Block-Sorted Quantified Conjunctive Queries

Hubie Chen[1] Dániel Marx[2]

[1]Universidad del País Vasco and IKERBASQUE, San Sebastián, Spain

[2]Institute for Computer Science and Control, Hungarian Academy of Sciences (MTA SZTAKI) Budapest, Hungary

40th International Colloquium on Automata, Languages and Programming (ICALP 2013)
Riga, Latvia
July 10, 2013

First-order model checking

We study the following problem:

FO MODEL CHECKING (FO-MC)	
Input:	first-oder formula ϕ, relational structure $\mathbf{A}$
Question:	does $\mathbf{A} \models \phi$?

Can model

- natural algorithmic problems (e.g., finding a k-clique),
- constraint satisfaction problems,
- database queries.

Bad news: The problem is PSPACE-complete in general.

First-oder model checking

For fixed ϕ, FO-MC(ϕ) is polynomial-time solvable, but exponent depends on the number of variables in ϕ.

The quest:

Find tractable fragments of FO-MC.

First-oder model checking

For fixed ϕ, FO-MC(ϕ) is polynomial-time solvable, but exponent depends on the number of variables in ϕ.

The quest:

Find tractable fragments of FO-MC.

- Find classes Φ of first-order formulas for which FO-MC(Φ) is polynomial-time solvable.
- Find classes Φ of first-order formulas for which FO-MC(Φ) is fixed-parameter tractable, e.g., can be solved in time $f(|\phi|) \cdot \|\mathbf{A}\|^{O(1)}$.
  - **Motivation:** in database queries, the query ϕ has small size, while $\mathbf{A}$ is large.

Existential conjunctive queries

We consider first sentences ϕ of the form

$$\exists x_1, x_2, x_3, x_4 : R_1(x_1, x_3) \wedge R_2(x_1, x_2, x_4) \wedge R_1(x_1, x_4),$$

that is, existential quantification followed by conjunction of atoms.

Existential conjunctive queries

We consider first sentences ϕ of the form

$$\exists x_1, x_2, x_3, x_4 : R_1(x_1, x_3) \wedge R_2(x_1, x_2, x_4) \wedge R_1(x_1, x_4),$$

that is, existential quantification followed by conjunction of atoms.

The formula can be described by a relational structure **A**.

Observation

B $\models \exists$**A** if and only if there is a homomorphism from **A** to **B**.

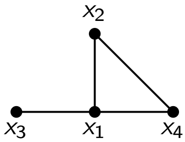
Task: classify which classes $\mathcal{A}$ of relational structures make the problem fixed-parameter tractable parameterized by the size of the query.

Graph-based view

Gaifman graph of a relational structure: two elements are adjacent if there is a relation containing a tuple containing both elements.

This way, with every existential conjunctive query $\exists A$, we can associate a graph.

$$\exists x_1, x_2, x_3, x_4 : R_1(x_1, x_3) \wedge R_2(x_1, x_2, x_4) \wedge R_1(x_1, x_4),$$

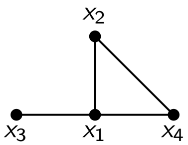


Graph-based view

Gaifman graph of a relational structure: two elements are adjacent if there is a relation containing a tuple containing both elements.

This way, with every existential conjunctive query $\exists A$, we can associate a graph.

$$\exists x_1, x_2, x_3, x_4 : R_1(x_1, x_3) \wedge R_2(x_1, x_2, x_4) \wedge R_1(x_1, x_4),$$



Same graph for

$$\exists x_1, x_2, x_3, x_4 : R(x_1, x_3) \wedge R(x_1, x_2, x_4) \wedge R(x_1, x_4)$$
$$\exists x_1, x_2, x_3, x_4 : R_1(x_1, x_3) \wedge R_1(x_1, x_3) \wedge R_2(x_1, x_4) \wedge R_3(x_2, x_4)$$

Two different views

Task: classify which classes $\mathcal{A}$ of relational structures make the problem fixed-parameter tractable parameterized by the size of the query.

- **Graph-based view:** For which classes $\mathcal{G}$ of graphs is the problem fixed-parameter tractable?
  (coarser view)
- **Structure-based view:** For which classes $\mathcal{A}$ of relational structures is the problem fixed-parameter tractable?
  (finer view)

Graph-based view

Complete characterization of graph classes that guarantee tractability:

Theorem [Grohe, Schwentick, Segoufin 2001]

Let $\mathcal{G}$ be a class of graphs.

