Approximation and semantic tree-width of conjunctive regular path queries

Rémi Morvan

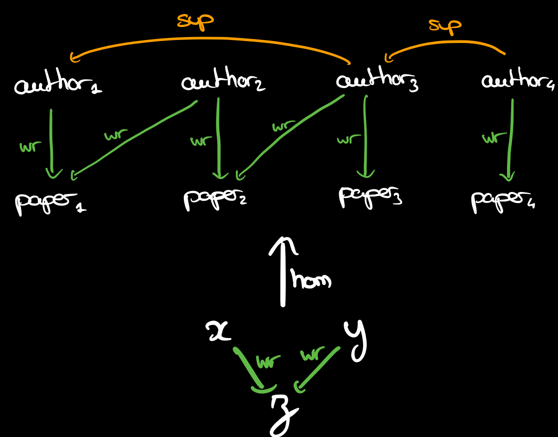
joint work with Diego Figueira

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

$\rightsquigarrow$ homomorphism semantic

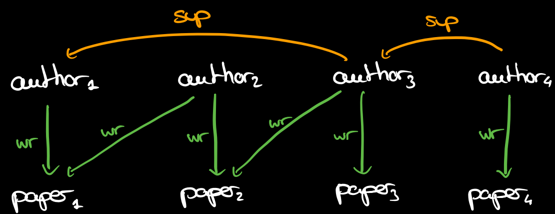
Prop Evaluation of CQs is NP-complete...
(combined complexity: input: database & query)

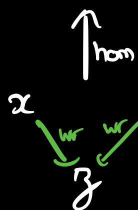
Upper bound:

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)

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

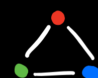
Evaluation:

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

$\rightsquigarrow$ homomorphism semantic

Lower bound:

3-COL $\leq_p$ CQ-EVAL

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

Boolean query

graph database

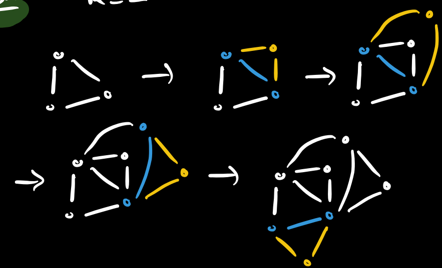
One solution: tree-width

Def k-trees:

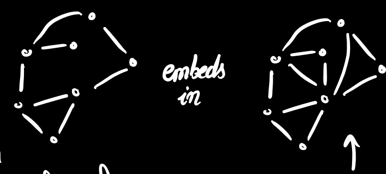
- start with a $(k+1)$-clique
- repeat:
  add a new node, and join it
  to a k-clique.

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

Ex $k=2$



Ex



↑ graph of tree-width 2

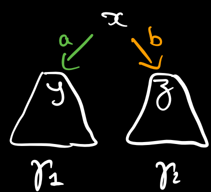
↑ 2-tree

Tree-width and CQs

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

Proof ($k=1$)

$\gamma =$



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

for $u \in G$:

∃? $v \in G$ st $u \xrightarrow{a} v$
and $\mathrm{EVAL}(\gamma_1, G)$
AND
∃? $w \in G$ st $u \xrightarrow{w} v$
and $\mathrm{EVAL}(\gamma_2, G)$

Complexity:

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

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

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

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 $\quad \gamma = \exists x\, \exists y. \exists z.$
$$\begin{tikzcd} & x \arrow[dl, "a"'] \arrow[dr, "a"] & \\ y \arrow[loop left, "b"] \arrow[rr, "b"'] & & z \end{tikzcd}$$
$\equiv \gamma' = \exists x. \exists y.$
$$\begin{tikzcd} x \arrow[d, "a"'] \\ y \arrow[loop left, "b"] \end{tikzcd}$$

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

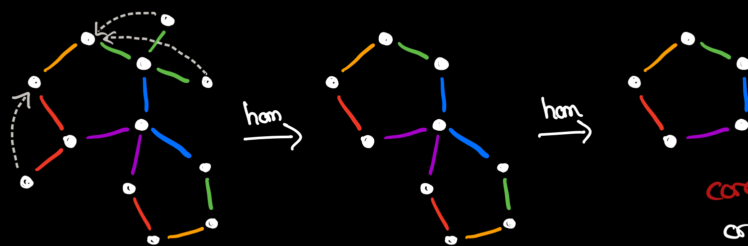
but semantic tree-width 1 !

