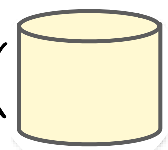
Logical Structures in Computation Reunion Workshop
Berkeley, CA
2017-12-11

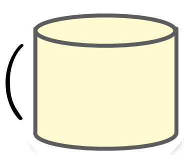
Semantic Tractability *of* Conjunctive Queries

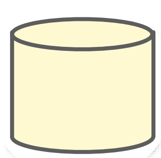
Diego Figueira
CNRS, LaBRI

Query optimization

☹ $\mathcal{C} \ni Q($  $)$ hard to evaluate

⤸

😀 $\mathcal{C}' \ni Q'$, with $Q \equiv Q'$ where $Q'($  $)$ efficient

 = relational structure

$\mathcal{C}$ = Conjunctive Queries

$\mathcal{C}'$ = Conjunctive Queries of ***small treewidth***

Conjunctive Queries

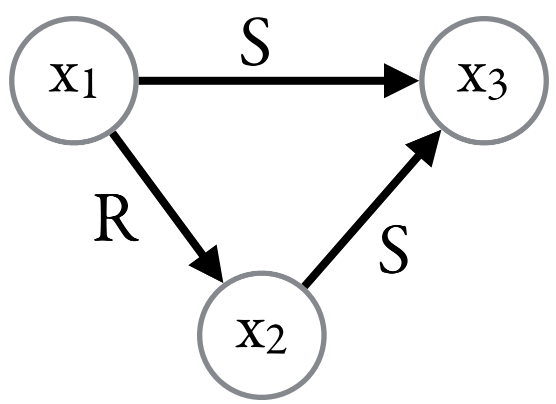
CQ = first-order formula

$$\exists\, x_1,\ldots,x_n \,.\, \bigwedge \,(\mathit{atoms})$$

Basic query language on databases

e.g.

$\exists\, x_1,x_2,x_3\ R(x_1,x_2) \wedge S(x_1,x_3) \wedge S(x_2,x_3)$



"Canonical structure"

evaluation of CQ ≈ ∃ homomorphism on relational structures

$\phi \dashrightarrow C_\phi$ s.t. $\forall S$: $S \models \phi$ *iff* $C_\phi \longrightarrow S$

$\phi_A \dashleftarrow A$ s.t. $\forall S$: $S \models \phi_A$ *iff* $A \longrightarrow S$

[Chandra, Merlin, '77]

Conjunctive Queries: syntactic restrictions

☹ Evaluation for CQ's is intractable: NP-complete, W[1]-hard

Evaluating ϕ on **A** takes $|\mathbf{A}|^{|\phi|}$ time $(\neq 2^{|\phi|} \cdot |\mathbf{A}|^{const})$

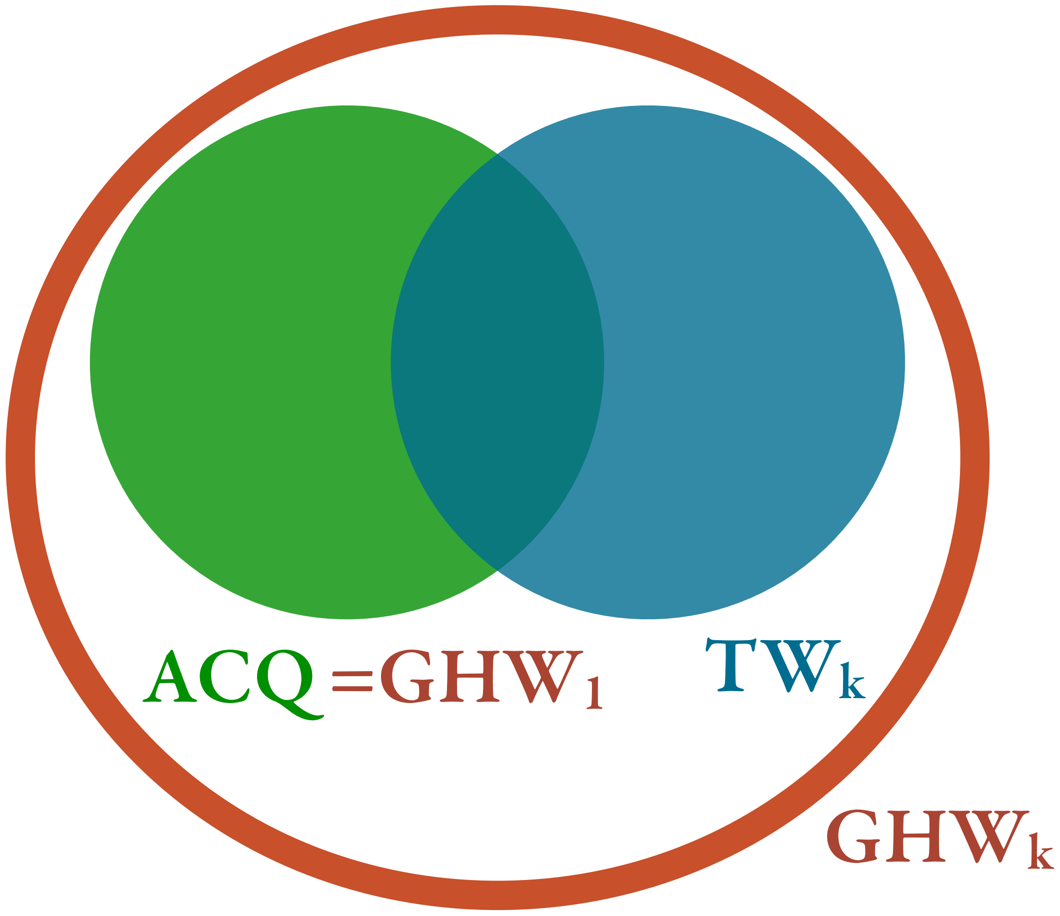
Effort to find tractable syntactic fragments:

