What we can show about the Multiset Semantics Query Containment

#### Jerzy Marcinkowski

joint work with

## Piotr Ostropolski-Nalewaja

(and, partially, also with Mateusz Orda)

a very, very, short talk

# What exactly are we talking about



Al art. Prompt: for each query language they invent, researchers study its Query Containment Problem.

#### What exactly are we talking about



Our setting: CQs, UCQs, CQs<sup>≠</sup>, only Boolean queries, multiset semantics.

Al art. Prompt: for each query language they invent, researchers study its Query Containment Problem.

#### What exactly are we talking about



Our setting: CQs, UCQs, CQs<sup>≠</sup>, only Boolean queries, multiset semantics. Input: relational structure. **Output:** 

natural number.

Al art. Prompt: for each query language they invent, researchers study its Query Containment Problem.

# Introducing Query Containment Problem (QCP)

They give us two queries,  $\Phi_s$  and  $\Phi_b$ , and ask whether for each D it holds that  $\Phi_s(D) \subseteq \Phi_b(D)$ 

They give us two queries,  $\Phi_s$  and  $\Phi_b$ , and ask whether for each D it holds that  $\Phi_s(D) \leq \Phi_b(D)$ 

They give us two queries,  $\Phi_s$  and  $\Phi_b$ , and ask whether for each D it holds that  $\Phi_s(D) \leq \Phi_b(D)$ 

If true, denoted as  $\Phi_s \subseteq_{\forall} \Phi_b$ 

They give us two queries,  $\Phi_s$  and  $\Phi_b$ , and ask whether for each D it holds that  $\Phi_s(D) \leq \Phi_b(D)$ 

If true, denoted as  $\Phi_s \subseteq_{\forall} \Phi_b$ 

The central open problem: decidability of QCP(CQ,CQ).

 $\Phi_s = (\exists x, y \ Dog(x), Cat(y))$ 

## $\Phi_s = (\exists x, y \ Dog(x), Cat(y)) \lor (\exists x, y \ Dog(x), Cat(y))$

 $\Phi_{s} = (\exists x, y \ Dog(x), Cat(y)) \lor (\exists x, y \ Dog(x), Cat(y))$  $\Phi_{b} = (\exists x, y \ Dog(x), Dog(y)) \lor (\exists x, y \ Cat(x), Cat(y))$  $Does \ \Phi_{s} \subseteq_{\forall} \Phi_{b} ?$ 

 $\Phi_{s} = (\exists x, y \ Dog(x), Cat(y)) \lor (\exists x, y \ Dog(x), Cat(y))$  $\Phi_{b} = (\exists x, y \ Dog(x), Dog(y)) \lor (\exists x, y \ Cat(x), Cat(y))$  $Does \ \Phi_{s} \subseteq_{\forall} \Phi_{b} ? Yes! \ 2cd \le c^{2} + d^{2}$ 

 $\Phi_{s} = (\exists x, y \ Dog(x), Cat(y)) \lor (\exists x, y \ Dog(x), Cat(y))$  $\Phi_{b} = (\exists x, y \ Dog(x), Dog(y)) \lor (\exists x, y \ Cat(x), Cat(y))$ Does  $\Phi_{s} \subseteq_{\forall} \Phi_{b}$ ? Yes!  $2cd \leq c^{2} + d^{2}$ 

#### **Corollary:**

We have proved that QCP(UCQ,UCQ) is undecidable:

any polynomial can be encoded like this (let us just take one species for each variable). And then use Hilbert's 10th problem.

After 30 years of research...

...a Sophisticated Technique (and beautiful) was invented...

...to prove that  $\Phi_{s} \subseteq_{\forall} \Phi_{b}$  for:

After 30 years of research...

...a Sophisticated Technique (and beautiful) was invented...

...to prove that  $\Phi_{s} \subseteq_{\forall} \Phi_{b}$  for:

 $\Phi_s(D) = \exists x, y, z \ E(x, y), E(y, z), E(z, x)$ 

 $\Phi_b(D) = \exists x, y, y' \ E(x, y), E(x, y')$ 

After 30 years of research...

...a Sophisticated Technique (and beautiful) was invented...

...to prove that  $\Phi_s \subseteq_{\forall} \Phi_b$  for:

 $\Phi_s(D) = \exists x, y, z \ E(x, y), E(y, z), E(z, x)$ 

 $\Phi_b(D) = \exists x, y, y' \ E(x, y), E(x, y')$ 





Jayram, Kolaitis, Vee, 2006: QCP( $CQ^{\neq}$ ,  $CQ^{\neq}$ ) is undecidable.

