

Approximation and semantic tree-width of conjunctive regular path queries

Rémi Morvan

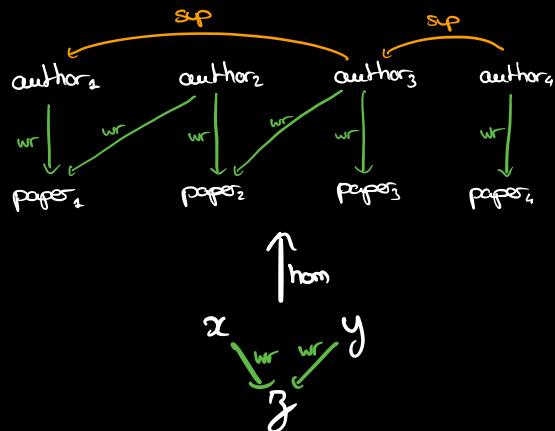
joint work with
Diego Figueira

LABRI, U. Bordeaux

6 December 2022

T2F seminar,
Bordeaux

(Graph) databases



Conjunctive queries (CQs)

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

Evaluation:

(author₁, author₂),
 (author₁, author₃),
 etc...

→ homomorphism semantic



Evaluation of CQs is
NP-complete ...

(combined complexity:
 input: database & query)

Upper bound:

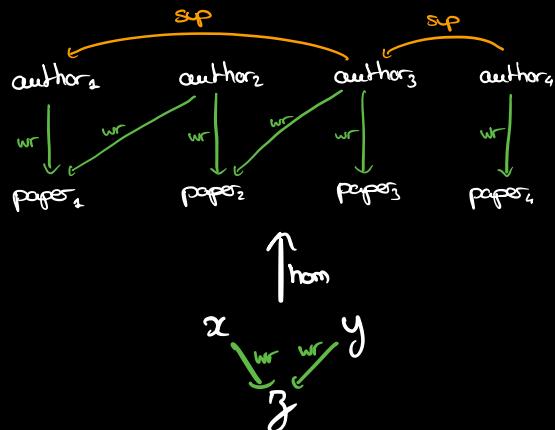
Input: $p(\bar{x})$ CQ

G database

\bar{a} tuple in G

Algo: Guess f : vars(p) $\rightarrow G$
 & check that $\left\{ \begin{array}{l} f(p(\bar{x})) = \bar{a} \\ f \text{ homomorphism.} \end{array} \right.$

(Graph) databases



Prop

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

Conjunctive queries (CQs)

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

Evaluation:

(author₁, author₂),
(author₁, author₃),
etc...

→ homomorphism semantic

Lower bound:

$$\begin{array}{c} 3\text{-COL} \leq_p \text{CQ-EVAL} \\ G \text{ 3-colorable?} \mapsto G \xrightarrow{\text{hom}} \begin{array}{c} \text{?} \\ \text{?} \\ \text{?} \end{array} \end{array}$$

↑ Boolean query ↑ graph database

One solution: tree-width

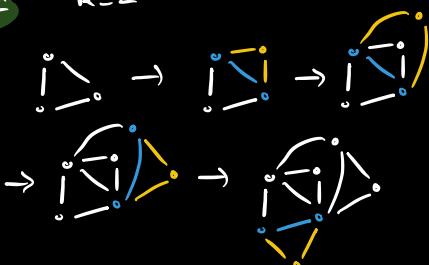
Def k -trees:

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

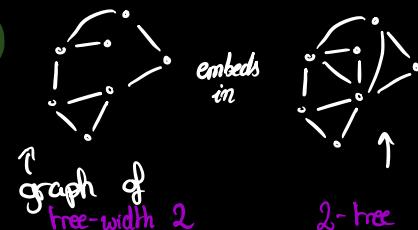
Def G has tree-width $\leq k$
if we can embed
it in a k -tree.
remove
nodes & edges

Ex

$k=2$



Ex



Tree-width and CQs

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

Proof ($k=1$)



Complexity:

$$|G|^2 \cdot (\|\text{EVAL}(\rho_1, G)\| + \|\text{EVAL}(\rho_2, G)\| + \text{cst}) \\ \rightsquigarrow \mathcal{O}(|G|^2 \cdot \|\rho\|)$$

