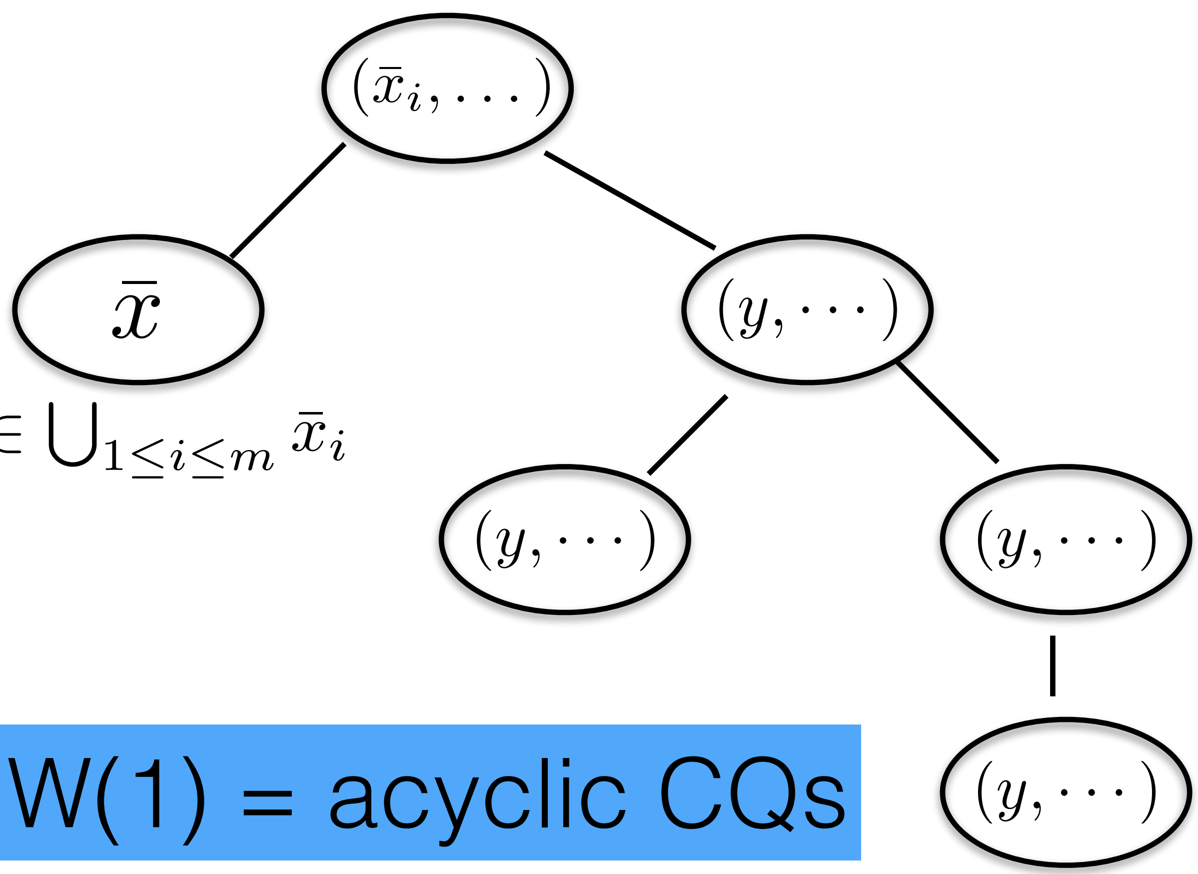
Open Problems: Semantic Optimization in Tractable Classes of CQs and CRPQs

Pablo Barceló
Center for Semantic Web Research &
DCC, University of Chile

CRPQs of Bounded Hypertreewidth

HW(k) := class of CQs that admit a *hypertree decomposition* of *width* at most k

Hypertree decomposition of a CQ $\exists \bar{y} \bigwedge_{1 \leq i \leq m} R_i(\bar{x}_i)$



$\bar{x} \in \bigcup_{1 \leq i \leq m} \bar{x}_i$

1. Each node is labeled with some variables from the CQ
2. The variables of each atom in the CQ appear together in a node
3. Appearances of variables are connected

Its width is:
max width of a node
(min set of atoms of q needed to cover the variables in the node)

The hypertreewidth of a CQ is the minimum width of its hypertree decompositions

HW(1) = acyclic CQs

Bounded hypertreewidth modulo equivalence

Given a CQ q, is there a $q' \in \mathsf{HW}(k)$ such that $q \equiv q'$?

Proposition

(B, Romero, Vardi, '16; similar ideas in Dalmau, Kolaitis, Vardi, '02)

The latter holds iff the <u>core</u> of q is in HW(k)

Core: Minimal subset of atoms of q that is equivalent to q

Corollary

If $q \equiv q'$ for $q' \in \mathsf{HW}(k)$, then $|q'| \leq |q|$ and can be computed in $2^{O(|q|)}$

Moreover, evaluation of q is *fixed-parameter tractable*

Checking bounded hypertreewidth modulo equivalence is NP-c

In the absence of constraints

Bounded hypertreewidth modulo equivalence

=

CQ minimization

Adding constraints

Yields a richer notion of semantic acyclicity

Example

The following CQ is not equivalent to an acyclic CQ:

$$\exists x, y, z\, A(x, y) \wedge B(y, z) \wedge C(z, x)$$

But it becomes so in the presence of the *full* tgd:

$$\forall x, y\, \big(A(x, y) \wedge B(y, z) \rightarrow C(z, x)\big)$$

It is equivalent to the following acyclic CQ under the tgd:

$$\exists x, y, z\, A(x, y) \wedge B(y, z)$$

Results for tgds
(B, Gottlob, Pieris, ’16)

Theorem

Being equivalent to a CQ in HW(1) under full tgds is undecidable

Theorem

Being equivalent to a CQ in HW(k) is decidable for guarded, sticky and non-recursive sets of tgds (2EXP, EXP, NEXP, resp)

Results for egds

Theorem (unpublished)
Being equivalent to a CQ in HW(1) under egds
is undecidable

Theorem (Figueira, ’16)
Being equivalent to a CQ in HW(k)
under unary keys over schemas of arity at most two
is decidable (2EXP)

Open question

Decidability status of
the problem under keys/FDs

Formal statement

Given a CQ q and a finite set Σ of keys/FDs, is there a $q' \in \mathsf{HW}(k)$ such that $q \equiv_{\Sigma} q'$?

Conjunctive Regular Path Queries (CRPQs)

Evaluated over *graph databases*
(Edge-labeled directed graphs,
or databases over a schema of binary relations)

CRPQs extend CQs over graph databases
they can check if a pair of nodes
is linked by a path labeled by a reg exp

CRPQs are expressions of the form:

$$\exists \bar{z} \bigwedge_{1 \leq i \leq m} L_i(x_i, y_i)$$

Its hypertreewidth corresponds to the one of
its *underlying* CQ

Open question

Bounded hypertreewidth modulo equivalence

Given a CRPQ q, is there a CRPQ $q' \in \mathsf{HW}(k)$ such that $q \equiv q'$?

If CRPQs are extended with unions and inverses (UC2RPQs):
EXPSPACE-c for k = 1 (B, Romero, Vardi, '16)