Finite Model Theory
Lecture 5: Query Containment

Spring 2025

Announcements

- Good job on homework 1!
- Homework 2 is posted and due next Friday, 4/25.

Where We Are

- Finite models: we may write $\varphi \equiv \psi$ to mean $\varphi \equiv_{\text{fin}} \psi$.
- Trakhtenbrot: $\texttt{SAT}_{\text{fin}}(\varphi)$ is undecidable.
- Model checking $\boldsymbol{A} \models \varphi$
  - FO: data complexity AC^0, expression complexity PSPACE-complete
  - CQ: data complexity AC^0, expression complexity NP-complete
- We often use "database", "query" instead of "structure", "formula".

Today: Query Containment/Equivalence

Problem Definition

Query Equivalence

Definition (Equivalence)

Q_1, Q_2 are equivalent if $\forall \boldsymbol{D}$, $Q_1(\boldsymbol{D}) = Q_2(\boldsymbol{D})$. Notation: $Q_1 \equiv Q_2$.

Query Equivalence

Definition (Equivalence)

Q_1, Q_2 are equivalent if $\forall \boldsymbol{D}$, $Q_1(\boldsymbol{D}) = Q_2(\boldsymbol{D})$. Notation: $Q_1 \equiv Q_2$.

Example: $Q_1 \equiv Q_2$ where,

$$Q_1(x) = \exists y \exists z \exists u (E(x,y) \wedge E(y,x) \wedge E(y,u))$$
$$Q_2(x) = \exists y \exists z \exists v (E(x,y) \wedge E(y,x) \wedge E(x,v))$$

Query Equivalence

Definition (Equivalence)

Q_1, Q_2 are equivalent if $\forall \boldsymbol{D}$, $Q_1(\boldsymbol{D}) = Q_2(\boldsymbol{D})$. Notation: $Q_1 \equiv Q_2$.

Example: $Q_1 \equiv Q_2$ where,

$$Q_1(x) = \exists y \exists z \exists u (E(x,y) \wedge E(y,x) \wedge E(y,u))$$
$$Q_2(x) = \exists y \exists z \exists v (E(x,y) \wedge E(y,x) \wedge E(x,v))$$

It suffices to study equivalence of Boolean queries, because:
$Q_1(\boldsymbol{x}) \equiv Q_2(\boldsymbol{y})$ iff $(|\boldsymbol{x}| = |\boldsymbol{y}|)$ and for constants $\boldsymbol{c}$, $Q_1[\boldsymbol{c}/\boldsymbol{x}] \equiv Q_2[\boldsymbol{c}/\boldsymbol{y}]$.

Query Containment

Definition (Containment)

Q_1 is contained in Q_2 if $\forall \boldsymbol{D}$, $Q_1(\boldsymbol{D}) \subseteq Q_2(\boldsymbol{D})$. Notation: $Q_1 \subseteq Q_2$

Again, suffices if Q_1, Q_2 are Boolean. Then $Q_1 \subseteq Q_2$ same as $Q_1 \Rightarrow Q_2$.

Query Containment

Definition (Containment)

Q_1 is contained in Q_2 if $\forall \boldsymbol{D}$, $Q_1(\boldsymbol{D}) \subseteq Q_2(\boldsymbol{D})$. Notation: $Q_1 \subseteq Q_2$

Again, suffices if Q_1, Q_2 are Boolean. Then $Q_1 \subseteq Q_2$ same as $Q_1 \Rightarrow Q_2$.

Example:

$$Q_1 = \exists x \exists y \exists z \exists u (E(x,y) \wedge E(y,z) \wedge E(z,u))$$
$$Q_2 = \exists x \exists y \exists z (E(x,y) \wedge E(y,z))$$

$$Q_1 \subseteq Q_2$$

Containment v.s. Equivalence

Fact

Equivalence and containment are (almost) the same problem:

$$\boxed{Q_1 \equiv Q_2} \text{ iff } \boxed{Q_1 \subseteq Q_2 \text{ and } Q_2 \subseteq Q_1}$$

$$\boxed{Q_1 \subseteq Q_2} \text{ iff}^1 \boxed{Q_1 \equiv Q_1 \wedge Q_2}$$

[1] Language must be closed under $\wedge$.

Containment for FO is Undecidable

Theorem

The problem Given Q_1, Q_2 in FO check whether $Q_1 \subseteq Q_2$ is undecidable.

Containment for FO is Undecidable

Theorem

The problem Given Q_1, Q_2 in FO check whether $Q_1 \subseteq Q_2$ is undecidable.

Proof By reduction from $\texttt{SAT}_{\text{fin}}$.

Let Φ be any sentence. (We want to check $\texttt{SAT}_{\text{fin}}(\Phi)$.)

Define $Q_1 \stackrel{\text{def}}{=} \Phi$ and $Q_2 \stackrel{\text{def}}{=} \texttt{FALSE}$. Then $Q_1 \subseteq Q_2$ iff $\neg\texttt{SAT}_{\text{fin}}(\Phi)$.

Containment for CQs

Containment for CQs

The containment problem for CQs is decidable; More precisely, NP-complete.

This is one of the oldest, most celebrated result in database theory. It is due to Chandra and Merlin, 1977.

Review: Equivalent Concepts

- Boolean Conjunctive Query:
  $Q = R(x, y, z) \wedge S(x, u) \wedge S(y, v) \wedge S(z, w) \wedge R(u, v, w)$

Review: Equivalent Concepts

- Boolean Conjunctive Query:
  $Q = R(x, y, z) \wedge S(x, u) \wedge S(y, v) \wedge S(z, w) \wedge R(u, v, w)$

- A Database $\boldsymbol{A}$ with domain $\texttt{Vars}(Q)$:

$\text{Dom}(\boldsymbol{A}) = \{x, y, z, u, v, w\}$

$R^A =$

x	y	z
u	v	w

$S^A =$

x	u
y	v
z	w

Review: Equivalent Concepts

- Boolean Conjunctive Query:
  $Q = R(x,y,z) \wedge S(x,u) \wedge S(y,v) \wedge S(z,w) \wedge R(u,v,w)$

- A Database $\boldsymbol{A}$ with domain $\texttt{Vars}(Q)$:

$\text{Dom}(\boldsymbol{A}) = \{x, y, z, u, v, w\}$

$R^A =$

x	y	z
u	v	w

$S^A =$

x	u
y	v
z	w

If $Q(\boldsymbol{x})$ has head variables $\boldsymbol{x}$, then we add named constants $\boldsymbol{x}^A$ to $\boldsymbol{A}$.