Minimisation of CQs

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

for the number of variables.

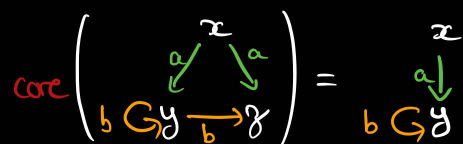
Def Core of an edge-labelled graph: minimal retraction



core of the original graph

Minimisation of CQs = Core of graphs.

Ex



CQs (not) of small semantic tree-width

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

$$\begin{tikzcd} & x \arrow[dl, "a"'] \arrow[dr, "a"] & \\ y \arrow[rr, "a"'] & & z \end{tikzcd}$$

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 $tw \leqslant 1$ of $\gamma(x)$ is the maximal under-approximation of $\gamma(x)$ by unions of CQs of $tw \leqslant 1$.

Ex

$$\begin{tikzcd} x \arrow[d, "a"'] \\ y \arrow[loop right, "a"] \end{tikzcd} \vee \begin{tikzcd} x \arrow[loop right, "a"] \arrow[d, bend right, "a"'] \\ y \arrow[u, bend right, "a"'] \end{tikzcd} \vee \begin{tikzcd} x \arrow[loop right, "a"] \arrow[d, "a"] \\ z \end{tikzcd} \subsetneq \begin{tikzcd} & x \arrow[dl, "a"'] \arrow[dr, "a"] & \\ y \arrow[rr, "a"'] & & z \end{tikzcd}$$

Conjunctive queries: summary

Evaluating CQs $\cong$ Homomorphism problem

$$\exists x,y,z,\ \begin{array}{l} x \xrightarrow{a} z \\ \wedge\ y \xrightarrow{b} z \end{array} \text{ in } G\,? \iff \left(x \xrightarrow{a} z \xleftarrow{b} y \right) \xrightarrow{\text{hom}\,?} G$$

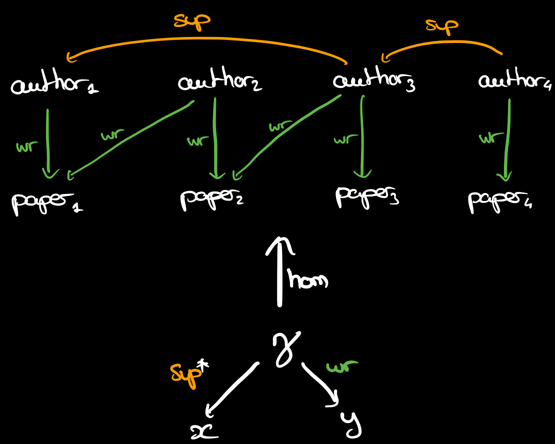
NP-complete

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}

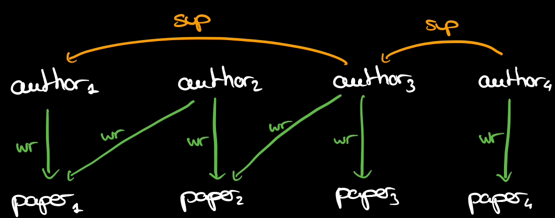
Atoms: $x \xrightarrow{L} y$

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

⇝ "homomorphism" semantic

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

Path queries (cont.)



$\uparrow$ hom

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

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 $y \xrightarrow{a} x$.