- ✻ **Acyclic** CQ (ACQ)
- ✻ **Bounded treewidth** CQ's (TW_k)
- ✻ **Bounded generalized hyper-treewidth / coverwidth** CQ's (GHW_k)

(CQ's that "resemble a tree")

😀 Evaluation for ACQ / TW_k / GHW_k queries is tractable

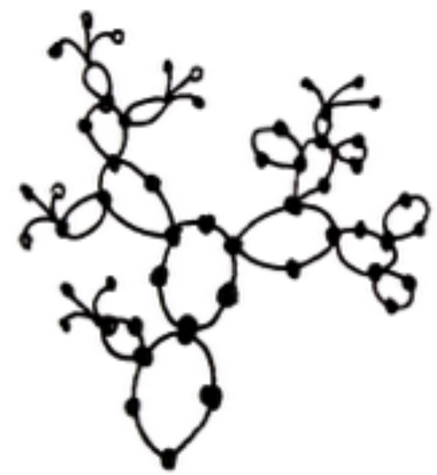
Conjunctive Queries: syntactic restrictions



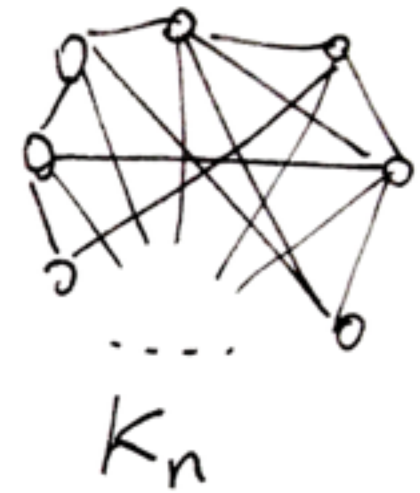
Treewidth, examples

(a tree)

