Approximation and semantic tree-width of conjunctive regular path queries

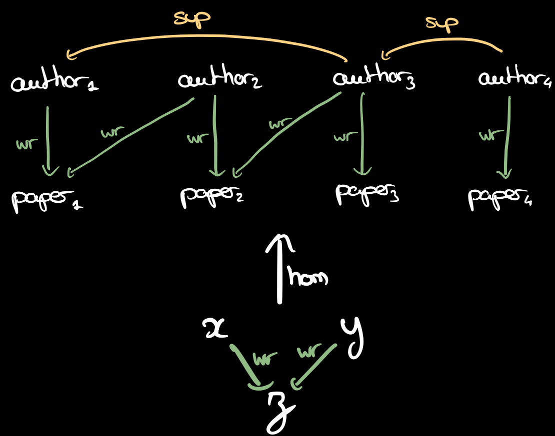
Rémi Morvan
www.morvan.xyz

joint work with
Diego Figueira

LABRI, U. Bordeaux

21 April 2023
GT DAAL, Le Kremlin-Bicêtre

(Graph) databases



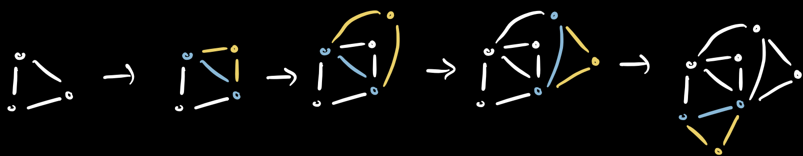
Conjunctive queries (CQs)

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

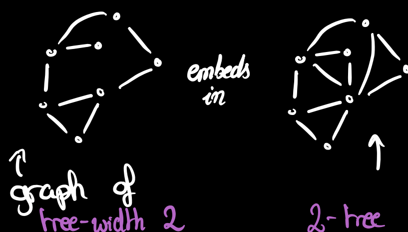
Prop Evaluation of CQs is NP-complete ...
(combined complexity:
input: database & query)

One solution: tree-width

Def / **Ex** k-trees $(k=2)$



Def / **Ex** tree-width $\leq k$



Tree-width and CQs

Thm [Yannakakis '81, Dechter & Pearl '89, indep. Freuder '90] Fix k. Evaluation of CQs of tree-width $\leq k$ is $O(|G|^{k+1} \cdot ||\gamma||) \in$ PTIME

database

query

Tree-width and CQs

Thm [Yannakakis '81
Dechtler & Pearl '89
indep. Freuder '90]

Fix k. Evaluation of CQs of tree-width $\leq k$ is $O(|G|^{k+1} \cdot \|\gamma\|) \in$ PTIME

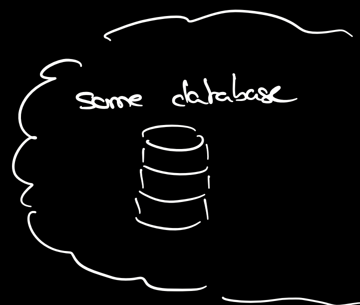
database

query

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

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

eval



Tree-width and CQs

Thm [Yannakakis '81, Dechter & Pearl '89, indep. Freuder '90]

Fix k. Evaluation of CQs of tree-width $\leq k$ is $O(|G|^{k+1} \cdot \|\varphi\|) \in$ PTIME

database, query

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

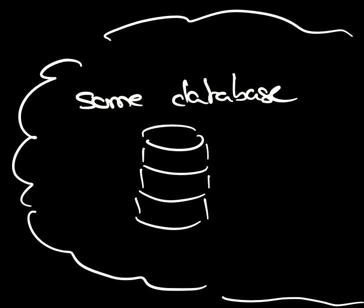
"semantic tree-width ≤ 1"

$$\equiv$$

$$\varphi' = \exists x. \exists y. \quad \begin{tikzcd} x \arrow[d, "a"'] \\ y \arrow[loop left, "b"] \end{tikzcd}$$

eval

eval

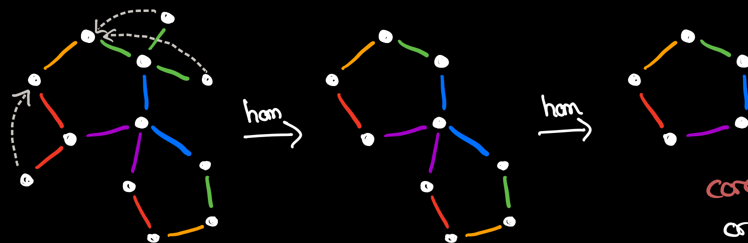


Minimisation of CQs

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

↙ for the number of variables.