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

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

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

Ex $\delta(x) = \exists y,z.$
$$\begin{tikzcd} x \arrow[rr, "b"] & & y \\ & z \arrow[ul, "a"] \arrow[ur, "a"'] & \end{tikzcd}$$
(minimal CQ)

$\equiv \delta'(x) =$
$$\begin{tikzcd} x \arrow[loop right, "ba^{-}a"] \end{tikzcd}$$

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 $\gamma(x, x', y, z) \hat{=}$

$$\begin{tikzcd} x \arrow[rr, "c"] \arrow[dr, "a"] \arrow[ddr, bend right, "a(bb)^*"'] & & x' \arrow[dl, "a"'] \arrow[ddl, bend left, "ab(bb)^*"] \\ & y \arrow[d, "b^*"] & \\ & z & \end{tikzcd}$$

has tree-width 3 ...

$$\equiv \quad \begin{tikzcd} x \arrow[rr, "c"] \arrow[dr, "a"] & & x' \arrow[dl, "a"'] \arrow[ddl, bend left, "ab(bb)^*"] \\ & y \arrow[d, "(bb)^*"'] & \\ & z & \end{tikzcd} \quad \vee \quad \begin{tikzcd} x \arrow[rr, "c"] \arrow[dr, "a"] \arrow[ddr, bend right, "a(bb)^*"'] & & x' \arrow[dl, "a"'] \\ & y \arrow[d, "b(bb)^*"] & \\ & z & \end{tikzcd}$$

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

$$\begin{tikzcd} x \arrow[rr, "b"] & & y \\ & z \arrow[ul, "a"] \arrow[ur, "a"'] & \end{tikzcd} \equiv \begin{tikzcd} x \arrow[loop right, "b\bar{a}a"] \end{tikzcd}$$

↑ 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!

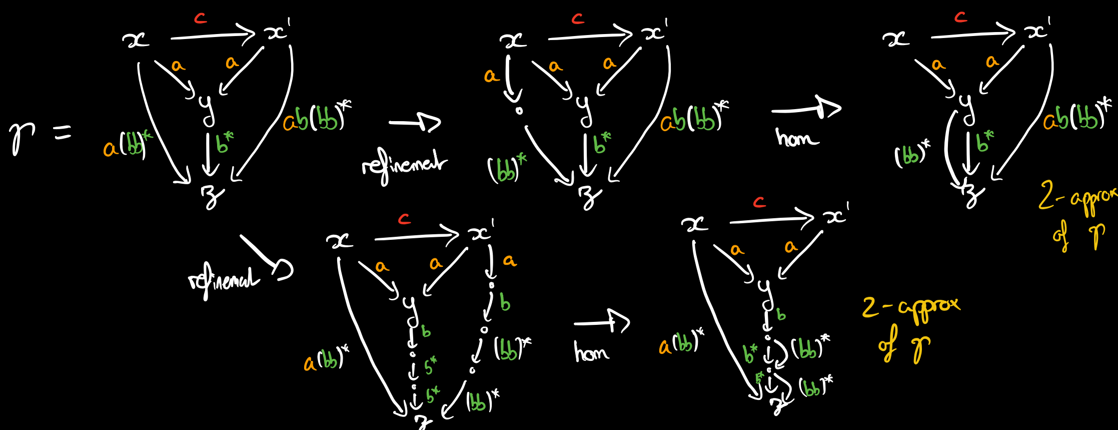
- DECIDABLE & EFFECTIVE for UC2RPQs when $k=1$ [Barceló, Romero & Vardi, PODS '13]
  ExpSpace-complete

- DECIDABLE & EFFECTIVE for UC2RPQs when $k \geq 2$ [Figueira, M., ICDT '23]
  2ExpSpace
  likely ExpSpace-complete
  & [Feier, Gogacz, Murlak, unpublished]