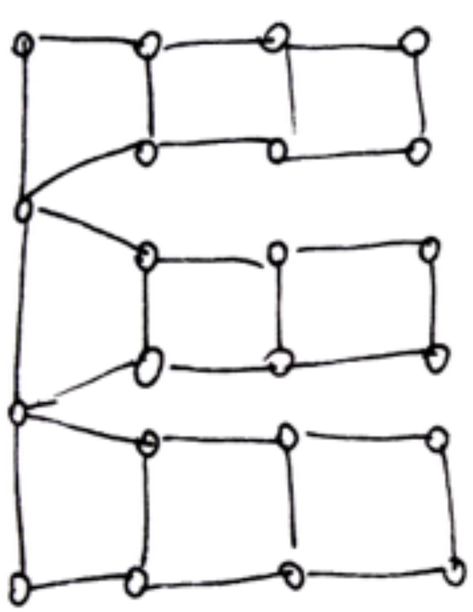
tree-width 1



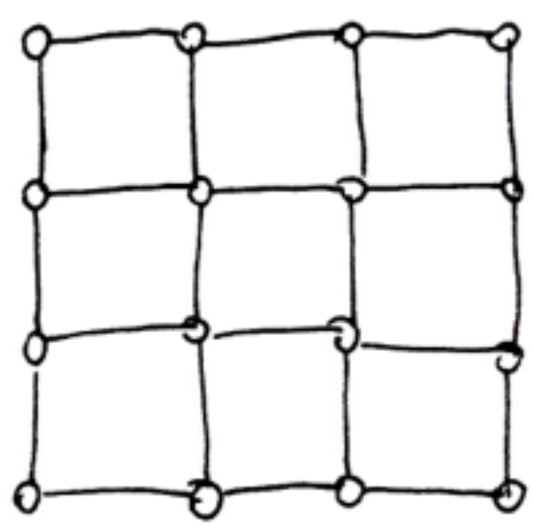
tree-width 2



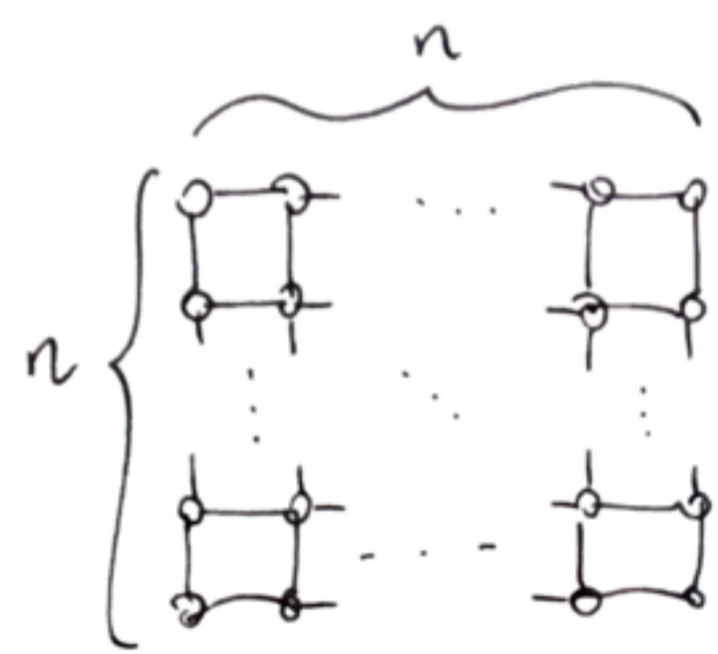
tree-width $n-1$



tree-width 2



tree-width 4



tree-width n

Conjunctive Queries: **semantic** restrictions

Evaluation for acyclic/TW_k/GHW_k CQs is $O(|\phi| \cdot |D|^{k+1})$

[Yannakakis, '81]
[Chekuri, Rajaraman, '00]
[Gottlob, Leone, Scarcello, '02]

What about *semantic* conditions?

If $\phi \equiv \psi$ with $\psi \in GHW_k$: instead of evaluating $\phi(A)$ in $TIME(|A|^{|\phi|})$, evaluate $\psi(A)$ in $TIME(|\psi| \cdot |A|^{const})$

(fixed-parameter tractable)

Optimization problem: "***find a well-behaved equivalent query***"

Semantic tree-width problem

Semantic width-k problem

Input: $\phi \in CQ$
Output: Is there an $\mathbf{GHW}_k$ CQ ψ so that $\psi \equiv \phi$?

The semantic width-k problem is NP-complete.

[Dalmau, Kolaitis, Vardi, '02]

Evaluation of semantically width-k queries is in PTime

[Chen, Dalmau, '05]

Semantic width-k problem under constraints

Semantic width-k problem under $\mathscr{C}$

Input: $\phi \in CQ, \Sigma \subseteq \mathscr{C}$

Output: Is there a CQ $\psi \in GHW_k$ so that $\psi \equiv_\Sigma \phi$?

Equivalence problem under $\mathscr{C}$ undecidable $\not\Leftarrow$ $\Rightarrow$ semantic GHW_k problem undecidable

[Barceló, Gottlob, Pieris, '16]

($\mathscr{C}$ = any class of tgds)

$$\forall x_1,\ldots,x_n \,.\, \phi(x_1,\ldots,x_n) \Rightarrow \exists y_1,\ldots,y_m\, \psi(x_1,\ldots,x_n,y_1,\ldots,y_m)$$

$\bigwedge$ Atoms

Semantic width-k problem under constraints

Semantic width-k problem under $\mathcal{C}$

Input: $\phi \in \mathrm{CQ}, \Sigma \subseteq \mathcal{C}$

Output: Is there a CQ $\psi \in \mathrm{GHW}_k$ so that $\psi \equiv_\Sigma \phi$?

When $\mathcal{C}$ is defined by **full tgds** the problem is **undecidable**.

[Barceló, Gottlob, Pieris, '16]

$$\forall\, x_1,\ldots,x_n \,.\, \phi(x_1,\ldots,x_n) \Rightarrow \psi(x_1,\ldots,x_n)$$

$\bigwedge$Atoms

Semantic width-k problem

Semantic width-k problem under $\mathscr{C}$

Input: $\phi \in \mathrm{CQ}, \Sigma \subseteq \mathscr{C}$

Output: Is there a CQ $\psi \in \mathrm{GHW}_k$ so that $\psi \equiv_\Sigma \phi$?

When $\mathscr{C}$ is defined by **non-recursive/guarded/sticky tgds** the **semantic width-k** under $\mathscr{C}$ is **decidable**

[Barceló, Gottlob, Pieris, '16]

guarded tgds: **2ExpTime** (**NP** if schema fixed)

non-recursive tgds: **NExpTime** (**NP** if schema fixed)

sticky tgds: **NExpTime** (**NP** if schema fixed)

Semantic width-k problem

Semantic width-k problem under $\mathcal{C}$

Input: $\phi \in CQ, \Sigma \subseteq \mathcal{C}$

Output: Is there a CQ $\psi \in GHW_k$ so that $\psi \equiv_\Sigma \phi$?

When $\mathcal{C}$ is defined by **egds** the problem is **undecidable**.

[Barceló, Gottlob, Pieris, '16]

When $\mathcal{C}$ is defined by **unary functional dependencies** it is **decidable**

[F, '16]

$$\forall x_1,\ldots,x_n,y_1,\ldots,y_n \,.\, R(x_1,\ldots,x_n) \wedge R(y_1,\ldots,y_n) \wedge (x_i = y_i) \Rightarrow (x_j = y_j)$$

"**R[i⟶j]**" : in relation R the ***i***-th component determines the ***j***-th component

Semantic TW/acyclicity problem

Semantic width-k problem under UFD

Input: $\phi \in CQ, \Sigma = \{R_1[i_1 \longrightarrow j_1], \ldots, R_n[i_n \longrightarrow j_n]\}$

Output: Is there a CQ $\psi \in GHW_k$ so that $\psi \equiv_\Sigma \phi$?

