Approximation and semantic tree-width of conjunctive regular path queries

Rémi Morvan

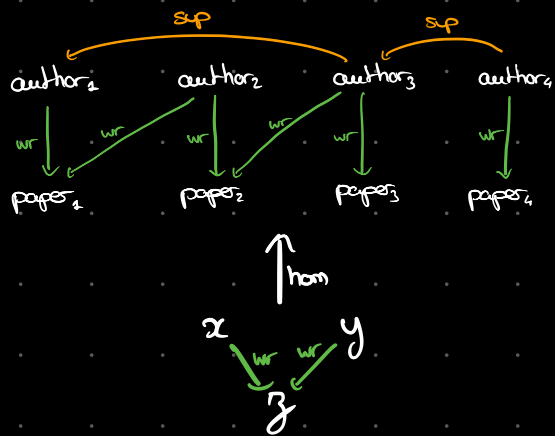
joint work with
Diego Figueira

LABRI, U. Bordeaux

6 December 2022

M2F seminar,
Bordeaux

(Graph) databases



Conjunctive queries (CQs)

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

Evaluation:
(author1, author2),
(author1, author1),
etc...

$\leadsto$ homomorphism semantic

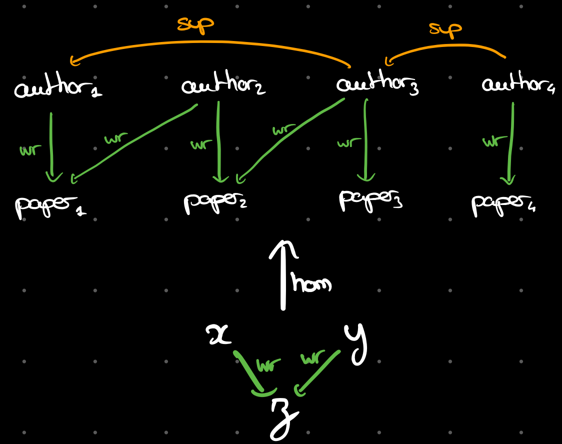
Prop Evaluation of CQs is **NP-complete**...
(combined complexity:
input: database & query)

Upper bound:

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

Algo: Guess f: vars(γ) $\longrightarrow G$
& 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)

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

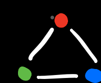
Evaluation:

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

$\rightsquigarrow$ homomorphism semantic

Lower bound:

3-COL $\leq_p$ CQ-EVAL

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

Boolean query

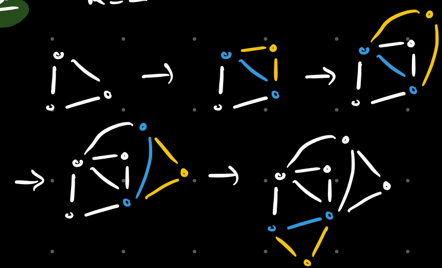
graph database

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.

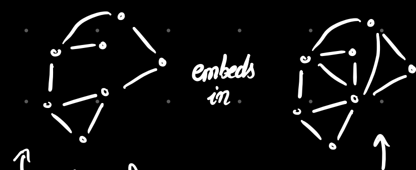
Ex $k=2$



Def G has tree-width $\leq k$ if we can embed it in a k-tree.

← remove nodes & edges

Ex



↑ graph of tree-width 2

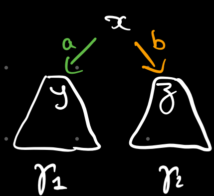
↑ 2-tree

Tree-width and CQs

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

Proof ($k=1$)

$\gamma =$



$\text{EVAL}(\gamma, G)$:

for $u \in G$:

- ∃? $v \in G$ st $u \xrightarrow{a} v$ and $\text{EVAL}(\gamma_1, G)$

  AND

  ∃? $w \in G$ st $u \xrightarrow{w} v$ and $\text{EVAL}(\gamma_2, G)$

Complexity:

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

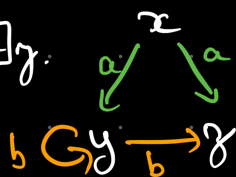
$$\rightsquigarrow O(|G|^2 \cdot \|\gamma\|)$$

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

Semantic tree-width

same evaluation on every database

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

<u>Ex</u> $\gamma = \exists x\, \exists y. \exists z.$ 

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

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

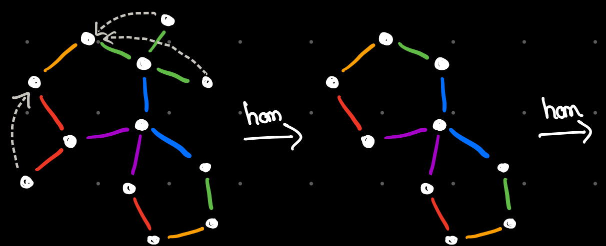
but semantic tree-width 1 !