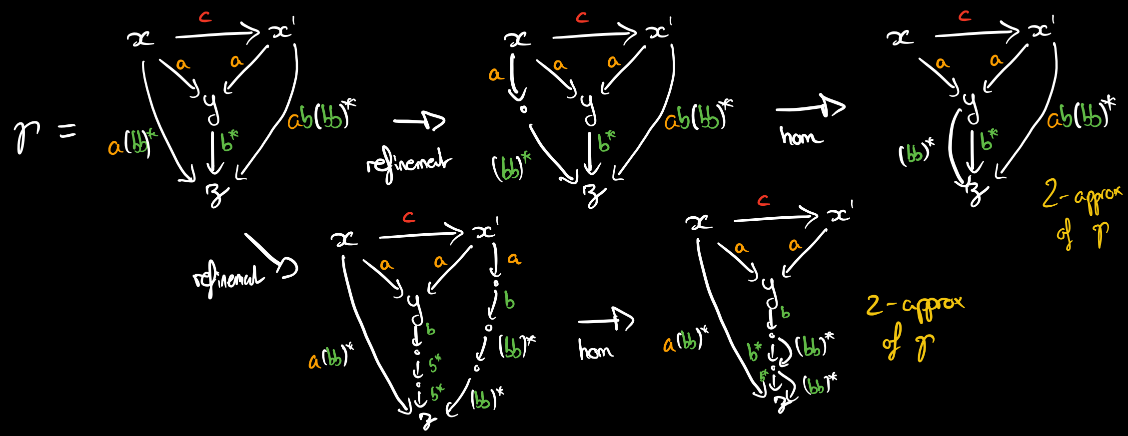
The case $k=1$ and $k \geq 2$ seem very different...

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 $tw \leq k$, 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.

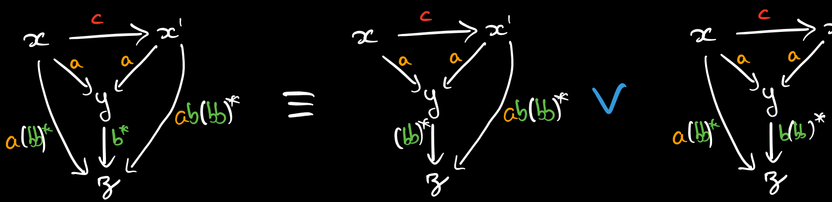
Properties of semantic tree-width

Theorem [Figueira, M., ICDT '23] Γ : UC2RPQ, $k \geq 2$. TFAE :

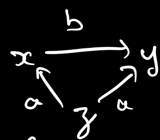
1) Γ is equivalent to an infinite union of CRPQs of $tw \leq k$
2) Γ is equivalent to a UC2RPQ of $tw \leq k$
3) Γ is equivalent to an infinite union of CQs of $tw \leq k$.

⊕ Closure property on the regular languages.

Ex $k=2$



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



$\equiv$ x ⟲ $b a^{-} a$

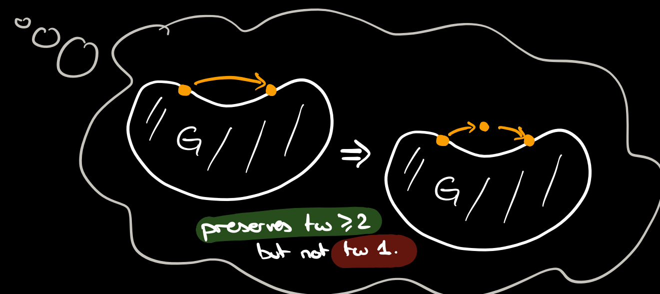
not expressible as an infinite set of CQs of $tw \leq 1$.

Semantic tree-width: overview

The following are decidable:

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

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



Simple regular expressions

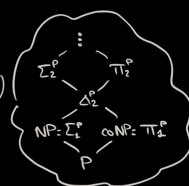
2EXPSPACE algo for deciding sem tw $\leqslant k$
EXPSPACE claimed by Feier, Gogacz & Murlak

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

UC2RPQ(SRE): $\begin{tikzcd} & \gamma \arrow[dl, "sup^*"'] \arrow[dr, "wr"] & \\ x & & y \end{tikzcd}$, 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 $\leq k$ $\to$ Compute equivalent query of tw $\leq k$ $\to$ Evaluate it

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

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

[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 $\neq$ 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