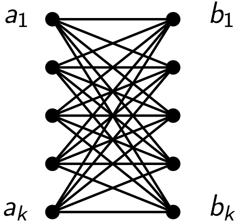
- If $\mathcal{G}$ has bounded treewidth, then EC-MO($\mathcal{G}$) is polynomial-time solvable.
- If $\mathcal{G}$ has unbounded treewidth, then EC-MO($\mathcal{G}$) is W[1]-hard.

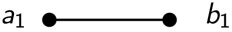
(The negative result is based on the Excluded Grid Theorem.)

Structure-based view

The graph-based view does not reveal some tractable cases as it bundles them together width hard cases.

$\mathbf{A}_1$		$\mathbf{A}_2$
$\exists a_1, \ldots, a_k, b_1, \ldots, b_k : \bigwedge_{1 \le i,j \le k} R(a_i, b_j)$	$\Longleftrightarrow$	$\exists a_1, b_1 : R(a_1, b_1)$





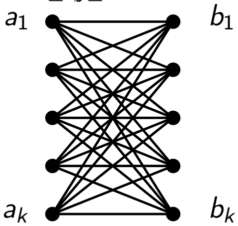
Structure-based view

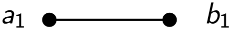
The graph-based view does not reveal some tractable cases as it bundles them together width hard cases.

$\mathbf{A}_1$

$$\exists a_1, \dots, a_k, b_1, \dots, b_k : \bigwedge_{1 \leq i,j \leq k} R(a_i, b_j) \iff \exists a_1, b_1 : R(a_1, b_1)$$

$\mathbf{A}_2$





Why are these two formulas equivalent?

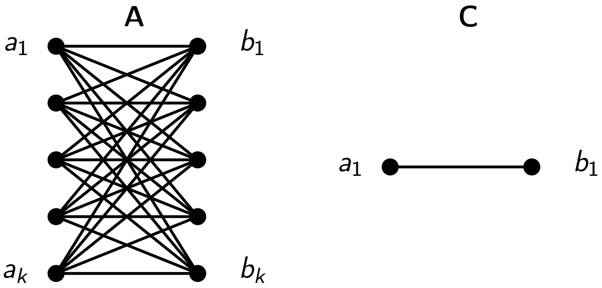
- **$\mathbf{A}_1$** implies **$\mathbf{A}_2$**: query **$\mathbf{A}_2$** is a substructure of **$\mathbf{A}_2$**.
- **$\mathbf{A}_2$** implies **$\mathbf{A}_1$**: there is a homomorphism from **$\mathbf{A}_1$** to **$\mathbf{A}_2$**.

Cores

Definition

A substructure **C** of **A** is a core of **A** if

- there is a homomorphism from **A** to **C**, and
- there is no homomorphism from **C** to a proper substructure of **C**.



- The core of **A** is unique up to isomorphism.
- If **C** is a core of **A**, then the queries $\exists$**A** and $\exists$**C** are equivalent.

Structure-based view

To understand complexity of existential conjunctive queries, we need to look at the cores of the structures.

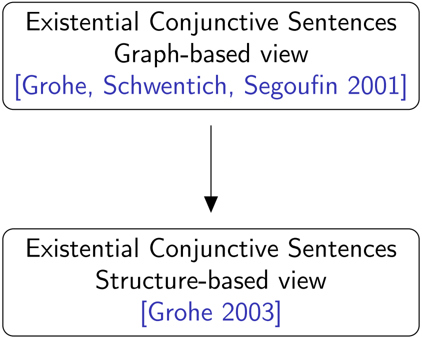
Complete characterization of classes of relational structures that guarantee tractability:

Theorem [Grohe 2003]

Let $\mathcal{A}$ be a class of relational structures of bounded arity.

- If the cores of the structures in $\mathcal{A}$ have bounded treewidth, then EC-MC($\mathcal{A}$) is polynomial-time solvable.
- If the cores of the structures in $\mathcal{A}$ have unbounded treewidth, then EC-MC($\mathcal{A}$) is W[1]-hard.

Classification results (bounded arity)



Quantified Conjunctive Model Checking

Let us look at more general quantified conjunctive sentences:

$$\exists x_1 \forall y_1, y_2 \exists x_2 : R_1(x_1, y_1) \wedge R_2(x_2, y_2) \wedge R_3(x_1, y_2)$$

The query can be described by a pair $(P, \mathbf{A})$ where

- P is the quantifier prefix (ordering and type of variables), and
- $\mathbf{A}$ is a relational structure.