The semantic width-k problem under **unary functional dependencies** is **decidable in 2ExpTime**, for every *k*

[F, '16]

(If possible, algorithm returns a CQ $\in GHW_k$)

Evaluation of CQs

☞ $\phi \equiv \psi$ iff $\text{core}(\mathbf{C}_\phi) \cong \text{core}(\mathbf{C}_\psi)$

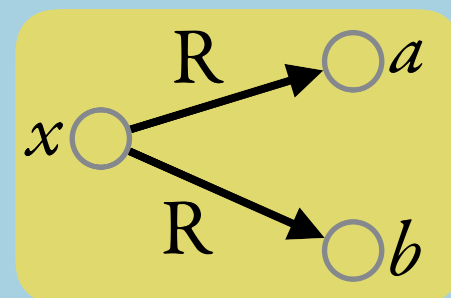
☞ And for Σ : UFD,

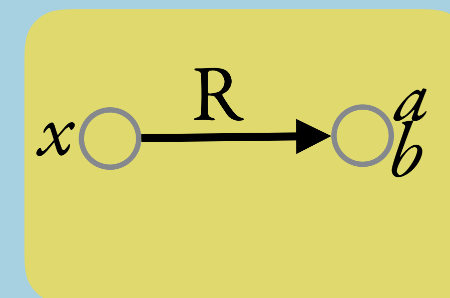
$$\phi \equiv_\Sigma \psi \quad \text{iff} \quad \text{core}(\textbf{chase}_\Sigma(\mathbf{C}_\phi)) \cong \text{core}(\textbf{chase}_\Sigma(\mathbf{C}_\psi))$$

Chase

chase$_\Sigma$(**A**) = The result of repeating:

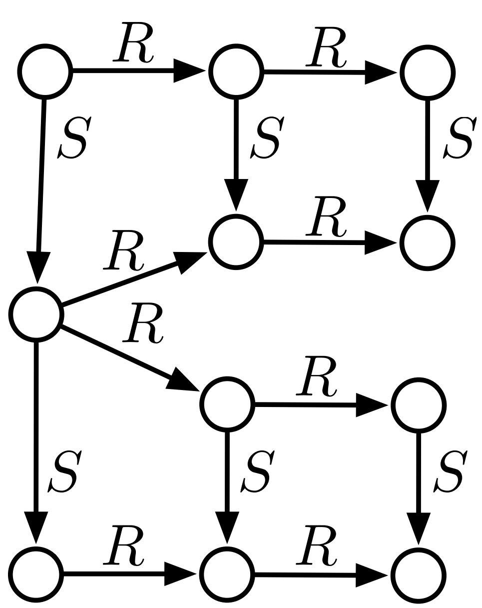
$(x,a) \in \mathrm{R}, (x,b) \in \mathrm{R}$

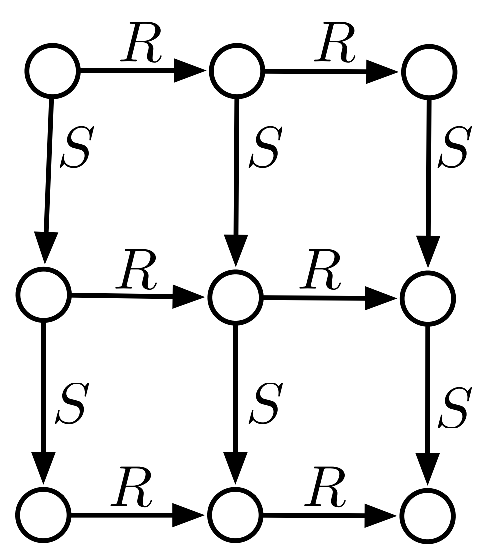
If and $\mathrm{R}[1 \longrightarrow 2] \in \Sigma$, and 

, collapse a,b 

.

e.g.

chase$_\Sigma$(

) = 

for $\Sigma = \{\mathrm{R}[1 \longrightarrow 2]\}$

Restatement of semantic width-k problem under UFD

$\phi \equiv_\Sigma \psi$ iff $\mathrm{core}(\mathbf{chase}_\Sigma(\mathbf{C}_\phi)) \cong \mathrm{core}(\mathbf{chase}_\Sigma(\mathbf{C}_\psi))$

Restatement of our problem:

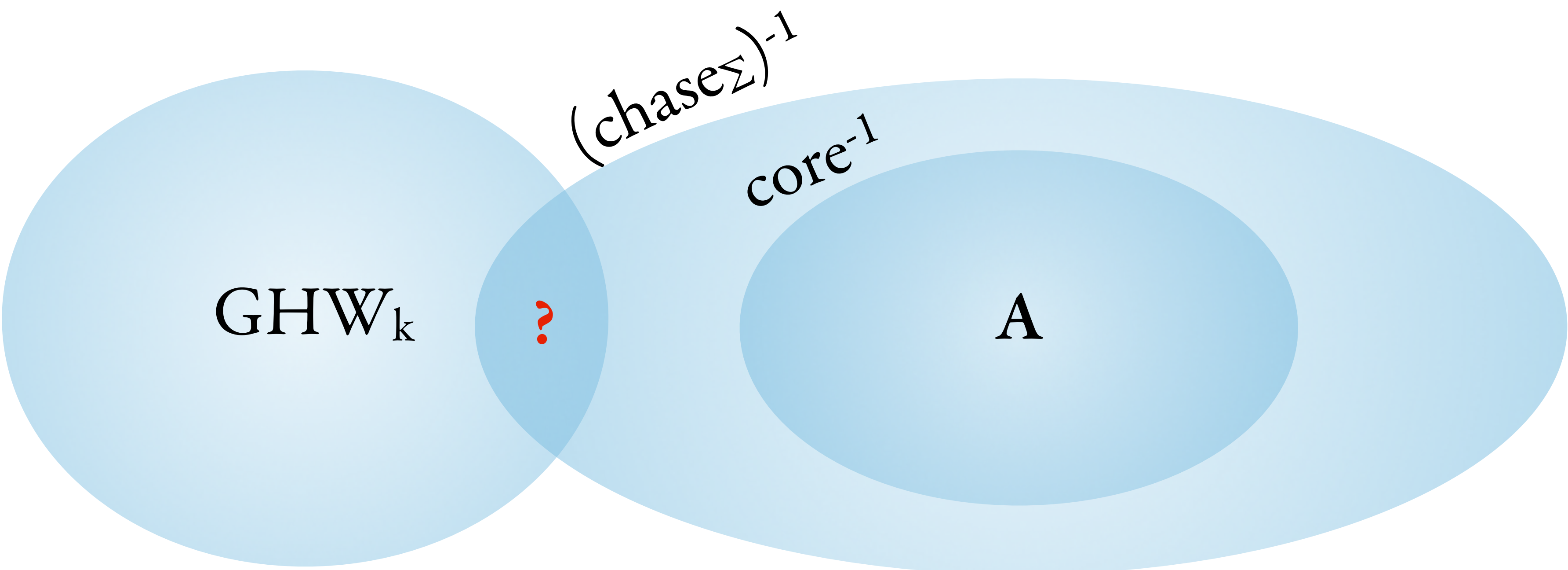
core-chase problem

Input: $\mathbf{A} \in \mathrm{STR}$, $\Sigma \subseteq \mathrm{UFD}$

Output: $\mathrm{GHW}_k \cap \{ \mathbf{B} \in \mathrm{STR} \mid \mathrm{core}(\mathrm{chase}_\Sigma(\mathbf{B})) \cong \mathbf{A} \} = \varnothing$?

