Approximation and semantic tree-width of conjunctive regular path queries

Rémi Morvan

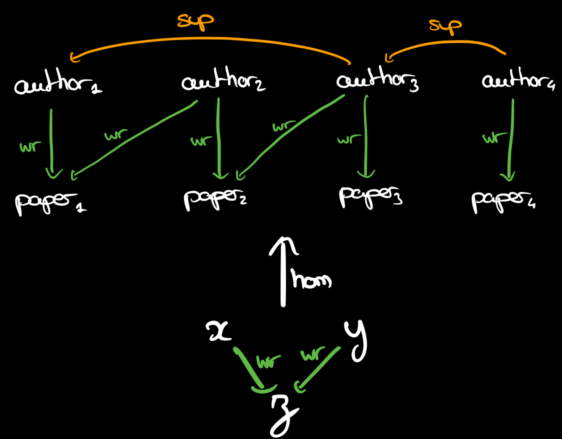
joint work with Diego Figueira

LABRI, U. Bordeaux

9 December 2022

LINKS seminar, Lille

(Graph) databases



Conjunctive queries (CQs)

$$\gamma(x,y) = \exists z.\ x \xrightarrow{wr} z \wedge y \xrightarrow{wr} z$$

Evaluation:
(author$_1$, author$_2$),
(author$_1$, author$_1$),
etc...

⇝ homomorphism semantic

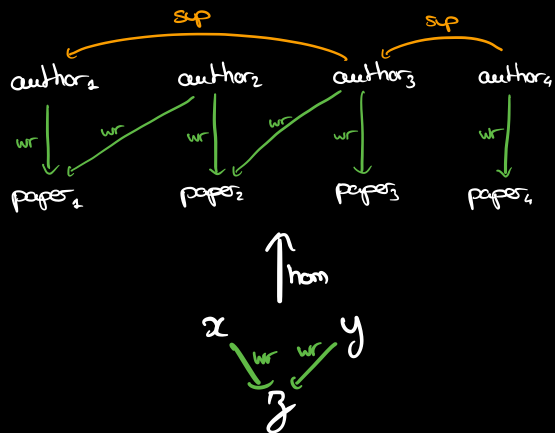
Prop Evaluation of CQs is **NP-complete**...
(combined complexity:
input: database & query)

Upper bound:

Input: $\gamma(\bar{x})$ CQ
G database
$\bar{a}$ tuple in G

Algo: Guess f: vars$(\gamma) \longrightarrow G$
& check that $\begin{cases} f(\bar{x}) = \bar{a} \\ f \text{ homomorphism.} \end{cases}$

(Graph) databases



Prop Evaluation of CQs is NP-complete ...
(combined complexity:
input: database & query)

Conjunctive queries (CQs)

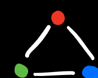
$$\gamma(x,y) = \exists z.\ x \xrightarrow{wr} z \wedge y \xrightarrow{wr} z$$

Evaluation:
(author$_1$, author$_2$),
(author$_1$, author$_1$),
etc...

⇝ homomorphism semantic

Lower bound:

$$\text{3-COL} \leq_p \text{CQ-EVAL}$$

G 3-colorable? $\mapsto$ G $\xrightarrow{hom}$  ?

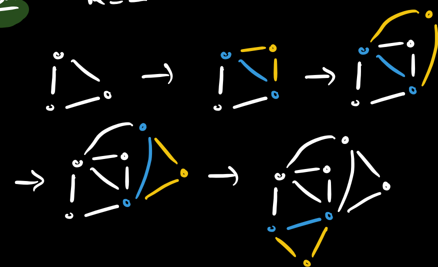
↑ Boolean query ↑ graph database

One solution: tree-width

Def k-trees:

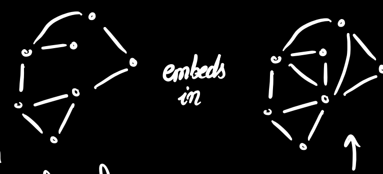
- start with a (k+1)-clique
- repeat:
  add a new node, and join it to a k-clique.