Quantified Conjunctive Model Checking

Let us look at more general quantified conjunctive sentences:

$$\exists x_1 \forall y_1, y_2 \exists x_2 : R_1(x_1, y_1) \wedge R_2(x_2, y_2) \wedge R_3(x_1, y_2)$$

The query can be described by a pair $(P, \mathbf{A})$ where

- P is the quantifier prefix (ordering and type of variables), and
- $\mathbf{A}$ is a relational structure.

Again two questions of structural characterization:

- **Graph-based view:** characterize the sets $\mathcal{G}$ of prefixed graphs (P, G) such that restriction to $\mathcal{G}$ is tractable.
- **Structure-based view:** characterize the sets $\mathcal{A}$ of prefixed structures $(P, \mathbf{A})$ such that restriction to $\mathcal{A}$ is tractable.

Note: the problem is PSPACE-hard already for trees!

Quantified Conjunctive Model Checking

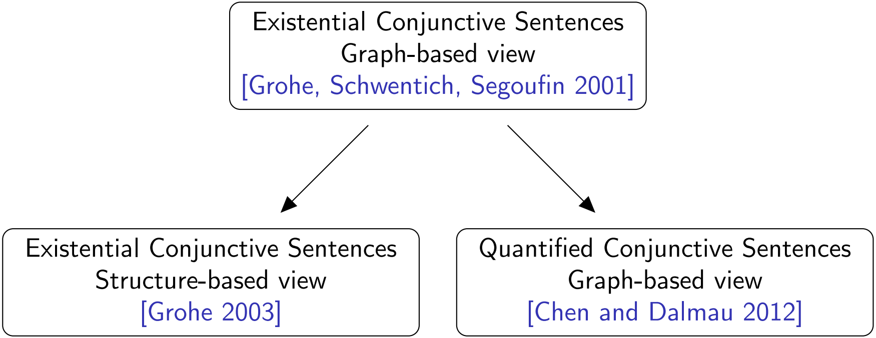
[Chen and Dalmau 2012] introduced a notion of width for prefixed graphs that generalizes treewidth ($\text{width}((\exists, G)) = \text{tw}(G)$).

Theorem [Chen and Dalmau 2012]

Let $\mathcal{G}$ be a class of prefixed graphs.

- If $\mathcal{G}$ has bounded width, then QC-MC($\mathcal{G}$) is polynomial-time solvable.
- If $\mathcal{G}$ has unbounded width, then QC-MC($\mathcal{G}$) is W[1]- or coW[1]-hard.

Classification results (bounded arity)



Quantified Conjunctive Sentences — Structure-based view

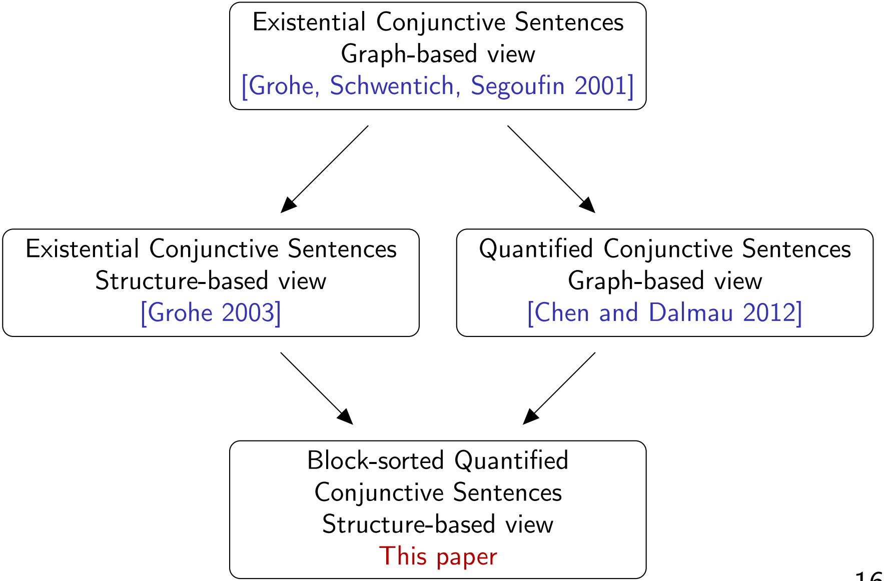
Natural next step: structured-based view for quantified conjunctive sentences.