$\mathbf{A} = \mathrm{core}(\mathrm{chase}_\Sigma(\mathbf{C}_\phi))$

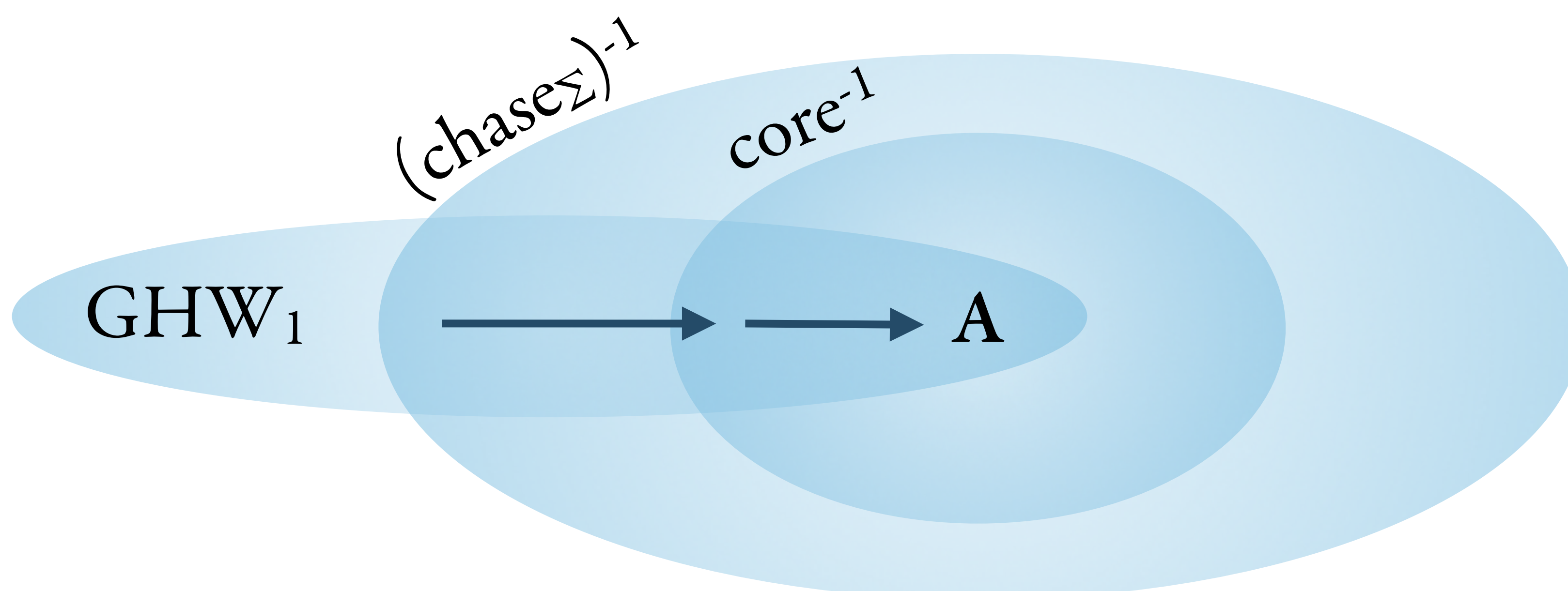
$\mathbf{B} = \mathbf{C}_\psi$



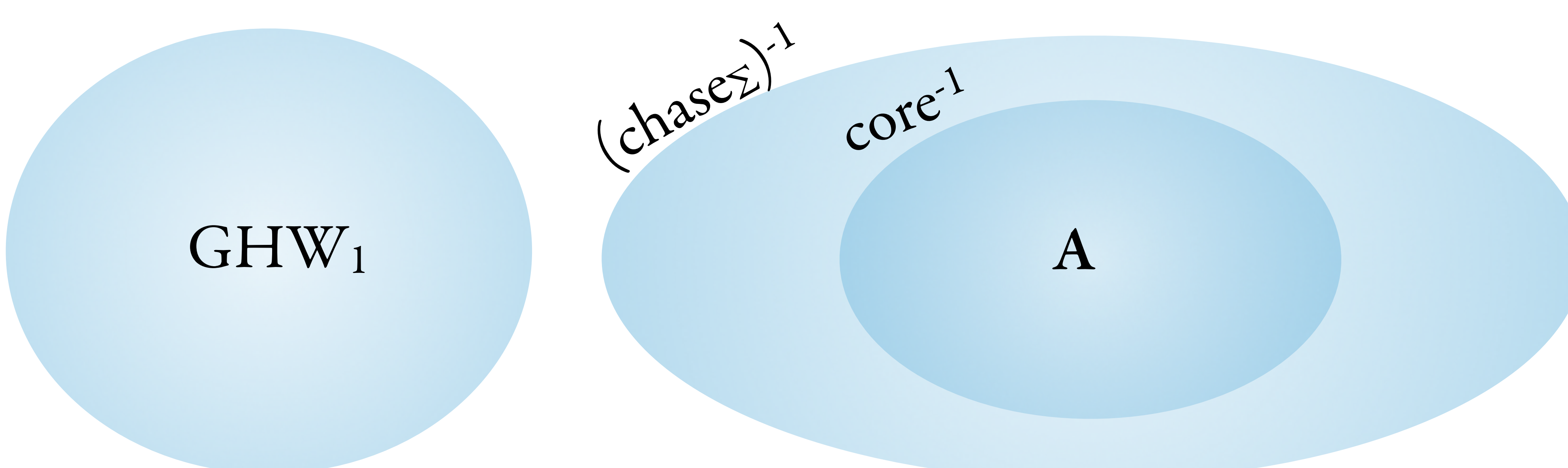
An easy case: GHW_1

core
$chase_\Sigma$ preserve GHW_1 $\Rightarrow$ simply check if input structure $\mathbf{A} \in GHW_1$

either



or

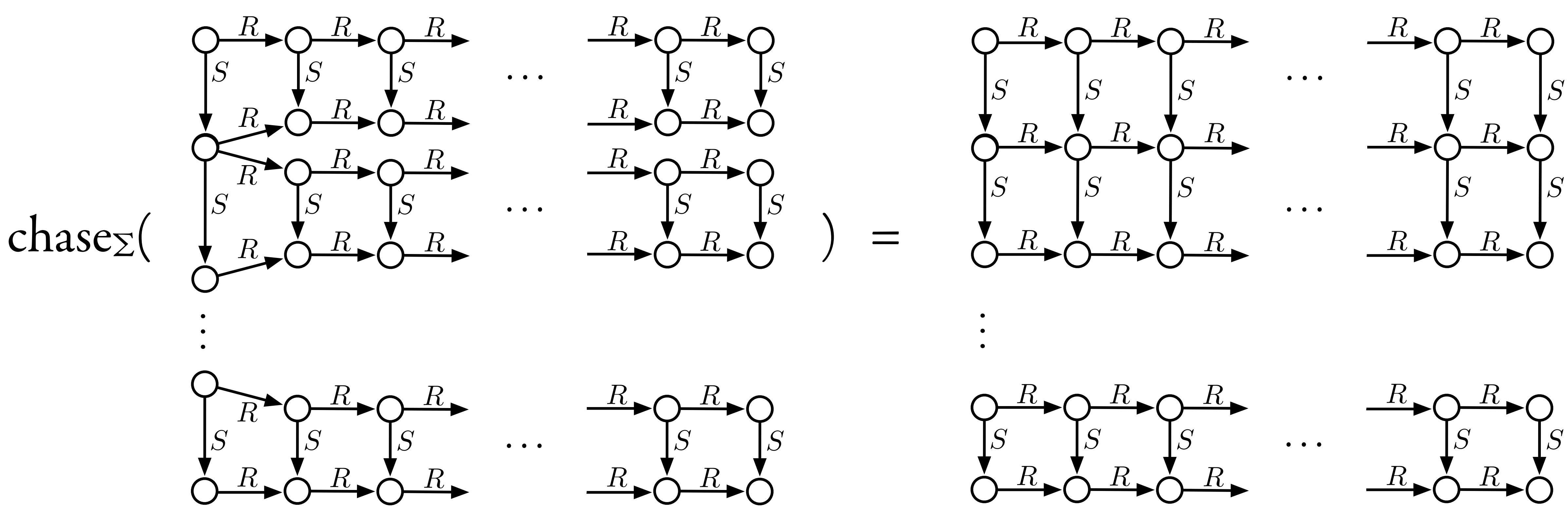


The general case, for arbitrary k

However, this does not hold in general

for any n : chase(sth of width 2) = (sth of width n)

e.g.