Minimisation of CQs

Thm [Folklore] Every CQ admits a unique minimal equivalent CQ.

for the number of variables.

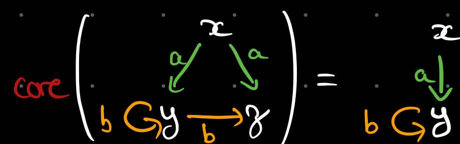
Def Core of an $\underbrace{\text{graph}}^{\text{edge-labelled}}$: minimal retraction



core of the original graph

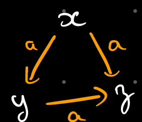
Minimisation of CQs = Core of graphs.

Ex



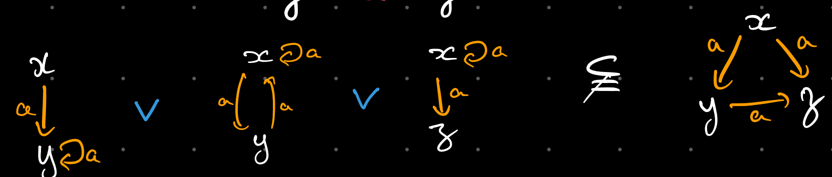
CQs (not) of small semantic tree-width

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 retractions of tw ≤ 1 of $\gamma(x)$ is the maximal under-approxim° of $\gamma(x)$ by unions of CQs of tw ≤ 1.

Ex



Conjunctive queries: summary

Evaluating CQs $\cong$ Homomorphism problem

$$\exists x, y, z, \; \begin{array}{l} x \xrightarrow{a} z \\ \wedge \; y \xrightarrow{b} z \end{array} \text{ in } G \,? \quad \Longleftrightarrow \quad \begin{tikzcd} x \arrow[dr, "a"] & & y \arrow[dl, "b"'] \\ & z & \end{tikzcd} \xrightarrow{\text{hom}\,?} G$$

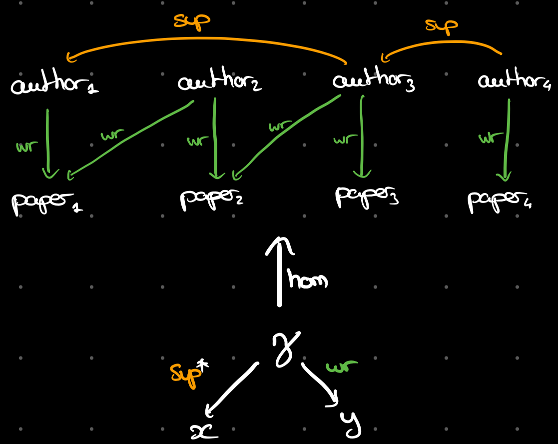
NP-complete

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)

regular lang on {wr, sup}

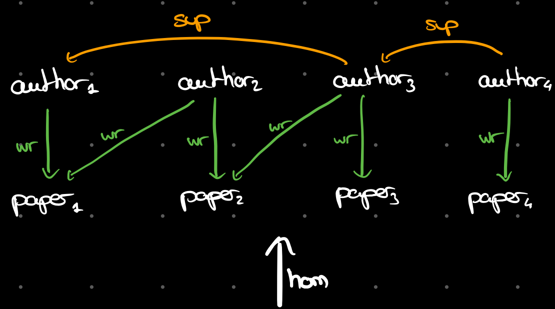
Atoms: $x \xrightarrow{L} y$

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

⤳ "homomorphism" semantic

Evaluation:
(author$_1$, paper$_4$),
(author$_4$, paper$_4$),
...

Path queries (cont.)



$$x \xrightarrow{(wr \cdot wr^-)^*} y$$

Prop Evaluation of C2RPQs is
- NP-complete ...
- PTIME when restricted to queries of tw $\leq k$.