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

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

for $u \in G$:

$\exists v \in G$ st $u \xrightarrow{a} v$
and $\text{EVAL}(\rho_1, G)$

AND

$\exists w \in G$ st $u \xrightarrow{b} w$
and $\text{EVAL}(\rho_2, G)$

Semantic tree-width

same evaluation
on every database

Q^o: Given $p(x) \in Q$, is $p(x)$ equivalent to
a $\in Q$ $p'(x)$ of $\text{tw} \leq k$?

$$\exists x \quad p = \exists x \exists y \exists z. \begin{array}{c} x \\ \swarrow a \quad \searrow a \\ b \xrightarrow{b} y \xrightarrow{b} z \end{array} \equiv p' = \exists x \exists y. \begin{array}{c} x \\ a \downarrow \\ b \xrightarrow{a} y \end{array}$$

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

but semantic tree-width 1 !

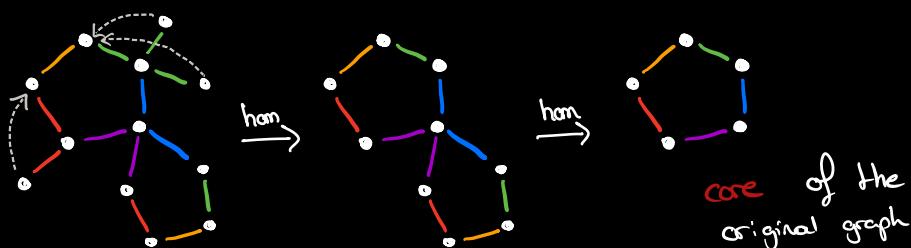
Minimisation of CQs

for the number of variables.

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

Def

Core of an edge-labelled graph : minimal retraction



Minimisation of CQs =
Core of graphs.

Ex

$$\text{core} \left(\xrightarrow{\text{b}} \xrightarrow{\text{Gy}} \xrightarrow{\text{b}} \gamma \right) = \xrightarrow{\text{a}} \xrightarrow{\text{Gy}} \gamma$$

CQs (not) of small semantic tree-width

Prop

γ has
semantic $tw \leq k$

IFF

$\text{core}(\gamma)$ 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 $tw \leq 1$ of $\gamma(x)$ by under-approximation of $tw \leq 1$ is the maximal union of CQs of $tw \leq 1$.

Ex

$$\begin{array}{c} x \\ \alpha \downarrow \\ y \end{array} \vee \begin{array}{c} x \supseteq a \\ \alpha \left(\right) a \\ y \end{array} \vee \begin{array}{c} x \supseteq a \\ \downarrow a \\ y \end{array} \subseteq \begin{array}{c} x \\ \alpha \downarrow \\ y \end{array}$$

Conjunctive queries: Summary

Evaluating CQs \approx Homomorphism problem

$$\exists x,y,z, \begin{array}{c} x \xrightarrow{a} y \\ \wedge y \xrightarrow{b} z \end{array} \text{ in } G? \Leftrightarrow \begin{array}{c} x \\ \downarrow a \\ y \\ \downarrow b \\ z \end{array} \xrightarrow{\text{hom}} G$$