Review: Equivalent Concepts

- Boolean Conjunctive Query:
  $Q = R(x,y,z) \wedge S(x,u) \wedge S(y,v) \wedge S(z,w) \wedge R(u,v,w)$

- A Database $\boldsymbol{A}$ with domain $\texttt{Vars}(Q)$:

$\text{Dom}(\boldsymbol{A}) = \{x,y,z,u,v,w\}$

$R^A =$

x	y	z
u	v	w

$S^A =$

x	u
y	v
z	w

If $Q(\boldsymbol{x})$ has head variables $\boldsymbol{x}$, then we add named constants $\boldsymbol{x}^A$ to $\boldsymbol{A}$.

We call $\boldsymbol{A}$ the Canonical Database of Q, denoted $\boxed{\boldsymbol{D}_Q}$

Review: Homomorphisms

A homomorphisms between two structures $h : \boldsymbol{A} \to \boldsymbol{B}$
is a function $h : \text{Dom}(\boldsymbol{A}) \to \text{Dom}(\boldsymbol{B})$ s.t.

$$h(R^A) \subseteq h(R^B) \text{ for all relation names } R\text{, and}$$
$$h(c^A) = c^B \text{ for all constant names } c.$$

Notice that $\boldsymbol{A}$ or $\boldsymbol{B}$ can be a CQ Q, or a database $\boldsymbol{D}$.

Review: Homomorphisms

A homomorphisms between two structures $h : \boldsymbol{A} \to \boldsymbol{B}$
is a function $h : \text{Dom}(\boldsymbol{A}) \to \text{Dom}(\boldsymbol{B})$ s.t.

$h(R^A) \subseteq h(R^B)$ for all relation names R, and
$h(c^A) = c^B$ for all constant names c.

Notice that $\boldsymbol{A}$ or $\boldsymbol{B}$ can be a CQ Q, or a database $\boldsymbol{D}$.

Main property: $\boxed{\boldsymbol{D} \models Q}$ iff $\boxed{\exists h : Q \to \boldsymbol{D}}$

Review: Homomorphisms

A homomorphisms between two structures $h : \boldsymbol{A} \to \boldsymbol{B}$
is a function $h : \text{Dom}(\boldsymbol{A}) \to \text{Dom}(\boldsymbol{B})$ s.t.

$h(R^A) \subseteq h(R^B)$ for all relation names R, and
$h(c^A) = c^B$ for all constant names c.

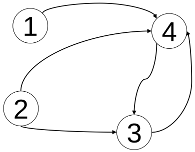
Notice that $\boldsymbol{A}$ or $\boldsymbol{B}$ can be a CQ Q, or a database $\boldsymbol{D}$.

Main property: $\boxed{\boldsymbol{D} \models Q}$ iff $\boxed{\exists h : Q \to \boldsymbol{D}}$

Example:

$Q = (E(x,y) \wedge E(y,z) \wedge E(x,z)$

$\boldsymbol{D} =$



Review: Homomorphisms

A homomorphisms between two structures $h : \boldsymbol{A} \to \boldsymbol{B}$
is a function $h : \text{Dom}(\boldsymbol{A}) \to \text{Dom}(\boldsymbol{B})$ s.t.

$h(R^A) \subseteq h(R^B)$ for all relation names R, and
$h(c^A) = c^B$ for all constant names c.

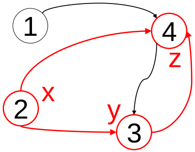
Notice that $\boldsymbol{A}$ or $\boldsymbol{B}$ can be a CQ Q, or a database $\boldsymbol{D}$.

Main property: $\boxed{\boldsymbol{D} \models Q}$ iff $\boxed{\exists h : Q \to \boldsymbol{D}}$

Example:

$Q = (E(x,y) \wedge E(y,z) \wedge E(x,z))$

$x \mapsto 2; y \mapsto 3; z \mapsto 4$

$\boldsymbol{D} =$ 

The Canonical Database

$\boldsymbol{D}_Q$ is just the query Q viewed as a database.

Main property: $\boxed{\boldsymbol{D}_Q \models Q}$

Proof: the identity function $Q \to \boldsymbol{D}_Q$ is a homomorphism.

Containment Theorem for Boolean CQs

Theorem (Chandra and Merlin, 1977)

The following are equivalent:

(1) Containment holds: $Q_1 \subseteq Q_2$. *(if* $\boldsymbol{D} \models Q_1$ *then* $\boldsymbol{D} \models Q_2$*)*

(2) $\boldsymbol{D}_{Q_1} \models Q_2$.

(3) There exists a homomorphism $h : Q_2 \to Q_1$

Containment Theorem for Boolean CQs

Theorem (Chandra and Merlin, 1977)

The following are equivalent:

(1) Containment holds: $Q_1 \subseteq Q_2$. *(if* $\boldsymbol{D} \vDash Q_1$ *then* $\boldsymbol{D} \vDash Q_2$*)*

(2) $\boldsymbol{D}_{Q_1} \vDash Q_2$.

(3) There exists a homomorphism $h : Q_2 \to Q_1$

Proof

(1) $\Rightarrow$ (2)

Containment Theorem for Boolean CQs

Theorem (Chandra and Merlin, 1977)

The following are equivalent:

(1) Containment holds: $Q_1 \subseteq Q_2$. *(if* $\boldsymbol{D} \vDash Q_1$ *then* $\boldsymbol{D} \vDash Q_2$*)*

(2) $\boldsymbol{D}_{Q_1} \vDash Q_2$.

(3) There exists a homomorphism $h : Q_2 \to Q_1$

Proof

(1) ⇒ (2) If $\boldsymbol{D}_{Q_1} \vDash Q_1$ and $Q_1 \subseteq Q_2$ then $\boldsymbol{D}_{Q_1} \vDash Q_2$

Containment Theorem for Boolean CQs

Theorem (Chandra and Merlin, 1977)

The following are equivalent:

(1) Containment holds: $Q_1 \subseteq Q_2$. *(if* $\boldsymbol{D} \models Q_1$ *then* $\boldsymbol{D} \models Q_2$*)*

(2) $\boldsymbol{D}_{Q_1} \models Q_2$.