Conjunctive 2-way regular path queries (C2RPQs)

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

regular lang on $\{wr, sp, wr^-, sp^-\}$

Atoms: $x \xrightarrow{L} y$

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

$$\gamma(x,y) = \exists y.\ \begin{tikzcd} & y \arrow[dr, "b^*"] & \\ x \arrow[ur, "a^*"] \arrow[rr, "c^*"'] & & z \end{tikzcd} \quad \equiv \quad \gamma'(x,y) = \begin{tikzcd} x \arrow[r, bend left, "a^* b^*"] \arrow[r, bend right, "c^*"'] & z \end{tikzcd}$$

Ex

$$\delta(x) = \exists y, z.\ \begin{tikzcd} x \arrow[rr, "b"] & & y \\ & z \arrow[ul, "a"] \arrow[ur, "a"'] & \end{tikzcd} \quad \equiv \quad \delta'(x) = \begin{tikzcd} x \arrow[loop right, "ba^{-}a"] \end{tikzcd}$$

(minimal CQ)

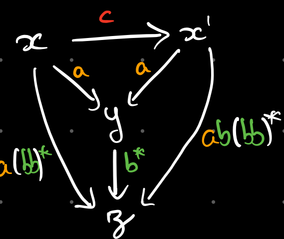
Pb C(2)RPQ <u>cannot</u> be minimised...

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 $\leq k$

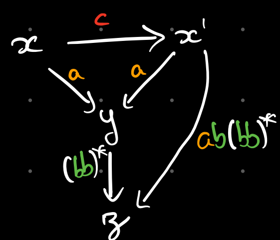
For CRPQs this is (probably) false...

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

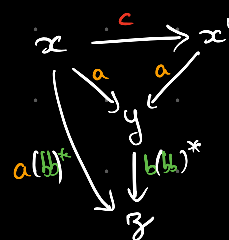


has tree-width 3 ...

$\equiv$



$\vee$



union of CRPQs tree-width ≤ 2

Deciding semantic tree-width

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

Ex $\gamma(x) = \exists y,z.$ $\begin{tikzcd} x \arrow[rr, "b"] & & y \\ & z \arrow[ul, "a"] \arrow[ur, "a"'] & \end{tikzcd}$ $\equiv$ $\begin{tikzcd} x \arrow[loop right, "b\,a^{-}a"] \end{tikzcd}$

sem. tw ≤ 1

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!

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!

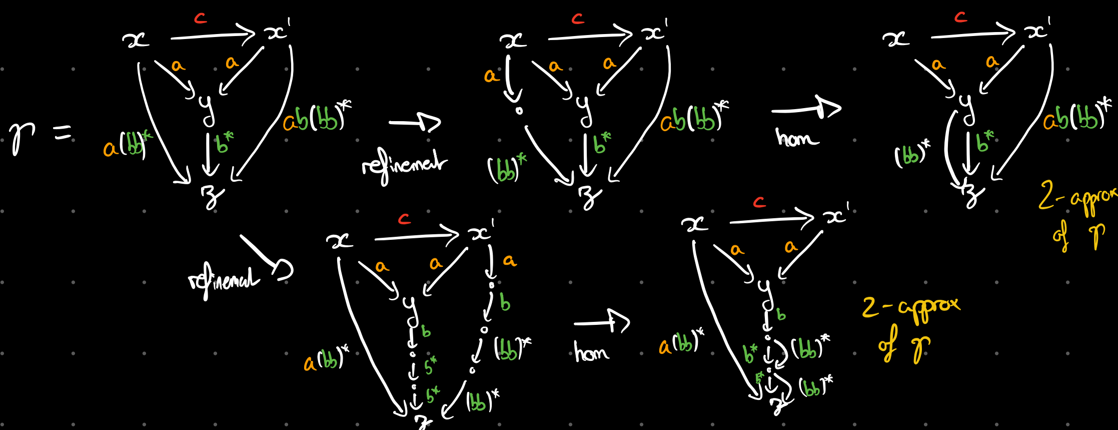
- DECIDABLE & EFFECTIVE for UC2RPQs when $k=1$ [Barceló, Romero & Vardi, PODS '13]
  ↖ ExpSpace-complete
- DECIDABLE & EFFECTIVE for UC2RPQs when $k \geq 2$ [Figueira, M., ICDT '23] & [Feier, Gogacz, Murlak, unpublished]
  ↖ 2ExpSpace
  likely ExpSpace-complete