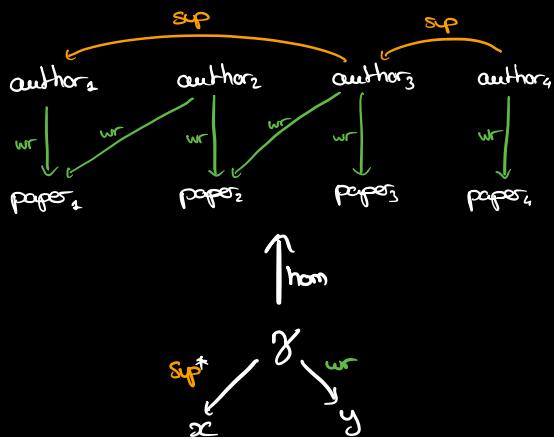
NP-complete

PTime when restricted
to queries of $\text{tw} \leq k$

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 $\text{tw} \leq k$

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 L_{wr}, sp

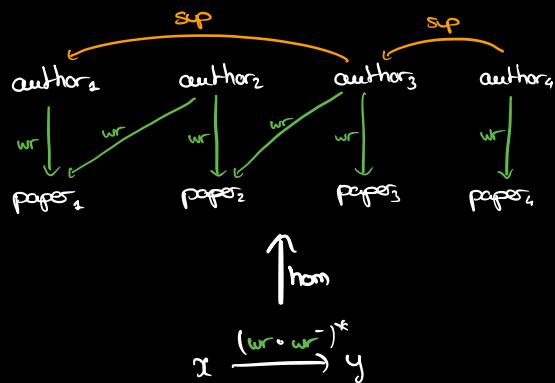
$$r(x,y) = \exists g. \, g \xrightarrow{sp^*} x \wedge g \xrightarrow{wr} y$$

~ "homomorphism" semantic

Evaluation:

(author₂, paper₄),
(author₄, paper₄),
...

Path queries (cont.)



Prop

Evaluation of C2RPQs is
NP-complete ...
PTIME when restricted to
queries of $\text{bw} \leq k$.

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 $\{wr, sp, wr^-, sp^-\}$

Ex: $p(x,y) = x \xrightarrow{(wr \cdot wr^-)^*} y$

⇒ "homomorphism" semantic

Evaluation:

(author₂, author₃),

...

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 $p(x,y) = \exists y. \ x \xrightarrow{a^*} y \quad \equiv \quad p'(x,y) = x \xrightarrow{a^* b^*} y$

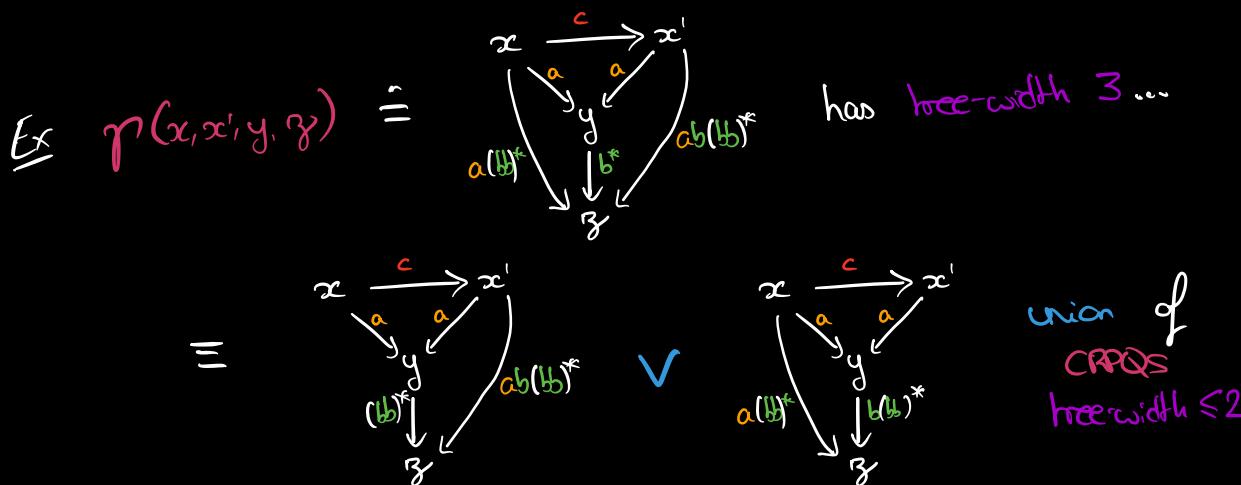
Ex $\delta(x) = \exists y. \ x \xrightarrow{b} y \quad \equiv \quad \delta'(x) = x \curvearrowleft^{ba^-a}$
(minimal CQ)

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

Union

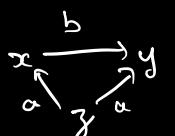
Fact For CQs, r is equivalent to a CQ of $\text{tw} \leq k$ iff r is equivalent to a union of CQs of $\text{tw} \leq k$

For CRPQs this is (probably) false...



Deciding semantic tree-width

Def: A UC2APQ Γ has semantic tree-width $\leq k$
if it is equivalent to a UC2RPQ of $\text{tw} \leq k$.

