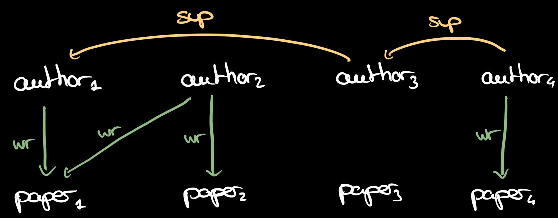
Approximation and semantic tree-width of conjunctive regular path queries

Diego Figueira & Rémi Morvan
LaBRI, Univ. Bordeaux

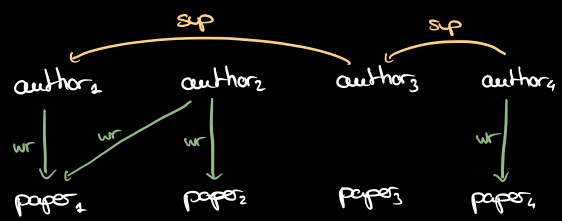
29 March 2023
ICDT '23, online/Ioannina

Path queries



graph database

Path queries



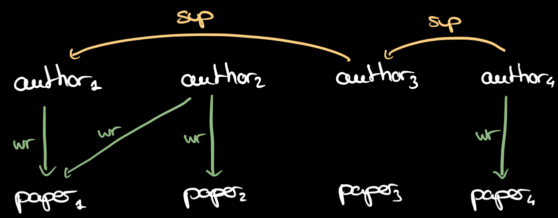
graph database

$$\gamma(x) = x \xrightarrow{sup^+} y \wedge y \xrightarrow{wr} z$$

"people with a productive descendance"

Conjunctive regular path queries (CRPQs)

Path queries



graph database

$\gamma(x) = x \xrightarrow{sup^+} y \wedge y \xrightarrow{wr} z$

"people with a productive descendance"

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

"same connected component of co-authorship"

Conjunctive regular path queries (CRPQs) $\subseteq$ CRPQs with 2-way navigation (C2RPQs)

Evaluation & containment of C(2)RPQs

Fact Evaluation of C(2)RPQs is NP-complete in combined complexity.

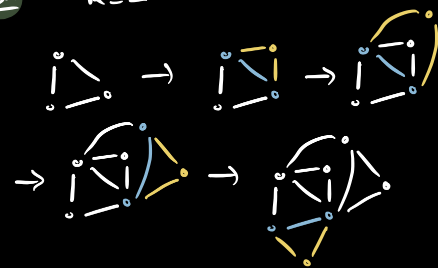
Prop [Florescu, Levy & Suciu PODS '98
indep. Calvanese, De Giacomo, Lenzerini & Vardi KR '00]
Containment of C(2)RPQs is ExpSPACE-complete

One solution: tree-width

Def k-trees:

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

Ex $k=2$



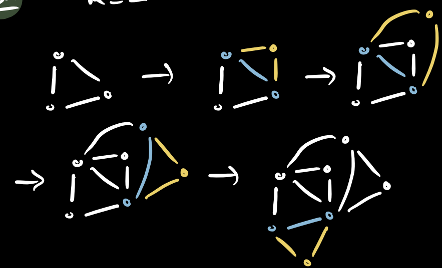
One solution: tree-width

Def k-trees:

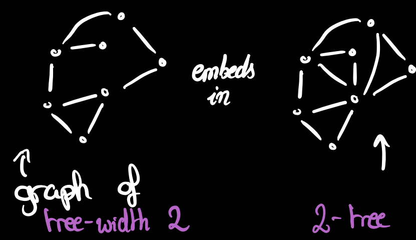
- start with a $(k+1)$-clique
- repeat:
  pick a k-clique, and join it to a new node.

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

Ex $k=2$



Ex



Semantic tree-width

Prop C(2)RPQ of tree-width $\leq k$ can be evaluated in polynomial time.

Semantic tree-width

Prop C(2)RPQ of tree-width $\leq k$ can be evaluated in polynomial time.

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

Ex

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

Ex

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

(minimal CQ)

Semantic tree-width

Prop C(2)RPQ of tree-width $\leq k$ can be evaluated in polynomial time.

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

Ex

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

Ex

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

(minimal CQ)

Pb C(2)RPQs 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...

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

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

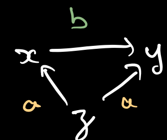
has tree-width 3...

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

union of CRPQs tree-width ≤ 2

Deciding semantic tree-width

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

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

$\equiv$ $x \circlearrowleft b a^{-} a$

$\uparrow$ sem. tw ≤ 1

DECIDING SEMANTIC TREE-WIDTH :

Input: Γ $\leftarrow$ fixed

Q°: Γ has sem tw $\leq k$?

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

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 & ExpSpace-hard

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

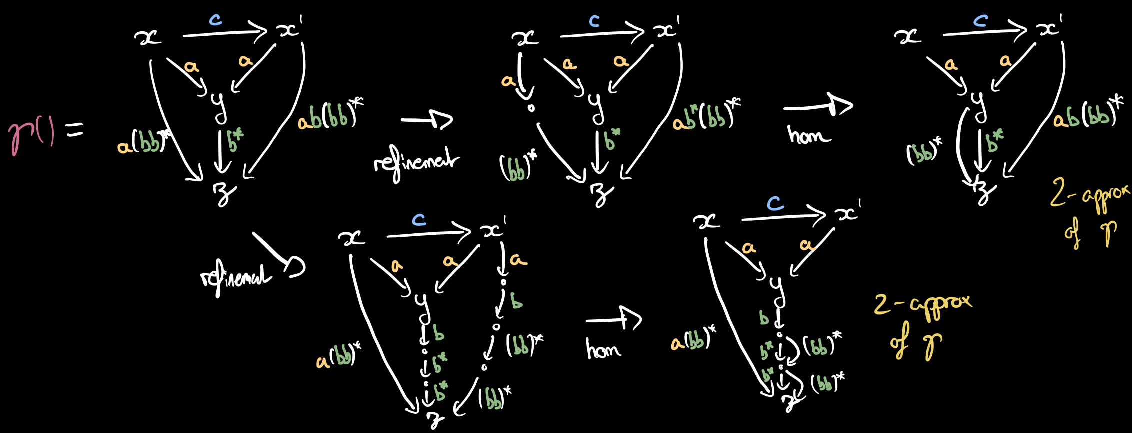
Deciding semantic tree-width ($k \geq 2$)