(3) There exists a homomorphism $h : Q_2 \rightarrow Q_1$

Proof

(1) $\Rightarrow$ (2) If $\boldsymbol{D}_{Q_1} \models Q_1$ and $Q_1 \subseteq Q_2$ then $\boldsymbol{D}_{Q_1} \models Q_2$

(2) $\Rightarrow$ (3)

Containment Theorem for Boolean CQs

Theorem (Chandra and Merlin, 1977)

The following are equivalent:

(1) Containment holds: $Q_1 \subseteq Q_2$. *(if* $\boldsymbol{D} \models Q_1$ *then* $\boldsymbol{D} \models Q_2$*)*

(2) $\boldsymbol{D}_{Q_1} \models Q_2$.

(3) There exists a homomorphism $h : Q_2 \to Q_1$

Proof

(1) $\Rightarrow$ (2) If $\boldsymbol{D}_{Q_1} \models Q_1$ and $Q_1 \subseteq Q_2$ then $\boldsymbol{D}_{Q_1} \models Q_2$

(2) $\Rightarrow$ (3) If $\boldsymbol{D}_{Q_1} \models Q_2$ then $\exists h : Q_2 \to \boldsymbol{D}_{Q_1}$, hence $\exists h : Q_2 \to Q_1$.

Containment Theorem for Boolean CQs

Theorem (Chandra and Merlin, 1977)

The following are equivalent:

(1) Containment holds: $Q_1 \subseteq Q_2$. (if $\boldsymbol{D} \models Q_1$ then $\boldsymbol{D} \models Q_2$)

(2) $\boldsymbol{D}_{Q_1} \models Q_2$.

(3) There exists a homomorphism $h : Q_2 \to Q_1$

Proof

(1) $\Rightarrow$ (2) If $\boldsymbol{D}_{Q_1} \models Q_1$ and $Q_1 \subseteq Q_2$ then $\boldsymbol{D}_{Q_1} \models Q_2$

(2) $\Rightarrow$ (3) If $\boldsymbol{D}_{Q_1} \models Q_2$ then $\exists h : Q_2 \to \boldsymbol{D}_{Q_1}$, hence $\exists h : Q_2 \to Q_1$.

(3) $\Rightarrow$ (1)

Containment Theorem for Boolean CQs

Theorem (Chandra and Merlin, 1977)

The following are equivalent:

(1) Containment holds: $Q_1 \subseteq Q_2$. *(if* $\boldsymbol{D} \vDash Q_1$ *then* $\boldsymbol{D} \vDash Q_2$*)*
(2) $\boldsymbol{D}_{Q_1} \vDash Q_2$.
(3) There exists a homomorphism $h : Q_2 \rightarrow Q_1$

Proof

(1) ⇒ (2) If $\boldsymbol{D}_{Q_1} \vDash Q_1$ and $Q_1 \subseteq Q_2$ then $\boldsymbol{D}_{Q_1} \vDash Q_2$
(2) ⇒ (3) If $\boldsymbol{D}_{Q_1} \vDash Q_2$ then $\exists h : Q_2 \rightarrow \boldsymbol{D}_{Q_1}$, hence $\exists h : Q_2 \rightarrow Q_1$.
(3) ⇒ (1) If $\boldsymbol{D} \vDash Q_1$, then $\exists g : Q_1 \rightarrow \boldsymbol{D}$;

Containment Theorem for Boolean CQs

Theorem (Chandra and Merlin, 1977)

The following are equivalent:

(1) Containment holds: $Q_1 \subseteq Q_2$. *(if* $\boldsymbol{D} \vDash Q_1$ *then* $\boldsymbol{D} \vDash Q_2$*)*

(2) $\boldsymbol{D}_{Q_1} \vDash Q_2$.

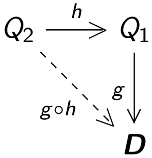
(3) There exists a homomorphism $h : Q_2 \to Q_1$

Proof

(1) ⇒ (2) If $\boldsymbol{D}_{Q_1} \vDash Q_1$ and $Q_1 \subseteq Q_2$ then $\boldsymbol{D}_{Q_1} \vDash Q_2$

(2) ⇒ (3) If $\boldsymbol{D}_{Q_1} \vDash Q_2$ then $\exists h : Q_2 \to \boldsymbol{D}_{Q_1}$, hence $\exists h : Q_2 \to Q_1$.

(3) ⇒ (1) If $\boldsymbol{D} \vDash Q_1$, then $\exists g : Q_1 \to \boldsymbol{D}$; $g \circ h : Q_2 \to \boldsymbol{D}$ implies $\boldsymbol{D} \vDash Q_2$.



Example: Containment

Recall: $\boxed{Q_1 \subseteq Q_2}$ iff $\boxed{\exists h : Q_2 \to Q_1}$

$$
\begin{aligned}
Q_1 =& E(x, y) \wedge E(y, z) \wedge E(z, u) \\
Q_2 =& E(x, y) \wedge E(y, z)
\end{aligned}
$$

$Q_1 \subseteq Q_2$ why???

Example: Containment

Recall: $\boxed{Q_1 \subseteq Q_2}$ iff $\boxed{\exists h : Q_2 \to Q_1}$

$$
\begin{aligned}
Q_1 =& E(x, y) \wedge E(y, z) \wedge E(z, u) \\
Q_2 =& E(x, y) \wedge E(y, z)
\end{aligned}
$$

$Q_1 \subseteq Q_2$ because of the homomorphism $x \mapsto x; y \mapsto y; z \mapsto z$.

Example: Containment

Recall: $\boxed{Q_1 \subseteq Q_2}$ iff $\boxed{\exists h : Q_2 \to Q_1}$

$$
\begin{aligned}
Q_1 &= E(x,y) \wedge E(y,z) \wedge E(z,u) \\
Q_2 &= E(x,y) \wedge E(y,z)
\end{aligned}
$$

$Q_1 \subseteq Q_2$ because of the homomorphism $x \mapsto x; y \mapsto y; z \mapsto z$.