Ex $\gamma^{(x)} = \exists y z.$  $\equiv x \rightsquigarrow b \bar{a} a$

\uparrow
sem. tw ≤ 1

DECIDING SEMANTIC TREE-WIDTH:

Input: Γ
Q^o: Γ has sem tw $\leq k$? \leftarrow fixed

Motiv^o:
UC2RPQs of $\text{tw} \leq k$
can be evaluated in PRIME!

Deciding semantic tree-width (cont.)

DECIDING SEMANTIC TREE-WIDTH:

Input: Γ

Q^o: Γ has sem tw $\leq k$?

fixed

Motiv^o:

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

- DECIDABLE & EFFECTIVE for UC2RPQs when $k \leq 1$ [Barceló, Romero & Vardi, PODS '13]
ExpSPACE-complete
- DECIDABLE & EFFECTIVE for UC2RPQs when $k \geq 2$ [Figueira, M., ICDT '23]
2ExpSPACE
likely Exp-Space-complete

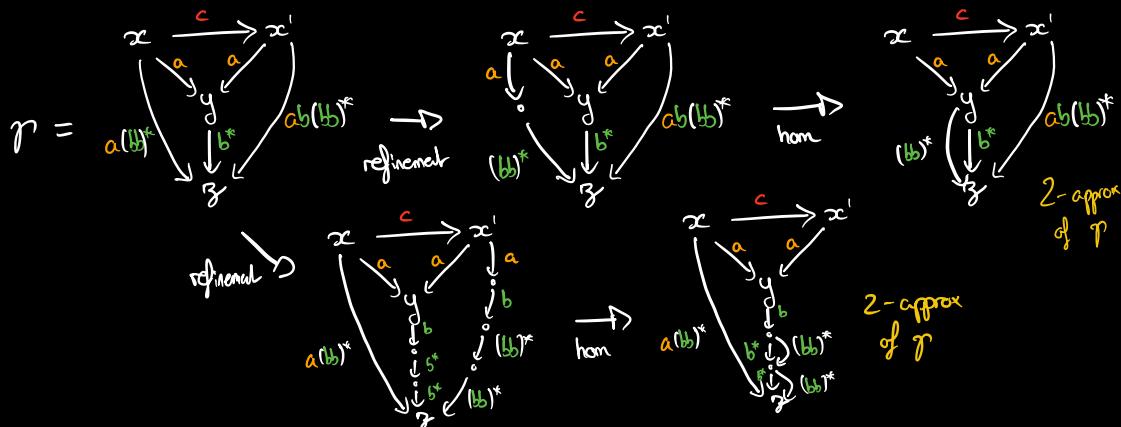
& [Feier, Gagacy, Murak, unpublished]

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

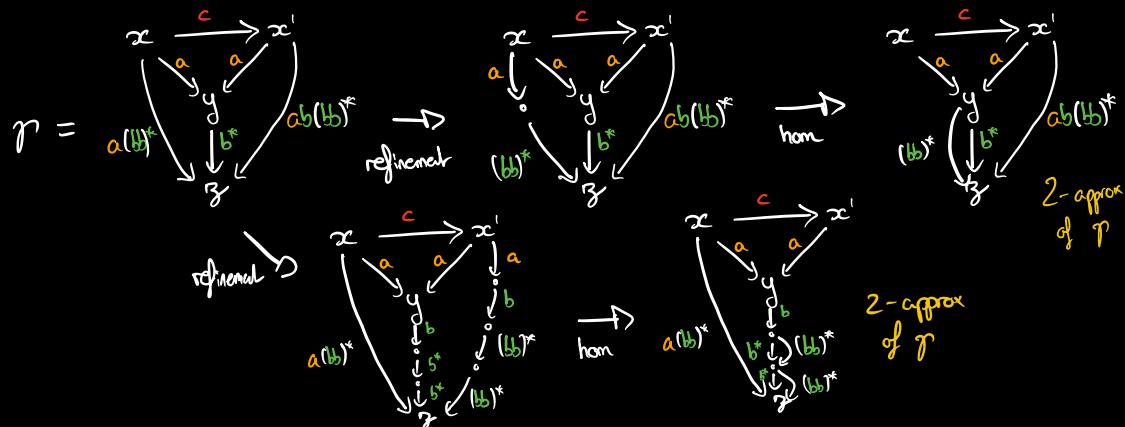
Deciding semantic tree-width ($k \geq 2$)

Idea: Start with a UC2RPQ.