for $\Sigma = \{ R[1 \longrightarrow 2] \}$

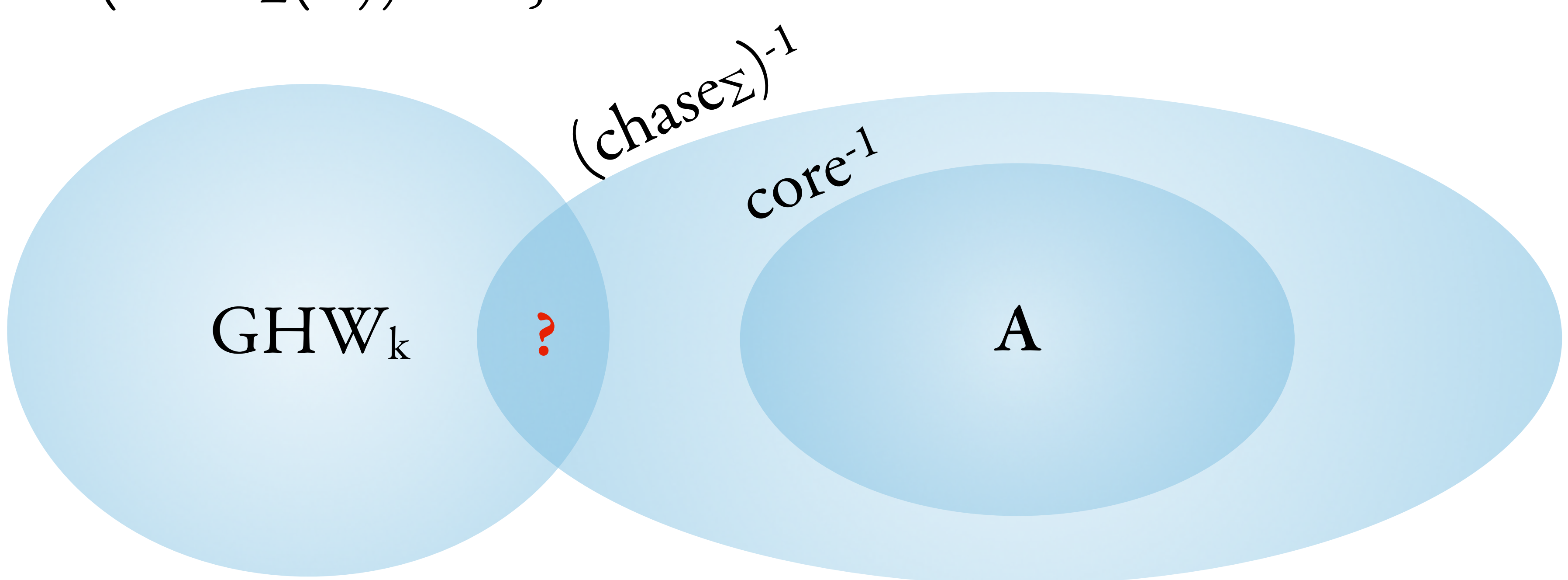
Solving the problem

$GHW_k \cap \{ \mathbf{B} \in STR \mid core(chase_\Sigma(\mathbf{B})) \cong \mathbf{A} \} = \varnothing$?

First approach:

- define $\{\mathbf{B} \mid core(chase_\Sigma(\mathbf{B})) \cong \mathbf{A}\}$ with $\phi_A \in MSO$
- test $GHW_k \ni C \vDash \phi_A$ for some C [Seese '91]

But $\{\mathbf{B} \mid core(chase_\Sigma(\mathbf{B})) \cong \mathbf{A}\}$ is not MSO-definable ☹



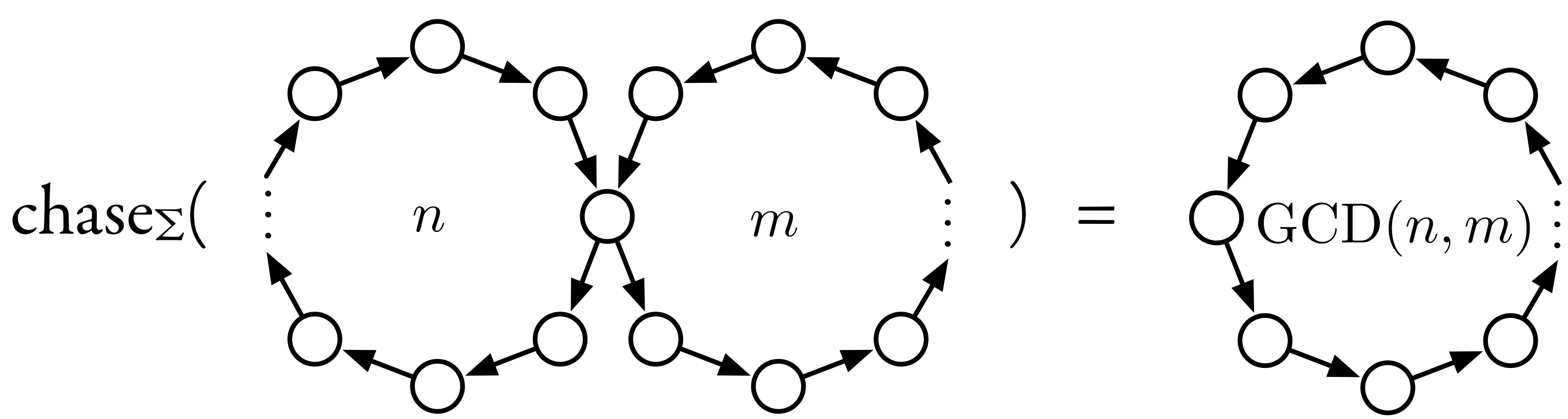
Solving the problem

$\mathrm{GHW}_k \cap \{\, \mathbf{B} \in \mathrm{STR} \mid \mathrm{core}(\mathrm{chase}_\Sigma(\mathbf{B})) \cong \mathbf{A} \,\} = \varnothing$?