Jayram, Kolaitis, Vee, 2006: QCP( $CQ^{\neq}$ ,  $CQ^{\neq}$ ) is undecidable.

They need 10<sup>18</sup> inequalities.

Jayram, Kolaitis, Vee, 2006: QCP( $CQ^{\neq}$ ,  $CQ^{\neq}$ ) is undecidable.

They need 10<sup>18</sup> inequalities.

"The containment problem for real CQs with inequalities"

What has been known on the negative side. Jayram, Kolaitis, Vee, 2006: QCP( $CQ^{\neq}$ ,  $CQ^{\neq}$ ) is undecidable. They need 10<sup>18</sup> inequalities.

"The containment problem for real CQs with inequalities"

Why was proving undecidability for the UCQ case easy ? monomials naturally translate to CQs, polynomials naturally translate to UCQ. What has been known on the negative side. Jayram, Kolaitis, Vee, 2006: QCP( $CQ^{\neq}$ ,  $CQ^{\neq}$ ) is undecidable.

They need 10<sup>18</sup> inequalities.

"The containment problem for real CQs with inequalities"

Why was proving undecidability for the UCQ case easy ? monomials naturally translate to CQs, polynomials naturally translate to UCQ.

[JKV06] invent a complicated mechanism... ...to encode polynomials as CQs  $\pi_s$  and  $\pi_b$ . But it only works for some "correct" databases. So they find a way to "reward"  $\pi_b$  if *D* is incorrect.

We introduce a technique called cq-ization. It takes a UCQ  $\Psi$  and returns a CQ  $cq(\Psi)$ .

We introduce a technique called cq-ization.

It takes a UCQ  $\Psi$  and returns a CQ  $cq(\Psi)$ .

The idea is that:  $(cq(\Psi))(D)$  somehow depends on  $\Psi(D)$ .

We introduce a technique called cq-ization.

It takes a UCQ  $\Psi$  and returns a CQ  $cq(\Psi)$ .

The idea is that:  $(cq(\Psi))(D)$  somehow depends on  $\Psi(D)$ .

What it is good for:

We introduce a technique called cq-ization. It takes a UCQ  $\Psi$  and returns a CQ cq( $\Psi$ ).

The idea is that:  $(cq(\Psi))(D)$  somehow depends on  $\Psi(D)$ .

What it is good for:

Theorem 1: QCP(CQ,UCQ) is as decidable as QCP(CQ,CQ).

We introduce a technique called cq-ization. It takes a UCQ  $\Psi$  and returns a <u>CQ</u> cq( $\Psi$ ).

The idea is that:  $(cq(\Psi))(D)$  somehow depends on  $\Psi(D)$ .

#### What it is good for:

Theorem 1: QCP(CQ,UCQ) is as decidable as QCP(CQ,CQ).

**Theorem 2:** For every  $\varepsilon > 0$  this problem is undecidable: given CQs  $\psi_s$  and  $\psi_b$ does  $(1 + \varepsilon)(\psi_s(D)) \le \psi_b(D)$  hold for every *D*?

We introduce a technique called cq-ization. It takes a UCQ  $\Psi$  and returns a CQ cq( $\Psi$ ).

The idea is that:  $(cq(\Psi))(D)$  somehow depends on  $\Psi(D)$ .

#### What it is good for:

Theorem 1: QCP(CQ,UCQ) is as decidable as QCP(CQ,CQ).

Theorem 2: For every  $\varepsilon > 0$  this problem is undecidable: given CQs  $\psi_s$  and  $\psi_b$ does  $(1 + \varepsilon)(\psi_s(D)) \le \psi_b(D)$  hold for every\* D?

We introduce a technique called cq-ization. It takes a UCQ  $\Psi$  and returns a CQ  $cq(\Psi)$ .

The idea is that:  $(cq(\Psi))(D)$  somehow depends on  $\Psi(D)$ .

What it is good for:

Theorem 1: QCP(CQ,UCQ) is as decidable as QCP(CQ,CQ).

**Theorem 2:** For every  $\varepsilon > 0$  this problem is undecidable: given CQs  $\psi_s$  and  $\psi_b$ does  $(1 + \varepsilon)(\psi_s(D)) \le \psi_b(D)$  hold for every\* *D*?

We introduce a technique called cq-ization. It takes a UCQ  $\Psi$  and returns a CQ  $cq(\Psi)$ .