Def



core of the original graph

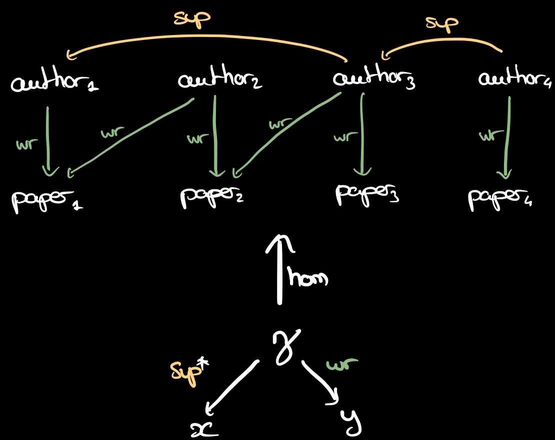
Minimisation of CQs = Core of graphs.

Ex

$$\mathrm{core}\left(\begin{tikzcd} & x \arrow[dl, "a"'] \arrow[dr, "a"] & \\ y \arrow[loop left, "b"] \arrow[rr, "b"'] & & z \end{tikzcd}\right) = \begin{tikzcd} x \arrow[d, "a"] \\ y \arrow[loop left, "b"] \end{tikzcd}$$

Prop γ has semantic $tw \leq k$ IFF $\mathrm{core}(\gamma)$ has $tw \leq k$

Path queries



Conjunctive 2way regular path queries (C2RPQs)

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

Atoms: $x \xrightarrow{L} y$

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

also true for finite unions of C2RPQs

Prop Evaluation of C2RPQs is NP-complete...

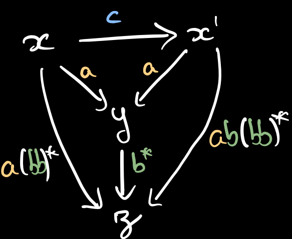
(combined complexity: input: database & query)

Semantic tree-width

Q: Given a C(2)RPQ, when is it equivalent to a finite union of C2RPQs of tree-width $\leq k$? "semantic tree-width $\leq k$"

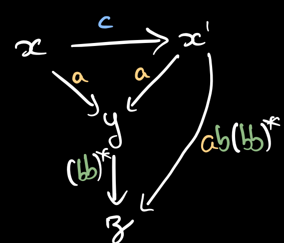
Pb CRPQs / C2RPQs cannot be minimized.

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

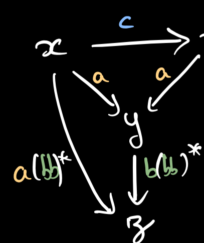


has tree-width 3...

$\equiv$



$\vee$



union of CRPQs tree-width ≤ 2

Deciding semantic tree-width

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

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

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

<u>Idea:</u> Show that the maximal under-approximation of γ by queries of tree-width $\leq k$ $\begin{cases} \text{exists} \\ \text{is computable.} \end{cases}$

How to build under-approximations of γ?

merging variables

refining atoms

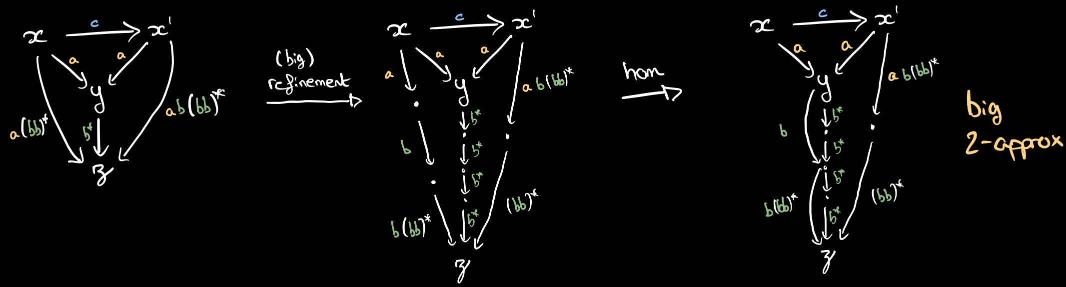
w/ $K_1 K_2 \subseteq L$

Def γ — $\boxed{\text{merging var. refinements}}$ $\to$ α — $\boxed{tw \leq k\,?}$ $\xrightarrow{\text{yes}}$ "2-approximation of γ"

The key Lemma

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

Proof goal Bound on size of refinement $\Rightarrow$ Bound on number of k-approximations



Properties of semantic tree-width

Theorem [Figueira, M., ICDT '23] γ : C2RPQ, $k \geq 2$.

γ is eq. to a simple query and is eq. to a query of $tw \leq k$

IFF

γ is eq. to a simple query of $tw \leq k$.

False for $k = 1$...

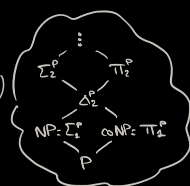
Simple regular expressions

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

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

UC2RPQ(SRE):
$$\begin{tikzcd} & \gamma \arrow[dl, "sup^*"'] \arrow[dr, "wr"] & \\ x & & y \end{tikzcd}$$
, etc.

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 $\leq k$ $\to$ 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]

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

Evaluation of $\mathcal{C}$ is FPT
IFF ?
$\mathcal{C}$ has bounded sem tree-width

holds for CQs

[Dalmau-Kolaitis-Vardi '02
& Grohe, '03]