<u>Idea:</u> Start with a C2RPQ, and "refine" it, then $\underbrace{\text{fold it}}_{=\text{ surjective homomorphism}}$ → if it has $tw \leq k$, it is a k-approxima°

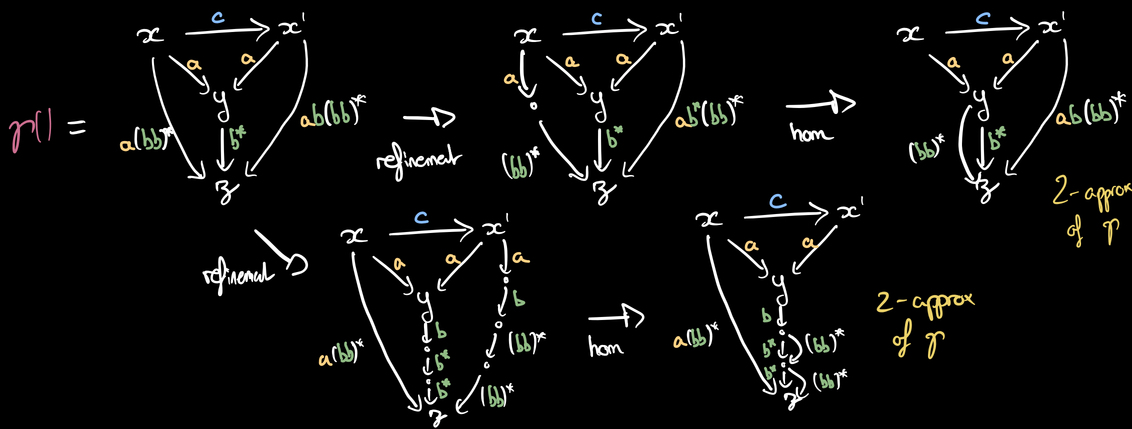
Deciding semantic tree-width $(k \geq 2)$

Idea: Start with a C2RPQ, and "refine" it, then fold it → if it has tw $\leq k$, it is a k-approximation

= surjective homomorphism

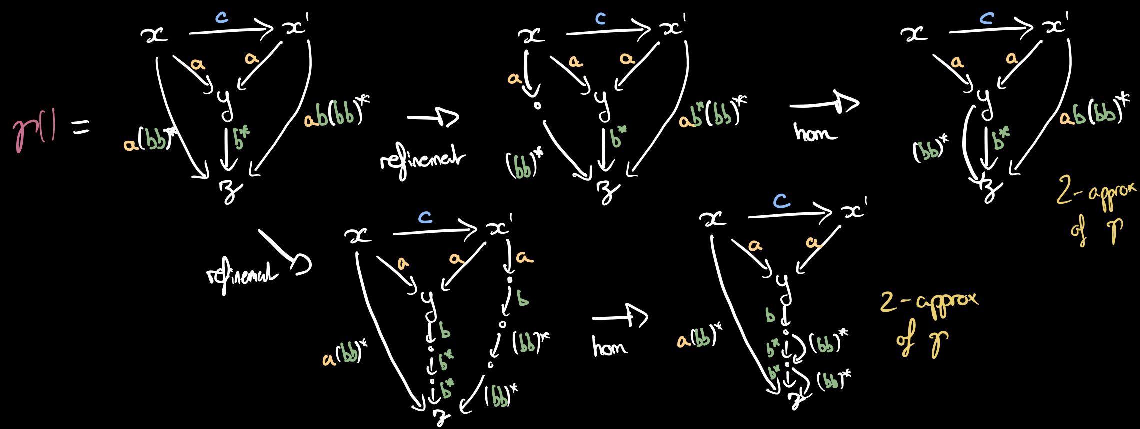


Deciding semantic tree-width (k≥2)



We obtain an infinite set of k-approximations.

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: UC2RPQ, $k \geq 2$. TFAE:

1) T is equivalent to an infinite union of C2RPQs 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 $tw \leq k$.

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 C2RPQs 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 $tw \leq k$.

Co-ex $(k=1)$

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

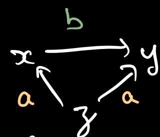
not expressible as an infinite set of CQs of $tw \leq 1$.

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 C2RPQs 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.

Co-ex ($k=1$) $\exists yz.$ 

$\equiv$ 

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

Simple regular expressions

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

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

Simple regular expressions

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

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

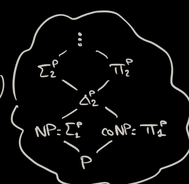
UC2RPQ(SRE) : 75% of all path queries "from real life"

[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 ...

Evaluation of queries of sem $tw \leq k$ $\rightsquigarrow$ $O(f(|\Gamma|) \cdot |G|^{k+1}) \in$ FPT in $|\Gamma|$

database

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

Open question: Let $\mathcal{C}$ be a r.e. class CRPQs / UC2RPQs.
Evaluation of $\mathcal{C}$ is FPT
IFF ?
$\mathcal{C}$ has bounded sem tree-width