Ex k=2



Def G has tree-width ≤ k if we can embed it in a k-tree. ← remove nodes & edges

Ex



↑ graph of tree-width 2

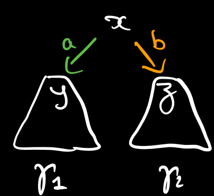
↑ 2-tree

Tree-width and CQs

Thm [Folklore] Fix k. Evaluation of CQs of tree-width $\leq k$ is PTIME. (in combined complexity)

<u>Proof</u> ($k=1$)

$\gamma =$



EVAL(γ, G):
for $u \in G$:
$\exists$? $v \in G$ st $u \xrightarrow{a} v$
and EVAL(γ_1, G)
AND
$\exists$? $w \in G$ st $u \xrightarrow{b} v$
and EVAL(γ_2, G)

Complexity:

$|G|^2 \cdot (\|\text{EVAL}(\gamma_1, G)\| + \|\text{EVAL}(\gamma_2, G)\| + cst)$

$\rightsquigarrow O(|G|^2 \cdot \|\gamma\|)$

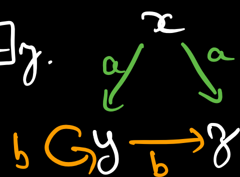
For $k \geq 1$: $O(|G|^{k+1} \cdot \|\gamma\|)$.

Semantic tree-width

same evaluation on every database

Q°: Given $\gamma(\bar{x})$ CQ, is $\gamma(\bar{x})$ equivalent to a CQ $\gamma'(\bar{x})$ of $tw \leq k$?

Ex $\gamma = \exists x \, \exists y. \exists z.$



$\equiv \gamma' = \exists x. \exists y.$

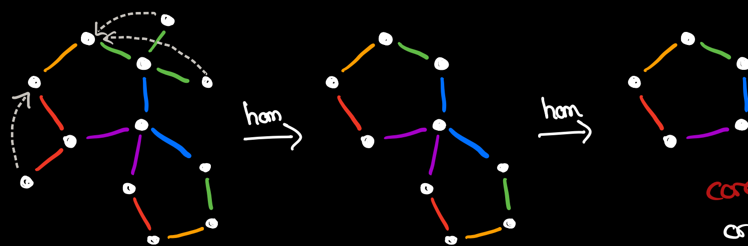
$\rightsquigarrow$ $\gamma(x)$ has tree-width 2

but semantic tree-width 1 !

Minimisation of CQs

Thm [Folklore] Every CQ admits a unique minimal equivalent CQ.
(minimal: for the number of variables.)

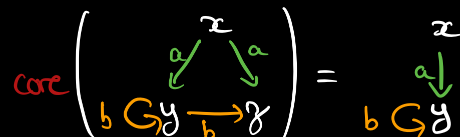
Def Core of an graph (edge-labelled graph): minimal retraction



core of the original graph

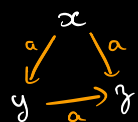
Minimisation of CQs = Core of graphs.

Ex



CQs (not) of small semantic tree-width

Prop γ has semantic tw $\leq k$ IFF core(γ) has tw $\leq k$

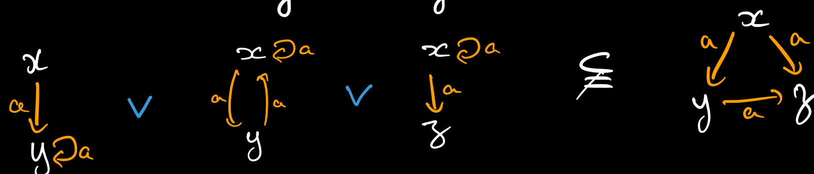


is minimal ...

What if we really want a query of tree-width 1?

Def-Prop [Barceló, Libkin & Romero - PODS '12]
Given a CQ $\gamma(x)$, the union of all hom. images of tw ≤ 1 of $\gamma(x)$ is the maximal under-approximation of $\gamma(x)$ by unions of CQs of tw ≤ 1.