Notice: h must preserve head variables when present.

$$
\begin{aligned}
Q_1'(x) &= E(x,y) \wedge E(y,z) \wedge E(z,u) \\
Q_2'(x) &= E(x,y) \wedge E(y,z)
\end{aligned}
$$

$Q_1' \subseteq Q_2'$

Example: Containment

Recall: $\boxed{Q_1 \subseteq Q_2}$ iff $\boxed{\exists h : Q_2 \to Q_1}$

$$Q_1 = E(x,y) \wedge E(y,z) \wedge E(z,u)$$
$$Q_2 = E(x,y) \wedge E(y,z)$$

$Q_1 \subseteq Q_2$ because of the homomorphism $x \mapsto x; y \mapsto y; z \mapsto z$.

Notice: h must preserve head variables when present.

$$Q_1'(x) = E(x,y) \wedge E(y,z) \wedge E(z,u)$$
$$Q_2'(x) = E(x,y) \wedge E(y,z)$$

$Q_1' \subseteq Q_2'$

$$Q_1''(u) = E(x,y) \wedge E(y,z) \wedge E(z,u)$$
$$Q_2''(y) = E(x,y) \wedge E(y,z)$$

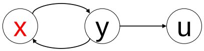
$Q_1'' \not\subseteq Q_2''$

$y \mapsto u; \quad x \mapsto z; \quad z \mapsto$????

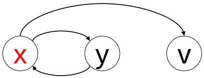
Example: Equivalence

Recall: $\boxed{Q_1 \equiv Q_2}$ iff $\boxed{Q_1 \subseteq Q_2 \text{ and } Q_2 \subseteq Q_1}$

$$Q_1(x) = E(x,y) \wedge E(y,x) \wedge E(y,u)$$



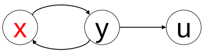
$$Q_2(x) = E(x,y) \wedge E(y,x) \wedge E(x,v)$$



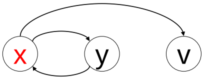
Example: Equivalence

Recall: $\boxed{Q_1 \equiv Q_2}$ iff $\boxed{Q_1 \subseteq Q_2 \text{ and } Q_2 \subseteq Q_1}$

$$Q_1(x) = E(x, y) \wedge E(y, x) \wedge E(y, u)$$



$$Q_2(x) = E(x, y) \wedge E(y, x) \wedge E(x, v)$$

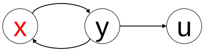


$h_1 : Q_2 \to Q_1$ $h_2 : Q_1 \to Q_2$

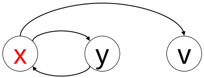
Example: Equivalence

Recall: $\boxed{Q_1 \equiv Q_2}$ iff $\boxed{Q_1 \subseteq Q_2 \text{ and } Q_2 \subseteq Q_1}$

$$Q_1(x) = E(x, y) \wedge E(y, x) \wedge E(y, u)$$



$$Q_2(x) = E(x, y) \wedge E(y, x) \wedge E(x, v)$$



$h_1 : Q_2 \to Q_1$

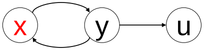
$x \mapsto x,\ y \mapsto y,\ v \mapsto y$

$h_2 : Q_1 \to Q_2$

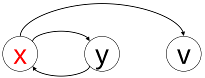
Example: Equivalence

Recall: $\boxed{Q_1 \equiv Q_2}$ iff $\boxed{Q_1 \subseteq Q_2 \text{ and } Q_2 \subseteq Q_1}$

$$Q_1(x) = E(x, y) \wedge E(y, x) \wedge E(y, u)$$



$$Q_2(x) = E(x, y) \wedge E(y, x) \wedge E(x, v)$$



$h_1 : Q_2 \rightarrow Q_1$

$x \mapsto x,\ y \mapsto y,\ v \mapsto y$

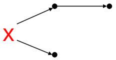
$h_2 : Q_1 \rightarrow Q_2$

$x \mapsto x,\ y \mapsto y,\ u \mapsto x$

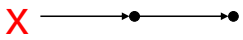
More Examples

Which pairs of queries are contained? Equivalent?

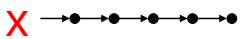
$Q_1(x) = E(x, y) \wedge E(y, z) \wedge E(x, w)$



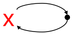
$Q_2(x) = E(x, u) \wedge E(u, v)$



$Q_3(x) = E(x, u_1) \wedge E(u_1, u_2) \wedge \cdots \wedge E(u_4, u_5)$



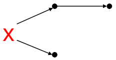
$Q_4(x) = E(x, y) \wedge E(y, x)$



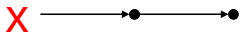
More Examples

Which pairs of queries are contained? Equivalent?

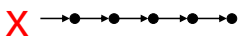
$Q_1(x) = E(x, y) \wedge E(y, z) \wedge E(x, w)$



$Q_2(x) = E(x, u) \wedge E(u, v)$



$Q_3(x) = E(x, u_1) \wedge E(u_1, u_2) \wedge \cdots \wedge E(u_4, u_5)$

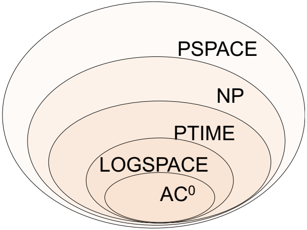


$Q_4(x) = E(x, y) \wedge E(y, x)$

$Q_4 \subseteq Q_3 \subsetneq Q_1 \equiv Q_2$

Complexity

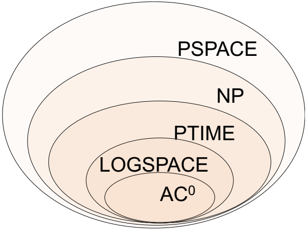
What is the complexity of this problem?
Given Boolean Q_1, Q_2, check if $Q_1 \subseteq Q_2$



Complexity

What is the complexity of this problem?

Given Boolean Q_1, Q_2, check if $Q_1 \subseteq Q_2$



NP-complete!

Proof: same as the model checking for CQs: $Q_1 \subseteq Q_2$ iff $\boldsymbol{D}_{Q_1} \models Q_2$.