We focus on a restricted, but fairly robust version: block-sorted quantified formulas.

$$(\exists \overbrace{x_1 x_2 x_3}^{S_1} \overbrace{x_4 x_5}^{S_2} \forall \overbrace{y_1 y_2}^{S_3} \exists \overbrace{x_6 x_7}^{S_4} \overbrace{x_8 x_9}^{S_5}, \mathbf{A})$$

- The conjunctive query setting of [Grohe 2003] can be thought of as a query with a single existential sort.
- The graph-based view of [Chen and Dalmau 2012] for quantified formulas can be thought of as having a separate sort for each variable.

Classification results (bounded arity)



Main result

Theorem [this paper]

Let $\mathcal{A}$ be a class of relational structures.

- If $\mathcal{A}$ has property X, then QC-MC($\mathcal{A}$) is FPT.
- If $\mathcal{A}$ does not have property X, then QC-MC($\mathcal{A}$) is W[1]- or coW[1]-hard.

Main result

Theorem [this paper]

Let $\mathcal{A}$ be a class of relational structures.

- If $\mathcal{A}$ has property X, then QC-MC($\mathcal{A}$) is FPT.
- If $\mathcal{A}$ does not have property X, then QC-MC($\mathcal{A}$) is W[1]- or coW[1]-hard.

What is this property X?

The "core" (in an appropritate sense) of every structure has bounded width (in the sense of [Chen and Dalmau 2012]).

Cores for block-sorted quantified formulas

What is the right notion of core?

Problem 1:

Recall:
If there is a homomorphism from $\mathbf{A}$ to $\mathbf{B}$, then $\exists\mathbf{B}$ implies $\exists\mathbf{A}$.

No longer true for quantified formulas:

$\forall y_1 \exists x_1 : R(x_1, y_1, y_1)$ does not imply $\forall y_1, y_2 \exists x_1 : R(x_1, y_1, y_2)$.

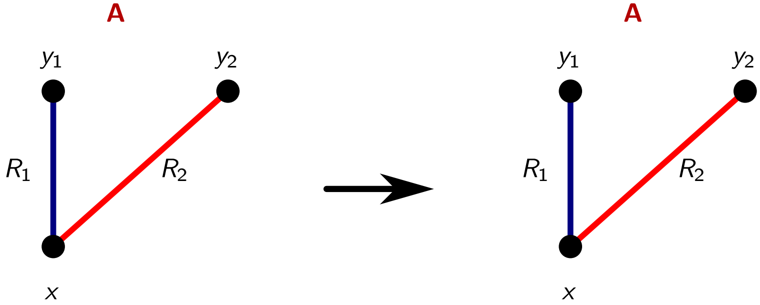
Lemma

If there is a homomorphishm from $\mathbf{A}$ to $\mathbf{B}$ that is *injective on the universal sorts,* then $(P, \mathbf{B})$ implies $(P, \mathbf{A})$.

Cores for block-sorted quantified formulas

Problem 2:

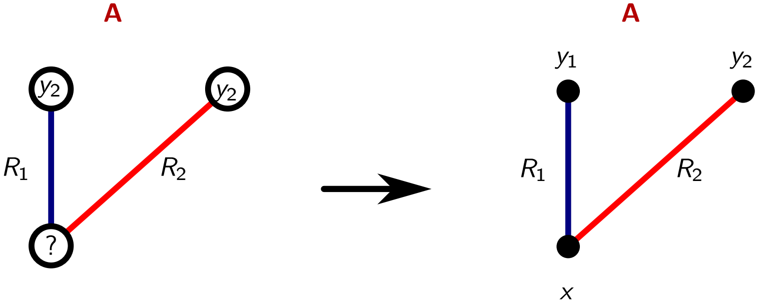
$\exists\mathbf{A}$ is trivially true on $\mathbf{A}$. But in general, $(P, \mathbf{A})$ is not true on $\mathbf{A}$!