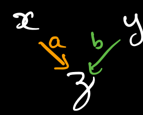
Ex



Conjunctive queries: summary

Evaluating CQs $\cong$ Homomorphism problem

$\exists x, y, z, \; x \xrightarrow{a} z \wedge y \xrightarrow{b} z$ in G? $\iff$



$\xrightarrow{\text{hom}?}$ G

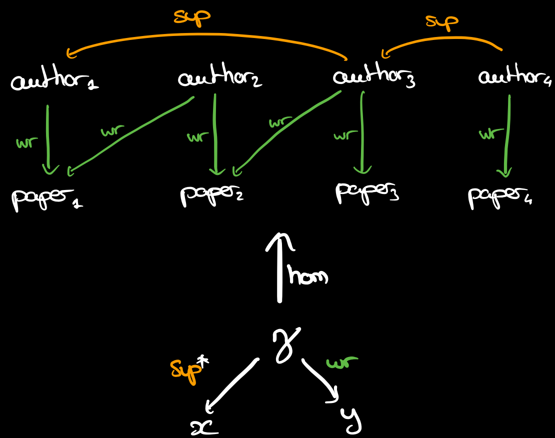
NP-complete

PTIME when restricted to queries of $tw \leq k$

Given a query, we can EFFECTIVELY decide if it has semantic tree-width $\leq k$ (minimisation / core)

We can also approximate queries by union of CQs of $tw \leq k$

Path queries



Prop Evaluation of CRPQs is NP-complete...

(combined complexity:
input: database & query)

Conjunctive regular path queries (CRPQs)

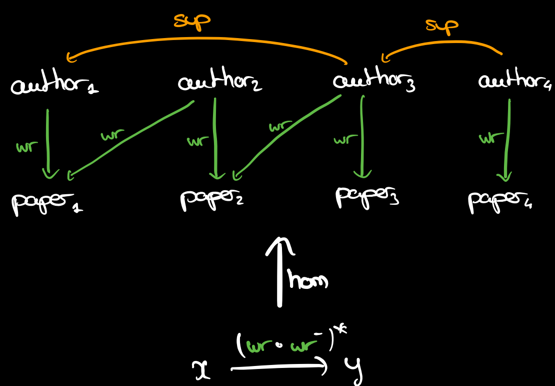
Atoms: $x \xrightarrow{L} y$ ← regular lang on {wr, sup}

$$\gamma(x,y) = \exists z.\ z \xrightarrow{sup^*} x \wedge z \xrightarrow{wr} y$$

⇝ "homomorphism" semantic

Evaluation:
(author1, paper4),
(author4, paper4),
...

Path queries (cont.)



Prop Evaluation of C2RPQs is
- NP-complete…
- PTIME when restricted to queries of tw ≤ k.

Conjunctive 2-way regular path queries (C2RPQs)

__Def__ $x \xrightarrow{a^-} y$ holds iff $y \xrightarrow{a} x$.

Atoms: $x \xrightarrow{L} y$ ← regular lang on $\{wr, sup, wr^-, sup^-\}$

__Ex__ $\gamma(x,y) = x \xrightarrow{(wr \cdot wr^-)^*} y$

⇝ "homomorphism" semantic

__Evaluation__:

(author$_1$, author$_3$),

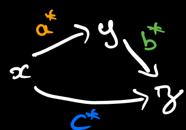
…

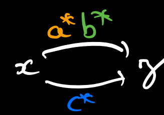
Prop Equivalence of C2RPQs is decidable.

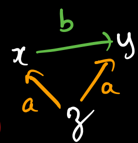
↗ ExpSpace - complete

Semantic tree-width

Q: Given a C(2)RPQ, when is it equivalent to a C(2)RPQ of tree-width $\leq k$?