Containment Theorem for CQs – Revised

Theorem (Chandra and Merlin, 1977)

Let $Q_1(\boldsymbol{x}_1)$, $Q_2(\boldsymbol{x}_2)$ be CQ's with $|\boldsymbol{x}_1| = |\boldsymbol{x}_2|$. The following are equivalent:

(1) $Q_1 \subseteq Q_2$ $\qquad$ $(\forall \boldsymbol{D}, Q_1(\boldsymbol{D}) \subseteq Q_2(\boldsymbol{D}))$

(2) $\boldsymbol{x}_1^{\boldsymbol{D}_{Q_1}} \in Q_2(\boldsymbol{D}_{Q_1})$ $\qquad$ *($Q_2(\boldsymbol{D}_{Q_1})$ returns the canonical tuple $\boldsymbol{x}_1$)*

(3) $\exists h : Q_2(\boldsymbol{x}_2) \to Q_1(\boldsymbol{x}_1)$. $\qquad$ *(must map $\boldsymbol{x}_2 \mapsto \boldsymbol{x}_1$)*

Containment of UCQs

$Q(\boldsymbol{x}) = Q_1(\boldsymbol{x}) \vee Q_2(\boldsymbol{x}) \vee \cdots$

$Q'(\boldsymbol{x}') = Q'_1(\boldsymbol{x}') \vee Q'_2(\boldsymbol{x}') \vee \cdots$

Containment of UCQs

$Q(\boldsymbol{x}) = Q_1(\boldsymbol{x}) \vee Q_2(\boldsymbol{x}) \vee \cdots$ $\qquad$ $Q'(\boldsymbol{x}') = Q'_1(\boldsymbol{x}') \vee Q'_2(\boldsymbol{x}') \vee \cdots$

Theorem

The following are equivalent:

- $Q \subseteq Q'$ *(Containment holds)*
- $\forall i \exists j, Q_i \subseteq Q'_j$ *(every Q_i is contained in some Q'_j)*

Proof Assume w.l.o.g. Q, Q' are Boolean UCQs.

$\boxed{Q \subseteq Q'}$:

Containment of UCQs

$Q(\boldsymbol{x}) = Q_1(\boldsymbol{x}) \vee Q_2(\boldsymbol{x}) \vee \cdots \qquad Q'(\boldsymbol{x}') = Q'_1(\boldsymbol{x}') \vee Q'_2(\boldsymbol{x}') \vee \cdots$

Theorem

The following are equivalent:

- $Q \subseteq Q'$ *(Containment holds)*
- $\forall i \exists j, Q_i \subseteq Q'_j$ *(every Q_i is contained in some Q'_j)*

Proof Assume w.l.o.g. Q, Q' are Boolean UCQs.

$\boxed{Q \subseteq Q'}$: then for every i: $\boldsymbol{D}_{Q_i} \models Q'$ (because $\boldsymbol{D}_{Q_i} \models Q_i$, since $\boldsymbol{D}_{Q_i} \models Q$)

Containment of UCQs

$Q(\boldsymbol{x}) = Q_1(\boldsymbol{x}) \vee Q_2(\boldsymbol{x}) \vee \cdots$ $\qquad$ $Q'(\boldsymbol{x}') = Q'_1(\boldsymbol{x}') \vee Q'_2(\boldsymbol{x}') \vee \cdots$

Theorem

The following are equivalent:

- $Q \subseteq Q'$ *(Containment holds)*
- $\forall i \exists j, Q_i \subseteq Q'_j$ *(every Q_i is contained in some Q'_j)*

Proof Assume w.l.o.g. Q, Q' are Boolean UCQs.

$\boxed{Q \subseteq Q'}$: then for every i: $\boldsymbol{D}_{Q_i} \vDash Q'$ (because $\boldsymbol{D}_{Q_i} \vDash Q_i$, since $\boldsymbol{D}_{Q_i} \vDash Q$)

Then $\exists j$, $\boldsymbol{D}_{Q_i} \vDash Q'_j$.

Containment of UCQs

$Q(\boldsymbol{x}) = Q_1(\boldsymbol{x}) \vee Q_2(\boldsymbol{x}) \vee \cdots$ $\qquad$ $Q'(\boldsymbol{x}') = Q'_1(\boldsymbol{x}') \vee Q'_2(\boldsymbol{x}') \vee \cdots$

Theorem

The following are equivalent:

- $Q \subseteq Q'$ *(Containment holds)*
- $\forall i \exists j, Q_i \subseteq Q'_j$ *(every Q_i is contained in some Q'_j)*

Proof Assume w.l.o.g. Q, Q' are Boolean UCQs.

$\boxed{Q \subseteq Q'}$: then for every i: $\boldsymbol{D}_{Q_i} \models Q'$ (because $\boldsymbol{D}_{Q_i} \models Q_i$, since $\boldsymbol{D}_{Q_i} \models Q$)

Then $\exists j$, $\boldsymbol{D}_{Q_i} \models Q'_j$. We have shown $\boxed{Q_i \subseteq Q'_j}$

Containment of UCQs

$Q(\boldsymbol{x}) = Q_1(\boldsymbol{x}) \vee Q_2(\boldsymbol{x}) \vee \cdots$ $\quad Q'(\boldsymbol{x}') = Q'_1(\boldsymbol{x}') \vee Q'_2(\boldsymbol{x}') \vee \cdots$

Theorem

The following are equivalent:

- $Q \subseteq Q'$ *(Containment holds)*
- $\forall i \exists j, Q_i \subseteq Q'_j$ *(every Q_i is contained in some Q'_j)*

Proof Assume w.l.o.g. Q, Q' are Boolean UCQs.

$\boxed{Q \subseteq Q'}$: then for every i: $\boldsymbol{D}_{Q_i} \vDash Q'$ (because $\boldsymbol{D}_{Q_i} \vDash Q_i$, since $\boldsymbol{D}_{Q_i} \vDash Q$)

Then $\exists j$, $\boldsymbol{D}_{Q_i} \vDash Q'_j$. We have shown $\boxed{Q_i \subseteq Q'_j}$

$\boxed{\forall i \exists j, Q_i \subseteq Q'_j}$:

Containment of UCQs

$Q(\boldsymbol{x}) = Q_1(\boldsymbol{x}) \vee Q_2(\boldsymbol{x}) \vee \cdots$ $\qquad$ $Q'(\boldsymbol{x}') = Q'_1(\boldsymbol{x}') \vee Q'_2(\boldsymbol{x}') \vee \cdots$

Theorem

The following are equivalent:

- $Q \subseteq Q'$ *(Containment holds)*
- $\forall i \exists j, Q_i \subseteq Q'_j$ *(every Q_i is contained in some Q'_j)*

Proof Assume w.l.o.g. Q, Q' are Boolean UCQs.