The idea is that:  $(cq(\Psi))(D)$  somehow depends on  $\Psi(D)$ .

What it is good for:

Theorem 1: QCP(CQ,UCQ) is as decidable as QCP(CQ,CQ).

**Theorem 2:** For every  $\varepsilon > 0$  This problem is undecidable: given CQs  $\psi_s$  and  $\psi_b$ ; does  $2 \cdot (\psi_s(D)) \le \psi_b(D)$  hold for every<sup>\*</sup> D?

We introduce a technique called cq-ization. It takes a UCQ  $\Psi$  and returns a CQ  $cq(\Psi)$ .

The idea is that:  $(cq(\Psi))(D)$  somehow depends on  $\Psi(D)$ .

What it is good for:

Theorem 1: QCP(CQ,UCQ) is as decidable as QCP(CQ,CQ).

**Theorem 2:** For every  $\varepsilon > 0$  This problem is undecidable: given CQs  $\psi_s$  and  $\psi_b$ ; does  $2 \cdot (\psi_s(D)) \le \psi_b(D)$  hold for every\* *D*?

Corollary: QCP(UCQ,CQ) is undecidable<sup>\*</sup>: It is undecidable, for given CQs  $\psi_s$  and  $\psi_b$ , whether  $(\psi_s \lor \psi_s)(D) \le \psi_b(D)$  holds for every<sup>\*</sup> D.

<sup>(\*)</sup> Fine print: only non-trivial structures D, which satisfy  $\sigma \neq \varphi$ , are allowed.

[JKV06]: QCP(CQ<sup> $\neq$ </sup>, CQ<sup> $\neq$ </sup>) is undecidable. (10<sup>18</sup> inequalities)

**Theorem 2:** It is undecidable, for given CQs  $\psi_s$  and  $\psi_b$  whether  $2 \cdot (\psi_s(D)) \leq \psi_b(D)$  holds for each non-trivial D.

[JKV06]: QCP(CQ<sup> $\neq$ </sup>, CQ<sup> $\neq$ </sup>) is undecidable. (10<sup>18</sup> inequalities)

**Theorem 2:** It is undecidable, for given CQs  $\psi_s$  and  $\psi_b$  whether  $2 \cdot (\psi_s(D)) \leq \psi_b(D)$  holds for each non-trivial D.

Plan: we will show how so solve Theorem 2 using [JKV06].

[JKV06]: QCP(CQ<sup> $\neq$ </sup>, CQ<sup> $\neq$ </sup>) is undecidable. (2 inequalities)

**Theorem 2:** It is undecidable, for given CQs  $\psi_s$  and  $\psi_b$  whether  $2 \cdot (\psi_s(D)) \leq \psi_b(D)$  holds for each non-trivial D.

Plan: we will show how so solve Theorem 2 using [JKV06].

[JKV06]: QCP(CQ<sup> $\neq$ </sup>, CQ<sup> $\neq$ </sup>) is undecidable. (2 inequalities)

**Theorem 2:** It is undecidable, for given CQs  $\psi_s$  and  $\psi_b$ whether  $2 \cdot (\psi_s(D)) \leq \psi_b(D)$  holds for each non-trivial D.

Plan: we will show how so solve Theorem 2 using [JKV06].

For given CQs  $\psi_s$  and  $\psi_b$  we will construct CQs  $\phi_s$  and  $\phi_b$  such that the following two conditions are equivalent:

•  $2 \cdot \psi_s(D) \leq \psi_b(D)$  holds for each D such that  $D \models \sigma \neq Q$ 

• 
$$(\sigma \neq \varrho) \land \phi_s \subseteq_\forall x \neq x' \land \phi_b$$

For given CQs  $\psi_s$  and  $\psi_b$  we will construct CQs  $\phi_s$  and  $\phi_b$  such that the following two conditions are equivalent:

- $2 \cdot \psi_s(D) \leq \psi_b(D)$  holds for each D such that  $D \models \sigma \neq Q$
- $(\sigma \neq \mathfrak{P}) \land \phi_s \subseteq_\forall x \neq x' \land \phi_b$

For given CQs  $\psi_s$  and  $\psi_b$  we will construct CQs  $\phi_s$  and  $\phi_b$  such that the following two conditions are equivalent:

2 · ψ<sub>s</sub>(D) ≤ ψ<sub>b</sub>(D) holds for each D such that D ⊨ σ ≠ ♀
(σ ≠ ♀) ∧ φ<sub>s</sub> ⊆<sub>∀</sub> x ≠ x' ∧ φ<sub>b</sub>