Ex $\gamma(x,y) = \exists y.$ 

$\equiv$ $\gamma'(x,y) =$ 

Ex $\delta(x) = \exists y, z.$ 

$\equiv$ $\delta'(x) =$ 

(minimal CQ)

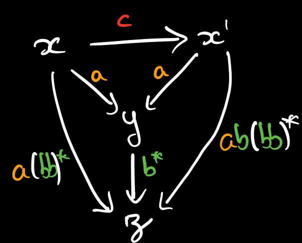
Pb C(2)RPQ cannot be minimised...

Union

Fact For CQs, γ is equivalent to a CQ of $tw \leq k$ iff γ is equivalent to a union of CQs of $tw \leq k$

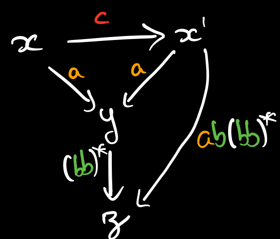
For CRPQs this is (probably) false...

Ex $\gamma(x, x', y, z) \doteq$

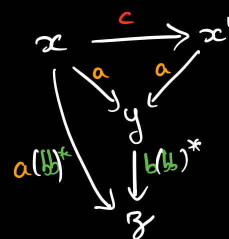


has tree-width 3...

$\equiv$



$\vee$

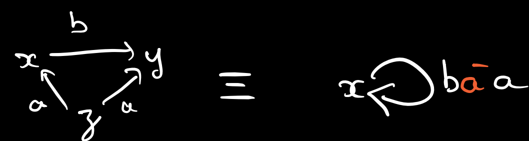


union of CRPQs tree-width ≤ 2

Deciding semantic tree-width

Def A UC2RPQ Γ has semantic tree-width $\leqslant k$ if it is equivalent to a UC2RPQ of $tw \leqslant k$.

Ex $\gamma(x) = \exists y, z.$



↑ sem. tw $\leqslant 1$

DECIDING SEMANTIC TREE-WIDTH :

Input: Γ

Q°: Γ has sem tw $\leqslant k$? ← fixed

Motiva°:

UC2RPQs of $tw \leqslant k$ can be evaluated in PTIME!

Deciding semantic tree-width (cont.)

DECIDING SEMANTIC TREE-WIDTH:
Input: Γ
Q°: Γ has sem tw $\leq k$? ← fixed

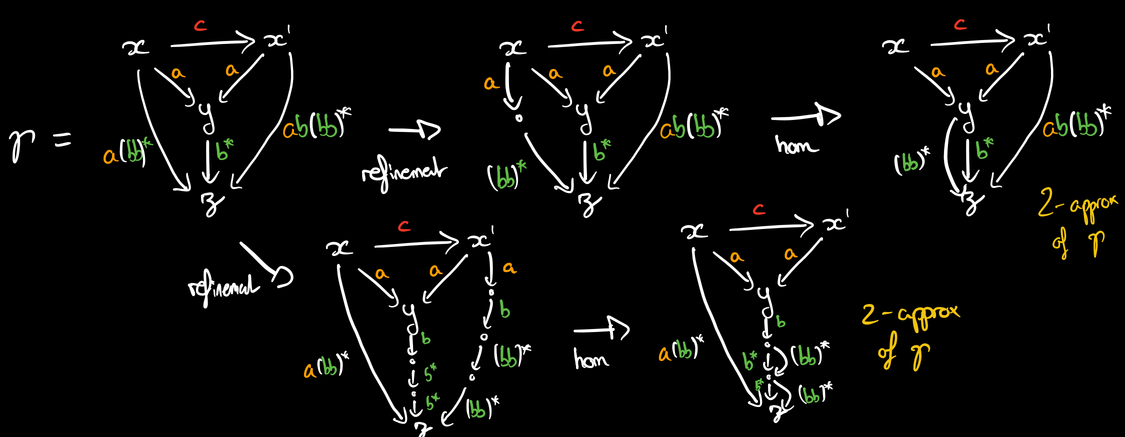
Motiva°:
UC2RPQs of tw $\leq k$
can be evaluated in PTIME!