First approach:

- define $\{\mathbf{B} \mid \mathrm{core}(\mathrm{chase}_\Sigma(\mathbf{B})) \cong \mathbf{A}\}$ with $\phi_A \in \mathrm{MSO}$
- test $\mathrm{GHW}_k \ni \mathrm{C} \models \phi_A$ for some C [Seese '91]

But $\{\mathbf{B} \mid \mathrm{core}(\mathrm{chase}_\Sigma(\mathbf{B})) \cong \mathbf{A}\}$ is not MSO-definable ☹



for $\Sigma = \{\, R[1 \longrightarrow 2] \,\}$

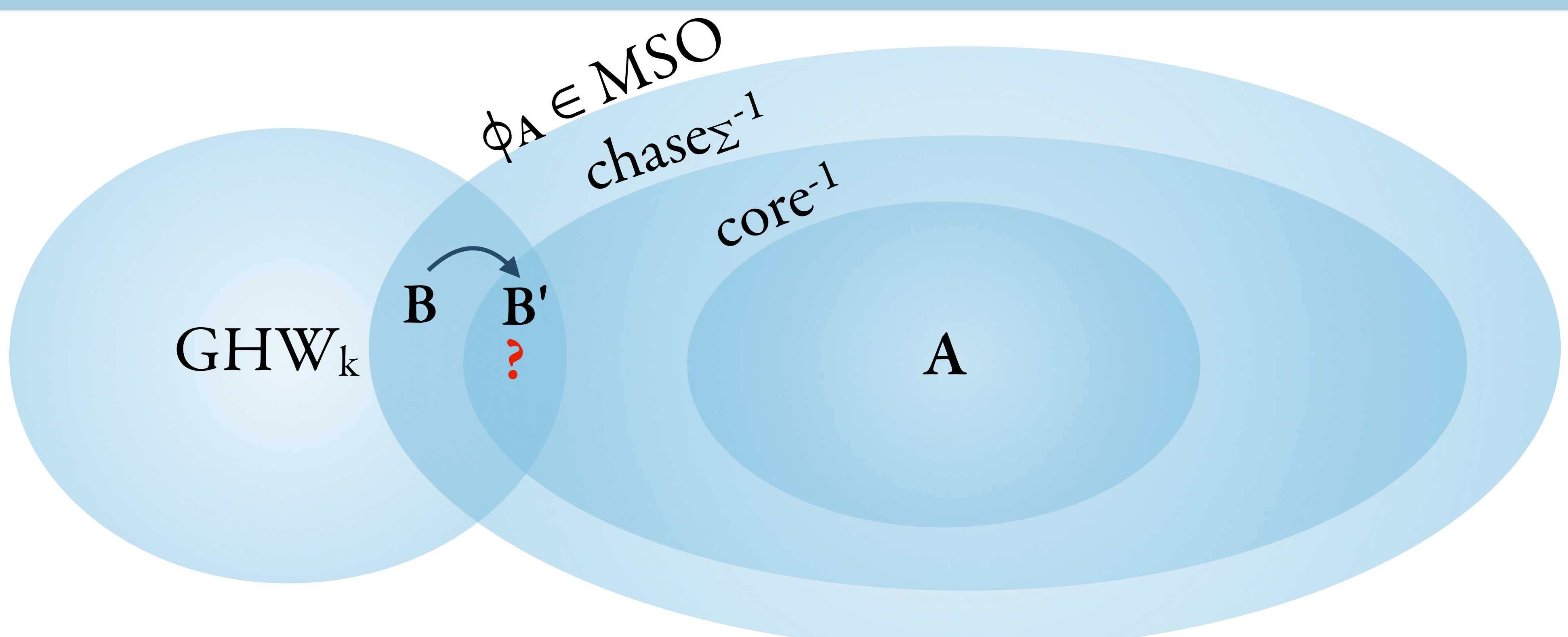
Solving the problem, now for real

$$GHW_k \cap \{ \mathbf{B} \in STR \mid core(chase_\Sigma(\mathbf{B})) \cong \mathbf{A} \} = \varnothing\ ?$$

How it's actually solved: define $\phi_A \in MSO$ so that

- if $core(chase_\Sigma(\mathbf{B})) \cong \mathbf{A}$ then $\mathbf{B} \models \phi_A$,
- if $\mathbf{B} \in TW_k$, $\mathbf{B} \models \phi_A$ then there is some extension $\mathbf{B'}$ of $\mathbf{B}$ so that
  - ✻ tree-width($\mathbf{B'}$) = tree-width($\mathbf{B}$)
  - ✻ $core(chase_\Sigma(\mathbf{B'})) \cong \mathbf{A}$

. . . and test $GHW_k \ni \mathbf{C} \models \phi$ for some $\mathbf{C}$



Final comments

- Exact **complexity** for semantic width-k problem under UFD?

  (Between NP and 2ExpTime)

- Generalization to treating **constants**, **free variables**, UCQ
- **Non-unary** functional dependencies?
- **Working** optimization procedure?
- Extending to bounded fractional hyper-trewidth?