$\boxed{Q \subseteq Q'}$: then for every i: $\boldsymbol{D}_{Q_i} \vDash Q'$ (because $\boldsymbol{D}_{Q_i} \vDash Q_i$, since $\boldsymbol{D}_{Q_i} \vDash Q$)

Then $\exists j$, $\boldsymbol{D}_{Q_i} \vDash Q'_j$. We have shown $\boxed{Q_i \subseteq Q'_j}$

$\boxed{\forall i \exists j, Q_i \subseteq Q'_j}$: Assume $\boldsymbol{D} \vDash Q$. Then $\exists i$, $\boldsymbol{D} \vDash Q_i$.

Containment of UCQs

$Q(\boldsymbol{x}) = Q_1(\boldsymbol{x}) \vee Q_2(\boldsymbol{x}) \vee \cdots$ $\qquad$ $Q'(\boldsymbol{x}') = Q'_1(\boldsymbol{x}') \vee Q'_2(\boldsymbol{x}') \vee \cdots$

Theorem

The following are equivalent:

- $Q \subseteq Q'$ *(Containment holds)*
- $\forall i \exists j, Q_i \subseteq Q'_j$ *(every Q_i is contained in some Q'_j)*

Proof Assume w.l.o.g. Q, Q' are Boolean UCQs.

$\boxed{Q \subseteq Q'}$: then for every i: $\boldsymbol{D}_{Q_i} \models Q'$ (because $\boldsymbol{D}_{Q_i} \models Q_i$, since $\boldsymbol{D}_{Q_i} \models Q$)

Then $\exists j$, $\boldsymbol{D}_{Q_i} \models Q'_j$. We have shown $\boxed{Q_i \subseteq Q'_j}$

$\boxed{\forall i \exists j, Q_i \subseteq Q'_j}$: Assume $\boldsymbol{D} \models Q$. Then $\exists i$, $\boldsymbol{D} \models Q_i$.

Let j be such that $Q_i \subseteq Q'_j$. Then $\boldsymbol{D} \models Q'_j$. We have shown $\boxed{\boldsymbol{D} \models Q'}$

Discussion

- Homomorphism criterion for checking containment of CQs/UCQs
- A simple, little known consequence: the same criterion can be adapted to implication of positive CNF:[2]

$$\varphi_1 = \forall x \forall y \forall z (E(x,y) \vee E(y,z))$$
$$\varphi_2 = \forall x \forall y \forall z \forall u (E(x,y) \vee E(y,z) \vee E(z,u))$$

  then $\varphi_1 \Rightarrow \varphi_2$

- The problem given CQs Q_1, Q_2, does $Q_1 \subseteq Q_2$ hold? is NP-complete.

[2] h must go in the opposite direction: $h : \varphi_1 \to \varphi_2$.

Adding Inequalities: $<, \leq, \neq$

Inequalities

Extend CQ with $<, \leq, \neq$. E.g. $Q(x, y, z) = R(x, y) \wedge R(x, z) \wedge y \neq z$.

The extend languages is denoted $\text{CQ}^{<}$, or $\text{CQ}^{\leq,\neq}$, or $\text{CQ}(\leq, \neq)$.

We assume $\text{Dom}(\boldsymbol{D})$ is densely ordered, e.g. $\mathbb{Q}$.

Problems: containment.

Homomorphism is Sufficient

Idea: treat $x < y$ as another relational predicate $R(x, y)$

A homomorphism $h : Q' \to Q$ must map $x < y$ to $h(x) < h(y)$

Homomorphism is Sufficient

Idea: treat $x < y$ as another relational predicate $R(x, y)$

A homomorphism $h : Q' \to Q$ must map $x < y$ to $h(x) < h(y)$

Fact

If there exists a homomorphism $Q' \to Q$ then $Q \subseteq Q'$.

Homomorphism is Sufficient

Idea: treat $x < y$ as another relational predicate $R(x, y)$

A homomorphism $h : Q' \to Q$ must map $x < y$ to $h(x) < h(y)$

Fact

If there exists a homomorphism $Q' \to Q$ then $Q \subseteq Q'$.

The predicate $h(x) < h(y)$ need not appear in Q, just implied by Q. E.g.

$$Q = R(x, y, z) \wedge (x < y) \wedge (y < z) \qquad Q' = R(u, v, w) \wedge (u \leq w)$$

Homomorphism is Sufficient

Idea: treat $x < y$ as another relational predicate $R(x, y)$

A homomorphism $h : Q' \to Q$ must map $x < y$ to $h(x) < h(y)$

Fact

If there exists a homomorphism $Q' \to Q$ then $Q \subseteq Q'$.

The predicate $h(x) < h(y)$ need not appear in Q, just implied by Q. E.g.

$$Q = R(x, y, z) \wedge (x < y) \wedge (y < z) \qquad Q' = R(u, v, w) \wedge (u \leq w)$$

$h : (u, v, w) \mapsto (x, y, z)$ maps $\boxed{u \leq w}$ to $\boxed{x \leq z}$

Homomorphism is Sufficient

Idea: treat $x < y$ as another relational predicate $R(x, y)$

A homomorphism $h : Q' \to Q$ must map $x < y$ to $h(x) < h(y)$

Fact

If there exists a homomorphism $Q' \to Q$ then $Q \subseteq Q'$.

The predicate $h(x) < h(y)$ need not appear in Q, just implied by Q. E.g.

$$Q = R(x, y, z) \wedge (x < y) \wedge (y < z) \qquad Q' = R(u, v, w) \wedge (u \leq w)$$

$h : (u, v, w) \mapsto (x, y, z)$ maps $\boxed{u \leq w}$ to $\boxed{x \leq z}$

$\boxed{x \leq z}$ does not appear in Q, but is implied: $\boxed{Q \models x \leq z}$

Homomorphism is Sufficient

Idea: treat $x < y$ as another relational predicate $R(x, y)$

A homomorphism $h : Q' \to Q$ must map $x < y$ to $h(x) < h(y)$

Fact

If there exists a homomorphism $Q' \to Q$ then $Q \subseteq Q'$.