- DECIDABLE & EFFECTIVE for UC2RPQs when $k = 1$ [Barceló, Romero & Vardi, PODS '13]
  ↖ ExpSpace-complete
- DECIDABLE & EFFECTIVE for UC2RPQs when $k \geq 2$ [Figueira, M., ICDT '23] & [Feier, Gogacz, Murlak, unpublished]
  ↖ 2ExpSpace
  likely ExpSpace-complete

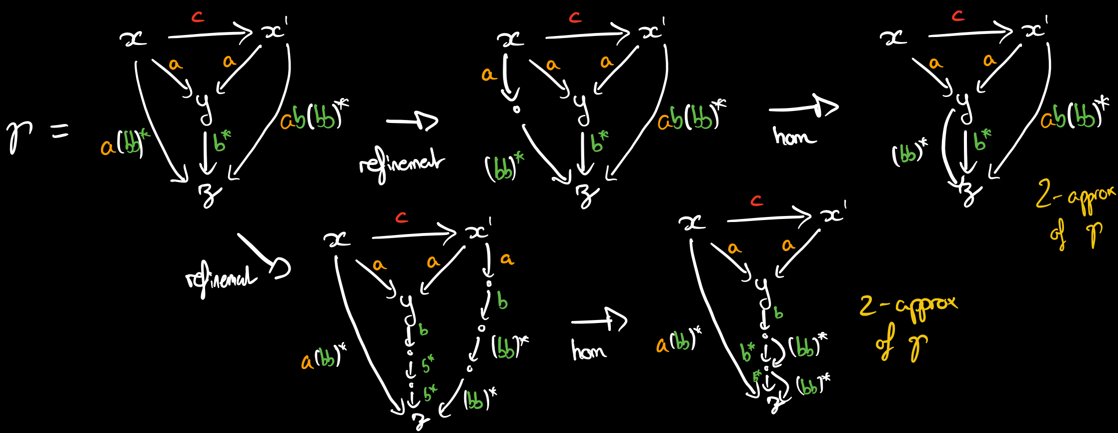
The case $k = 1$ and $k \geq 2$ seem very different...

Deciding semantic tree-width ($k \geq 2$)

Idea: Start with a UC2RPQ,
select any C2RPQ in the union and "refine" it,
then fold it → if it has $tw \leq k$, it is a k-approxima°
= surjective homomorphism



Deciding semantic tree-width ($k \geq 2$)



We obtain an infinite set of k-approximations.

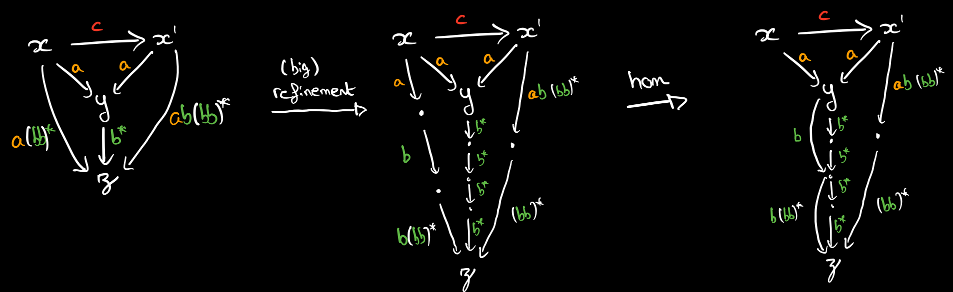
"Key Lemma" [Figueira, M., ICDT '23] This infinite set of C2RPQs is effectively expressible as a UC2RPQ.

→ Test if this UC2RPQ is equivalent to the original one.

The key lemma

"Key Lemma" [Figueira, M., ICDT '23] This infinite set of C2RPQs is effectively expressible as a UC2RPQ.

Proof idea Bound on size of refinement ⇒ Bound on number of k-approximations



big 2-approx

Look at a "tree decomposition" of the approximation, look where long path are sent, massage it → TADA!