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

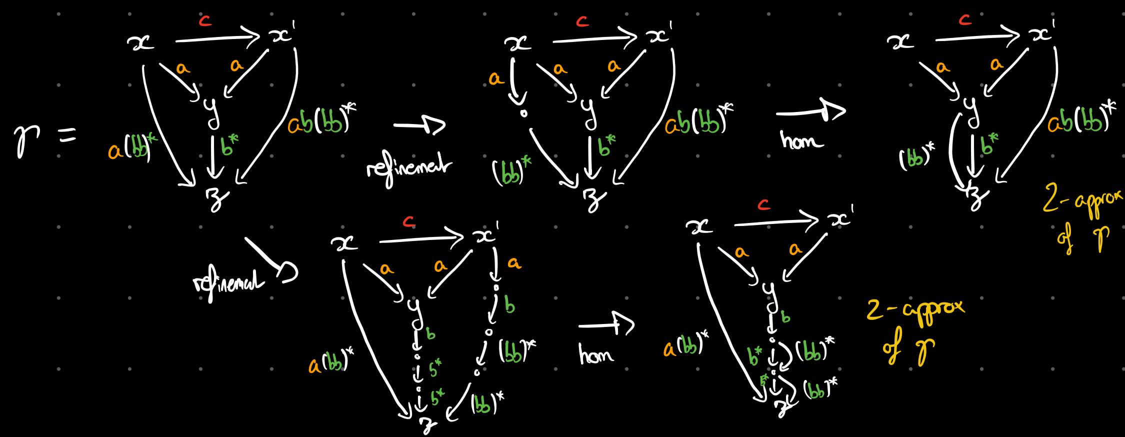
Deciding semantic tree-width ($k \geq 2$)

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-approximation

= 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.

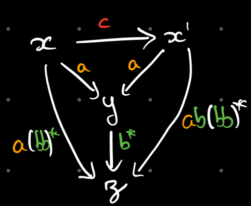
Properties of semantic tree-width

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

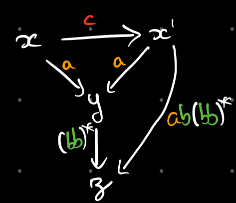
1) T is equivalent to an infinite union of CRPQs of $tw \leq k$
2) T is equivalent to a UC2RPQ of $tw \leq k$
3) T is equivalent to an infinite union of CQs of $tw \leq k$.

⊕ Closure property on the regular languages.

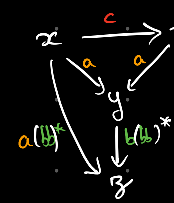
Ex $k=2$



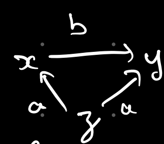
$\equiv$



$\vee$



$k=1$ $\exists y,z.$



$\equiv$

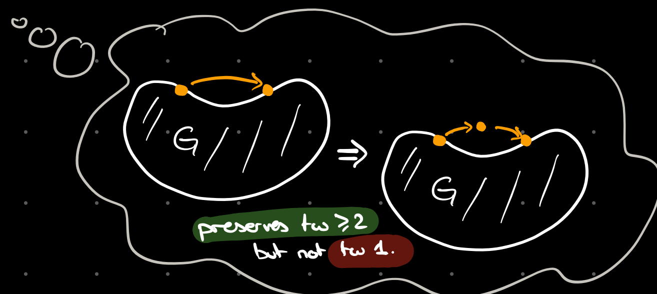
not expressible as an infinite set of CQs of $tw \leq 1$.

Semantic tree-width: overview

The following are decidable:

CQ	equivalent to a	CQ	of $tw \leq k$?	} minimisation or approxima°
UCQ	—"—	UCQ	—"—	
UC2RPQ	—"—	UC2RPQ	—"—	} approxima°

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



Simple regular expressions

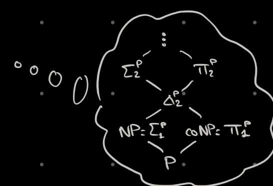
2EXPSPACE algo for deciding sem tw $\leqslant k$
EXPSPACE claimed by Feier, Gogacz & Murlak

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

UC2RPQ(SRE): $x \xleftarrow{sup^*} \gamma \xrightarrow{wr} y$, etc.

75% of all path queries "from real life"
[Bonifati, Martens, Timm, 2020]

Theorem [Figueira, M., ICDT '23]
Semantic tree-width $\leqslant k$ is in Π_2^P over UC2RPQ(SRE).



A glimpse beyond...

Query of sem tw $\leq k$ → Compute equivalent query of tw $\leq k$ → Evaluate it

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

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

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

Thm [Grohe, FOCS 2003] $\mathcal{C}$: class of CQs

Evaluation of $\mathcal{C}$ is FPT

IFF

Evaluation of $\mathcal{C}$ is PTIME

IFF

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

Open question:

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

Evaluation of $\mathcal{C}$ is FPT

IFF ?

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