The predicate $h(x) < h(y)$ need not appear in Q, just implied by Q. E.g.

$$Q = R(x, y, z) \wedge (x < y) \wedge (y < z) \qquad Q' = R(u, v, w) \wedge (u \leq w)$$

$h : (u, v, w) \mapsto (x, y, z)$ maps $\boxed{u \leq w}$ to $\boxed{x \leq z}$

$\boxed{x \leq z}$ does not appear in Q, but is implied: $\boxed{Q \vDash x \leq z}$ Thus, $Q \subseteq Q'$

Homomorphism is Not Necessary

Fact

$h : Q' \to Q$ is a sufficient, but not a necessary condition for $Q \subseteq Q'$.

Homomorphism is Not Necessary

Fact

$h : Q' \to Q$ is a sufficient, but not a necessary condition for $Q \subseteq Q'$.

Example: (Boolean queries):

$$Q = S(x, y) \wedge S(y, z) \wedge (x < z) \qquad Q' = S(u, v) \wedge (u < v)$$

Homomorphism is Not Necessary

Fact

$h : Q' \to Q$ is a sufficient, but not a necessary condition for $Q \subseteq Q'$.

Example: (Boolean queries):

$$Q = S(x,y) \wedge S(y,z) \wedge (x < z) \qquad Q' = S(u,v) \wedge (u < v)$$

$Q \subseteq Q'$. Why?, but there is no homomorphism $Q' \to Q$,

Review: Preorder

Definition

A relation $\leq$ on a set V is called a preorder if:

- It is reflexive: $x \leq x$.
- It is transitive: $x \leq y$, $y \leq z$ implies $x \leq z$.

Write $\boxed{a \equiv b}$ for $a \leq b$ and $b \leq a$.

The preorder is total if $\forall a, b \in V$, either $a \leq b$ or $b \leq a$ or $a \equiv b$.

Extending Q with a Total Preorder $\preceq$

Fix a total preorder $\preceq$ on $\mathrm{Vars}(Q) \cup \mathrm{Const}(Q)$,

$Q_{\preceq}$ denotes the extension of Q with $\preceq$.

Note $Q_{\preceq}$ may be inconsistent, i.e. $\equiv \texttt{False}$.

Extending Q with a Total Preorder $\leq$

Fix a total preorder $\leq$ on $\text{Vars}(Q) \cup \text{Const}(Q)$,

$Q_{\leq}$ denotes the extension of Q with $\leq$.

Note $Q_{\leq}$ may be inconsistent, i.e. $\equiv \texttt{False}$.

Example: $Q = R(x, y, 3) \wedge S(y, z, u, 9) \wedge (u \leq x)$

Extending Q with a Total Preorder $\leq$

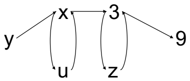
Fix a total preorder $\leq$ on $\text{Vars}(Q) \cup \text{Const}(Q)$,

$Q_{\leq}$ denotes the extension of Q with $\leq$.

Note $Q_{\leq}$ may be inconsistent, i.e. $\equiv \texttt{False}$.

Example: $Q = R(x, y, 3) \wedge S(y, z, u, 9) \wedge (u \leq x)$

$$y \prec x \equiv u \prec 3 \equiv z \prec 9$$



Extending Q with a Total Preorder $\preceq$

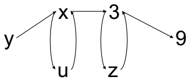
Fix a total preorder $\preceq$ on $\mathrm{Vars}(Q) \cup \mathrm{Const}(Q)$,

$Q_\preceq$ denotes the extension of Q with $\preceq$.

Note $Q_\preceq$ may be inconsistent, i.e. $\equiv$ `False`.

Example: $Q = R(x, y, 3) \wedge S(y, z, u, 9) \wedge (u \leq x)$

$$y \prec x \equiv u \prec 3 \equiv z \prec 9$$



$$Q_\preceq = R(x, y, 3) \wedge S(y, z, u, 9) \wedge (y < x) \wedge (x = u) \wedge (x < 3) \wedge (3 = z) \wedge \cdots$$

A Necessary and Sufficient Condition

Theorem

Let Q, Q' be $CQ^{<,\leq,\neq}$ queries. The following conditions are equivalent:

(1) $Q \subseteq Q'$ $\quad$ *($\forall \boldsymbol{D}$, if $\boldsymbol{D} \vDash Q$ then $\boldsymbol{D} \vDash Q'$)*

(2) For any consistent total preorder $\leq$ on Q, $\exists h : Q' \to Q_{\leq}$.

A Necessary and Sufficient Condition

Theorem

Let Q, Q' be $CQ^{<,\leq,\neq}$ queries. The following conditions are equivalent:

(1) $Q \subseteq Q'$ $\quad$ *($\forall \boldsymbol{D}$, if $\boldsymbol{D} \models Q$ then $\boldsymbol{D} \models Q'$)*

(2) For any consistent total preorder $\leq$ on Q, $\exists h : Q' \to Q_{\leq}$.

Proof:

(2) $\Rightarrow$ (1) If $\boldsymbol{D} \models Q$, then there exists a homomorphism:

$$h_0 : Q \to \boldsymbol{D}$$

This induces a total preorder $\leq$ on Q. Let h be a homomorphism:

$$h : Q' \to Q_{\leq}$$

Their composition is a homomorphism $Q' \to \boldsymbol{D}$, proving $Q'(\boldsymbol{D}) = \texttt{true}$.

A Necessary and Sufficient Condition

Theorem

Let Q, Q' be $CQ^{<,\leq,\neq}$ queries. The following conditions are equivalent:

(1) $Q \subseteq Q'$ $\quad$ *($\forall \boldsymbol{D}$, if $\boldsymbol{D} \models Q$ then $\boldsymbol{D} \models Q'$)*

(2) For any consistent total preorder $\leq$ on Q, $\exists h : Q' \to Q_{\leq}$.

Proof:

(2) $\Rightarrow$ (1) If $\boldsymbol{D} \models Q$, then there exists a homomorphism:

$$h_0 : Q \to \boldsymbol{D}$$

This induces a total preorder $\leq$ on Q. Let h be a homomorphism:

$$h : Q' \to Q_{\leq}$$

Their composition is a homomorphism $Q' \to \boldsymbol{D}$, proving $Q'(\boldsymbol{D}) = \mathtt{true}$.