**Define:**  $\gamma = P(x) \land P(x') \land P(\mathfrak{P}) \land P(\mathfrak{F})$  (*P* is new)

For given CQs  $\psi_s$  and  $\psi_b$  we must construct CQs  $\phi_s$  and  $\phi_b$  such that The following two conditions are equivalent:

- $2 \cdot \psi_s(D) \leq \psi_b(D)$  holds for each D such that  $D \models \sigma \neq Q$
- $(\sigma \neq \varrho) \land \gamma \land \psi_s \subseteq_\forall x \neq x' \land \gamma \land \psi_b$

**Define:**  $\gamma = P(x) \land P(x') \land P(\mathfrak{P}) \land P(\mathfrak{F})$  (*P* is new) and:  $\phi_s = \psi_s \land \gamma$  and  $\phi_b = \psi_b \land \gamma$ 

For given CQs  $\psi_s$  and  $\psi_b$  we must construct CQs  $\phi_s$  and  $\phi_b$  such that The following two conditions are equivalent:

- $2 \cdot \psi_s(D) \leq \psi_b(D)$  holds for each D such that  $D \models \sigma \neq Q$
- $(\sigma \neq \varrho) \land \gamma \land \psi_s \subseteq_\forall x \neq x' \land \gamma \land \psi_b$

**Define:**  $\gamma = P(x) \land P(x') \land P(\mathfrak{P}) \land P(\mathfrak{F})$  (*P* is new)

and:  $\phi_s = \psi_s \land \gamma$  and  $\phi_b = \psi_b \land \gamma$ 

Idea: using inequality "we can multiply by 2":  $((\sigma \neq \mathfrak{P}) \land \gamma)(D) \leq 2 \cdot ((x \neq x' \land \gamma)(D))$ 

For given CQs  $\psi_s$  and  $\psi_b$  we must construct CQs  $\phi_s$  and  $\phi_b$  such that The following two conditions are equivalent:

- $2 \cdot \psi_s(D) \leq \psi_b(D)$  holds for each D such that  $D \models \sigma \neq Q$
- $(\sigma \neq \varrho) \land \gamma \land \psi_s \subseteq_\forall x \neq x' \land \gamma \land \psi_b$

**Define:**  $\gamma = P(x) \land P(x') \land P(\mathfrak{P}) \land P(\mathfrak{S})$  (*P* is new)

and:  $\phi_s = \psi_s \land \gamma$  and  $\phi_b = \psi_b \land \gamma$ 

Idea: using inequality "we can multiply by 2":  $((\sigma \neq \varphi) \land \gamma)(D) \leq 2 \cdot ((x \neq x' \land \gamma)(D))$ 

Can we multiply by 2 without  $\neq$  ?

For given CQs  $\psi_s$  and  $\psi_b$  we must construct CQs  $\phi_s$  and  $\phi_b$  such that The following two conditions are equivalent:

- $2 \cdot \psi_s(D) \leq \psi_b(D)$  holds for each D such that  $D \models \sigma \neq Q$
- $(\sigma \neq \varrho) \land \gamma \land \psi_s \subseteq_\forall x \neq x' \land \gamma \land \psi_b$
- **Define:**  $\gamma = P(x) \land P(x') \land P(\mathfrak{P}) \land P(\mathfrak{F})$  (*P* is new)
- and:  $\phi_s = \psi_s \land \gamma$  and  $\phi_b = \psi_b \land \gamma$
- Idea: using inequality "we can multiply by 2":  $((\sigma \neq \varphi) \land \gamma)(D) \leq 2 \cdot ((x \neq x' \land \gamma)(D))$

Can we multiply by 2 without  $\neq$ ? No, we (provably) can't. If we could, QCP(CQ,CQ) would be undecidable. Conclusion.

So... what is the probability that the problem is decidable?



So... what is the probability that the problem is decidable?



So... what is the probability that the problem is decidable?

Phokion is a Bayesian:

More work was done trying to prove decidability. We failed so it is more likely to be undecidable.





So... what is the probability that the problem is decidable?

Phokion is a Bayesian:

More work was done trying to prove decidability. We failed so it is more likely to be undecidable.



Albert is a frequentist: It is either one or zero. We just do not know yet.





Phokion is a Bayesian:

More work was done trying to prove decidability. We failed so it is more likely to be undecidable.



Albert is a frequentist: It is either one or zero.

We just do not know yet.

(Which is absolutely correct and absolutely useless)