Cores for block-sorted quantified formulas

Problem 2:

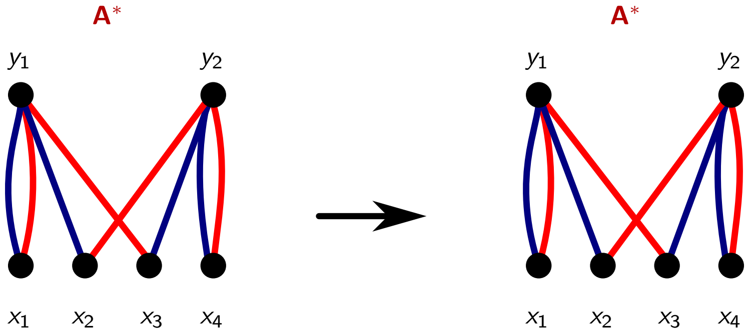
$\exists \mathbf{A}$ is trivially true on $\mathbf{A}$. But in general, $(P, \mathbf{A})$ is not true on $\mathbf{A}$!



Cores for block-sorted quantified formulas

Problem 2:

$\exists \mathbf{A}$ is trivially true on $\mathbf{A}$. But in general, $(P, \mathbf{A})$ is not true on $\mathbf{A}$!



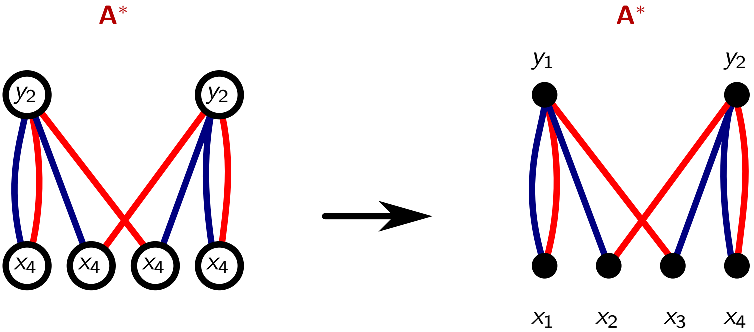
However, we can create an $\mathbf{A}^*$ such that

- $(P, \mathbf{A})$ and $(P^*, \mathbf{A}^*)$ are logically equivalent, and
- $\mathbf{A}^* \models (P^*, \mathbf{A}^*)$.

Cores for block-sorted quantified formulas

Problem 2:

$\exists \mathbf{A}$ is trivially true on $\mathbf{A}$. But in general, $(P, \mathbf{A})$ is not true on $\mathbf{A}$!



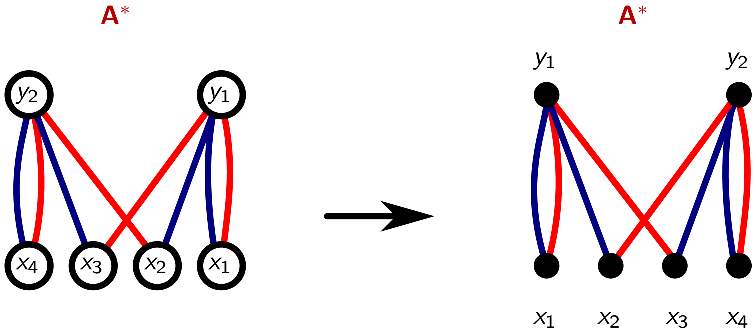
However, we can create an $\mathbf{A}^*$ such that

- $(P, \mathbf{A})$ and $(P^*, \mathbf{A}^*)$ are logically equivalent, and
- $\mathbf{A}^* \models (P^*, \mathbf{A}^*)$.

Cores for block-sorted quantified formulas

Problem 2:

$\exists\mathbf{A}$ is trivially true on $\mathbf{A}$. But in general, $(P, \mathbf{A})$ is not true on $\mathbf{A}$!



However, we can create an $\mathbf{A}^*$ such that

- $(P, \mathbf{A})$ and $(P^*, \mathbf{A}^*)$ are logically equivalent, and
- $\mathbf{A}^* \models (P^*, \mathbf{A}^*)$.

Cores for block-sorted quantified formulas

We can define the core $(P, \mathbf{A})$ as a $(P^*, \mathbf{C})$ such that

- $(P, \mathbf{A})$ and $(P^*, \mathbf{C})$ are logically equivalent,
- $\mathbf{C} \models (P^*, \mathbf{C})$, and
- there is no homomorphism injective on the universal sorts from $\mathbf{C}$ to a proper substructure of $\mathbf{A}$.

The tractability criterion is essentially whether these cores have bounded treewidth in the sense of [Chen and Dalmau 2012].

Classification results (bounded arity)