(1) $\Rightarrow$ (2) follows from $\boldsymbol{D}_{Q_{\leq}} \models Q$, hence $\boldsymbol{D}_{Q_{\leq}} \models Q'$, and $\exists h : Q' \to Q_{\leq}$.

Example

$$Q = S(x,y) \wedge S(y,z) \wedge (x < z)$$

$$Q' = S(u,v) \wedge (u < v)$$

Lets prove that $Q \subseteq Q'$.

Example

$$\boxed{Q = S(x,y) \wedge S(y,z) \wedge (x < z)} \qquad \boxed{Q' = S(u,v) \wedge (u < v)}$$

Lets prove that $\boxed{Q \subseteq Q'}$.

5 consistent total preorders on Q:

$$
\begin{aligned}
Q_1 =& S(x,y) \wedge S(y,z) \wedge (y < x) \wedge (y < z)\\
Q_2 =& S(x,y) \wedge S(y,z) \wedge (x = y) \wedge (y < z)\\
Q_3 =& S(x,y) \wedge S(y,z) \wedge (x < y) \wedge (y < z)\\
Q_4 =& S(x,y) \wedge S(y,z) \wedge (x < y) \wedge (y = z)\\
Q_5 =& S(x,y) \wedge S(y,z) \wedge (x < y) \wedge (z < y)
\end{aligned}
$$

Example

$$\boxed{Q = S(x,y) \wedge S(y,z) \wedge (x < z)} \qquad \boxed{Q' = S(u,v) \wedge (u < v)}$$

Lets prove that $\boxed{Q \subseteq Q'}$.

5 consistent total preorders on Q:

$$
\begin{aligned}
Q_1 =& S(x,y) \wedge S(y,z) \wedge (y < x) \wedge (y < z)\\
Q_2 =& S(x,y) \wedge S(y,z) \wedge (x = y) \wedge (y < z)\\
Q_3 =& S(x,y) \wedge S(y,z) \wedge (x < y) \wedge (y < z)\\
Q_4 =& S(x,y) \wedge S(y,z) \wedge (x < y) \wedge (y = z)\\
Q_5 =& S(x,y) \wedge S(y,z) \wedge (x < y) \wedge (z < y)
\end{aligned}
$$

In each case, either $(u,v) \mapsto (x,y)$ or $(u,v) \mapsto (y,z)$ is a homomorphism.

Complexity

Assume Q, Q' are CQ's that may contain $<, \leq, \neq$.

Theorem

The problem given Q, Q' determine whether $Q \subseteq Q'$ is Π_2^p-complete.

Complexity

Assume Q, Q' are CQ's that may contain $<, \leq, \neq$.

Theorem

The problem given Q, Q' determine whether $Q \subseteq Q'$ is Π_2^p-complete.

Review: there exists Q s.t. Given Q', check $Q \subseteq Q'$ is NP-complete.
Reduction from 3CNF Φ.

$$\Phi = (X \vee \neg Y \vee Z) \wedge (\neg X \vee Y \vee \neg Z) \wedge (\neg X \vee \neg Y \vee Z) \wedge \cdots.$$

$$Q'_\Phi = B(x, z, y) \wedge C(y, x, z) \wedge C(z, x, y) \wedge \cdots$$

$Q = A(0,0,1), A(0,1,0), \ldots$ (all tuples except $A(0,0,0)$; similarly B, C, D)

Complexity

Assume Q, Q' are CQ's that may contain $<, \leq, \neq$.

Theorem

The problem given Q, Q' determine whether $Q \subseteq Q'$ is Π_2^p-complete.

Proof: Membership in Π_2^p follows from:

$Q \subseteq Q'$ iff for all extensions $Q_{\leq}$, there exists a homomorphisms $Q' \to Q_{\leq}$.

Complexity

Assume Q, Q' are CQ's that may contain $<, \leq, \neq$.

Theorem

The problem given Q, Q' determine whether $Q \subseteq Q'$ is Π_2^p-complete.

Proof: Reduction from $\forall\exists 3CNF$: $\boxed{\Psi = \forall X_1 \cdots \forall X_k \exists X_{k+1} \cdots \exists X_n \Phi}$

Complexity

Assume Q, Q' are CQ's that may contain $<, \leq, \neq$.

Theorem

The problem given Q, Q' determine whether $Q \subseteq Q'$ is Π_2^p-complete.

Proof: Reduction from $\forall\exists 3CNF$: $\boxed{\Psi = \forall X_1 \cdots \forall X_k \exists X_{k+1} \cdots \exists X_n \Phi}$

- Q has 4 relations A, B, C, D each with 7 tuples.
- Q'_Φ has one atom/clause; e.g. $(X_i \vee \neg X_j \vee X_k)$ becomes $B(x_i, x_k, x_j)$.
- So far: $\exists X_1 \cdots \exists X_n \Phi$ iff $\exists h : Q'_\Phi \to Q$.

Complexity

Assume Q, Q' are CQ's that may contain $<, \leq, \neq$.

Theorem

The problem given Q, Q' determine whether $Q \subseteq Q'$ is Π_2^p-complete.

Proof: Reduction from $\forall\exists 3CNF$: $\boxed{\Psi = \forall X_1 \cdots \forall X_k \exists X_{k+1} \cdots \exists X_n \Phi}$

- Q has 4 relations A, B, C, D each with 7 tuples.
- Q'_Φ has one atom/clause; e.g. $(X_i \vee \neg X_j \vee X_k)$ becomes $B(x_i, x_k, x_j)$.
- So far: $\exists X_1 \cdots \exists X_n \Phi$ iff $\exists h : Q'_\Phi \to Q$.

For each universal variable X_i, add the following atoms:

- Add $S(0, u_i, v_i) \wedge S(1, v_i, w_i) \wedge (u_i < w_i)$ to Q.
- Add $S(x_i, a_i, b_i) \wedge (a_i < b_i)$ to Q'_Φ.

$\boxed{Q \subseteq Q'_\Phi}$ holds iff both $x_i \mapsto 0$, $x_i \mapsto 1$ lead to a homomorphisms.

Summary

- The big question: what other extensions of CQ can we allow and still be able to decide containment?
- The following have been studied: inequalities, safe negation $\neg$, certain aggregates $\text{sum}, \text{min}, \text{max}, \text{count}$.
- Containment/equivalence for pure CQ/UCQ is very elegant. Extensions add significant difficulties.