	data	**combined**	**data**	**combined**	**data**	**combined**
BaseFGTGDs	in coNP	2EXP-c	coNP-c	2EXP-c	undecidable	
BaseCovFGTGDs	P-c	2EXP-c	coNP-c	2EXP-c	coNP-c	2EXP-c

Also in paper:

- generalization to "guarded" logics that include disjunction and some negation (rather than just TGDs);
- lower bounds for QAtc and QAlin 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?

Tree decompositions

For FGTGD constraints Σ and a UCQ Q:
if $\mathcal{F}_0 \wedge \Sigma \wedge \neg Q$ 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 Σ 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.