Select any C2RPQ in the union and "refine" it,
 then fold it → if it has $\text{tw} \leq k$, it is a
 $= \begin{matrix} \text{surjective} \\ \text{homomorphism} \end{matrix}$ k -approxima^o



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.

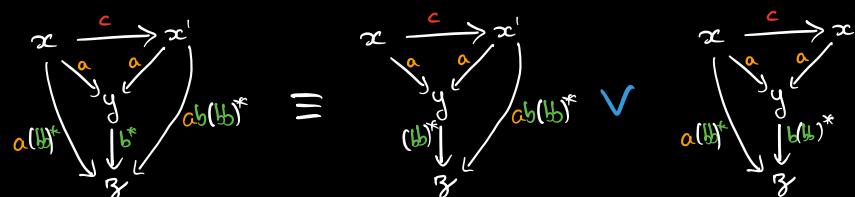
→ Test if this UC2RPQ is equivalent to the original one.

Properties of semantic tree-width

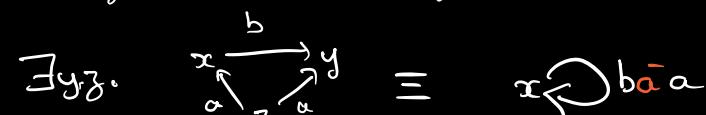
Theorem [Figueira, M., ICDT '23] T : UC2RQ, $k \geq 2$. TFAE:

- 1) T is equivalent to an infinite union of CRPQS of $\text{tw} \leq k$
 - 2) T is equivalent to a UC2RPQ of $\text{tw} \leq k$
 - 3) T is equivalent to an infinite union of CQs of $\text{tw} \leq k$.
- ⊕ Closure property on the regular languages.

Ex $\underline{k=2}$



$\underline{k=1}$



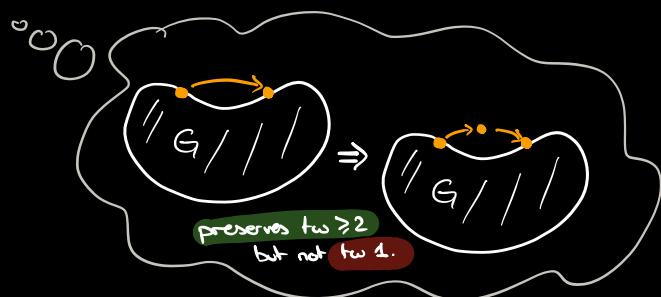
not expressible as an infinite set of CQs of $\text{tw} \leq 1$.

Semantic tree-width : overview

The following are decidable:

$$\begin{array}{lll} \text{CQ} & \text{equivalent to a CQ of } \text{tw} \leq k ? & \left. \begin{array}{l} \text{minimisation} \\ \text{or} \\ \text{approximate} \end{array} \right\} \\ \text{UCQ} & \sim \sim \sim & \text{UCQ} \sim \sim \sim \\ \text{UC2RPQ} & \sim \sim \sim & \text{UC2RPQ} \sim \sim \sim \left. \begin{array}{l} \text{approximate} \end{array} \right\} \end{array}$$

For UC2RPQ, the case $k=1$ and $k>2$ 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^* .

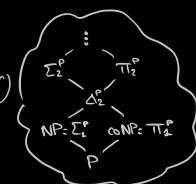
UC2RPQ(SRE) : Σ^* $\xrightarrow{\delta}$ y , etc.

75% of all path queries "from real life"

[Bonfatti, Martens, Timm, 2020]

Theorem [Figueira, M., ICDT '23]

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



A glimpse beyond ...

Query of sem $\text{tw} \leq k \rightarrow$ Compute equivalent query of $\text{tw} \leq k$ \rightarrow Evaluate it

in $|T|$

FPT^C algo for evaluation
of queries of sem $\text{tw} \leq k$. $\mathcal{O}(f(|T|) \cdot |G|^{k+1})$

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

Thm [Grohe, Focs 2003] \mathcal{C} : re. 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 $\text{FPT} \neq \text{W[1]}$

Open question:

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

Evaluation of \mathcal{C} is FPT
IFF ?

\mathcal{C} has bounded sem tree-width