Simple in principle / Detailed proof ≈ 10 pages

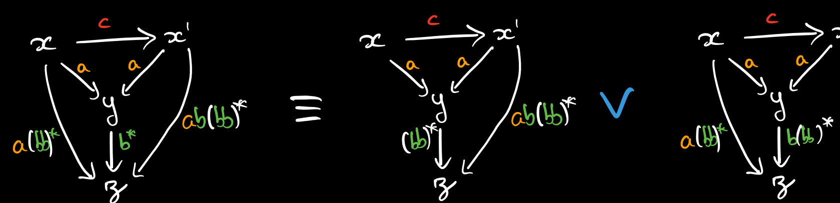
Properties of semantic tree-width

Theorem [Figueira, M., ICDT '23] T: UC2RQ, $k \geq 2$. TFAE:

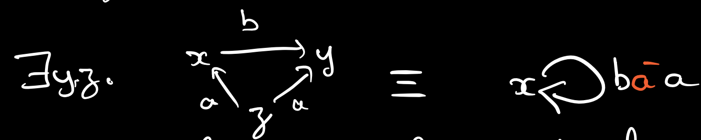
1) T is equivalent to an infinite union of CRPQs of $tw \leq k$
2) T is equivalent to a UC2RPQ of $tw \leq k$
3) T is equivalent to an infinite union of CQs of $tw \leq k$.

⊕ Closure property on the regular languages.

Ex $k=2$



$k=1$



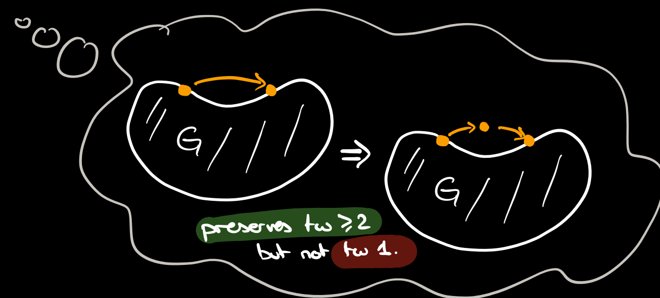
not expressible as an infinite set of CQs of $tw \leq 1$.

Semantic tree-width : overview

The following are decidable:

CQ equivalent to a CQ of tw ≤ k ? } minimisation or approxim°
UCQ —''— UCQ —''—
UC2RPQ —''— UC2RPQ —''— } approxim°

For UC2RPQ, the case $k=1$ and $k \geq 2$ seem to be very different problems...



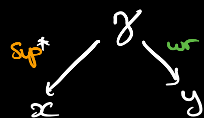
Simple regular expressions

2ExpSpace algo for deciding sem tw $\leq k$

ExpSpace claimed by Feier, Gogacz & Murlak

Simple regular expressions: $a_1 + a_2 + \dots + a_k$ or a_i^*.

UC2RPQ(SRE) :



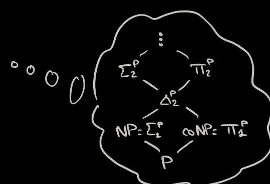
, etc.

75% of all path queries "from real life"

[Bonifati, Martens, Timm, 2020]

Theorem [Figueira, M., ICDT '23]

Semantic tree-width $\leq k$ is in Π_2^p over UC2RPQ(SRE).



A glimpse beyond ...

Query of sem tw ≤ k → Compute equivalent query of tw ≤ k → Evaluate it

FPT (in $|\Gamma|$) algo for evaluation of queries of sem tw ≤ k.

$$O(f(|\Gamma|) \cdot |G|^{k+1})$$

[Romero, Barceló, Vardi, LICS 2017]
improved in [Figueira, M., ICDT 2023]

Thm [Grohe, FoCS 2003] $\mathcal{C}$: r.e. class of CQs

Evaluation of $\mathcal{C}$ is FPT

IFF

Evaluation of $\mathcal{C}$ is PTIME

IFF

$\mathcal{C}$ has bounded sem tree-width.

assuming FPT ≠ W[1]

Open question:

Let $\mathcal{C}$ be a class CRPQs / UC2RPQs.

Evaluation of $\mathcal{C}$ is FPT

IFF ?

$\mathcal{C}$ has bounded sem tree-width