Open Problems in Consistent Query Answering

Open Problems in Database Theory, ICDT 2017

Benny Kimelfeld   Paris Koutris

Database Repairs [ABC99]

- **Inconsistent database**: database D, violates set Σ of constraints

Database Repairs [ABC99]

- **Inconsistent database**: database D, violates set Σ of constraints
- **Repair** of D: database D' that
  - satisfies Σ
  - obtained from D via a *minimal* set of tuple deletions/additions
    - (*minimal* w.r.t. set inclusion)

Database Repairs [ABC99]

- **Inconsistent database**: database D, violates set Σ of constraints
- **Repair** of D: database D' that
  - satisfies Σ
  - obtained from D via a *minimal* set of tuple deletions/additions
    - (*minimal* w.r.t. set inclusion)
- Special case: Σ is a set of *primary-key constraints*
  - Then, a repair selects one tuple for each key value

Consistent Query Answering

Consistent Query Answering (CQA)

Let Σ be a finite set of constraints, and Q a boolean query. **CQA** is the following decision problem:

Given an inconsistent D, is $Q(D')$ true for every repair D'?

Consistent Query Answering

Consistent Query Answering (CQA)

Let Σ be a finite set of constraints, and Q a boolean query. **CQA** is the following decision problem:

Given an inconsistent D, is $Q(D')$ true for every repair D'?

- For common classes of Σ (e.g., GAV-tgds, egds) and Q (e.g., UCQs), CQA is in coNP

Consistent Query Answering

Consistent Query Answering (CQA)

Let Σ be a finite set of constraints, and Q a boolean query. **CQA** is the following decision problem:

Given an inconsistent D, is $Q(D')$ true for every repair D'?

- For common classes of Σ (e.g., GAV-tgds, egds) and Q (e.g., UCQs), CQA is in coNP
- **Dichotomy** for such classes of Σ and Q refers to the conjecture that *for every Σ and Q, CQA is either in PTime or coNP-complete*
  - Ideally, we would also like to have an algorithm that determines the complexity of CQA for given Σ and Q

Consistent Query Answering

Consistent Query Answering (CQA)

Let Σ be a finite set of constraints, and Q a boolean query. **CQA** is the following decision problem:

Given an inconsistent D, is $Q(D')$ true for every repair D'?

- For common classes of Σ (e.g., GAV-tgds, egds) and Q (e.g., UCQs), CQA is in coNP
- **Dichotomy** for such classes of Σ and Q refers to the conjecture that *for every Σ and Q, CQA is either in PTime or coNP-complete*
  - Ideally, we would also like to have an algorithm that determines the complexity of CQA for given Σ and Q
- **FO rewritability**: Can CQA for Q and Σ be expressed as an ordinary query Q' in FO (hence, PTime)?

History on CQA Research for Primary-Key Constraints

- **2005**: First attempt to establish a dichotomy for *acyclic simple CQs* [FM05]
  - *simple* = no self joins

History on CQA Research for Primary-Key Constraints

- **2005**: First attempt to establish a dichotomy for *acyclic simple CQs* [FM05]
  - *simple* = no self joins
- **2010**: Dichotomy in FO rewritability for *acyclic simple CQs* [Wij10]

History on CQA Research for Primary-Key Constraints

- **2005**: First attempt to establish a dichotomy for *acyclic simple CQs* [FM05]
  - *simple* = no self joins
- **2010**: Dichotomy in FO rewritability for *acyclic simple CQs* [Wij10]
- **2012**: Dichotomy (PTime vs coNP-complete) for *simple CQs with two atoms* [KP12]

History on CQA Research for Primary-Key Constraints

- **2005**: First attempt to establish a dichotomy for *acyclic simple CQs* [FM05]
  - *simple* = no self joins
- **2010**: Dichotomy in FO rewritability for *acyclic simple CQs* [Wij10]
- **2012**: Dichotomy (PTime vs coNP-complete) for *simple CQs with two atoms* [KP12]
- **2014**: Dichotomy for *simple CQs over binary relations* [KS14]

History on CQA Research for Primary-Key Constraints

- **2005**: First attempt to establish a dichotomy for *acyclic simple CQs* [FM05]
  - *simple* = no self joins
- **2010**: Dichotomy in FO rewritability for *acyclic simple CQs* [Wij10]
- **2012**: Dichotomy (PTime vs coNP-complete) for *simple CQs with two atoms* [KP12]
- **2014**: Dichotomy for *simple CQs over binary relations* [KS14]
- **2015**: Dichotomy for *all simple CQs* [KW15]
  - In fact, a more refined classification: FO, PTime$\setminus$FO, coNP-complete

Open Problems

Are there dichotomies/trichotomies for broader classes of Q and Σ?

Σ \ Q	simple CQs			
key constraints	✓			

Open Problems

Are there dichotomies/trichotomies for broader classes of Q and Σ?

Σ \ Q	simple CQs	CQs		
key constraints	✓			

Open Problems

Are there dichotomies/trichotomies for broader classes of Q and Σ?

Σ \ Q	simple CQs	CQs	$\text{CQ}^{\neg}$	
key constraints	✓			

Open Problems

Are there dichotomies/trichotomies for broader classes of Q and Σ?

Σ \ Q	simple CQs	CQs	$\text{CQ}^{\neg}$	UCQs
key constraints	✓			

Open Problems

Are there dichotomies/trichotomies for broader classes of Q and Σ?

Σ \ Q	simple CQs	CQs	$\mathrm{CQ}^{\neg}$	UCQs
key constraints	✓			
FD				

Open Problems

Are there dichotomies/trichotomies for broader classes of Q and Σ?

Σ \ Q	simple CQs	CQs	$\mathsf{CQ}^{\neg}$	UCQs
key constraints	✓			
FD				
EGD				

Open Problems

Are there dichotomies/trichotomies for broader classes of Q and Σ?

Σ \ Q	simple CQs	CQs	$\text{CQ}^{\neg}$	UCQs
key constraints	✓			
FD				
EGD				
GAV				

Open Problems

Are there dichotomies/trichotomies for broader classes of Q and Σ?

Σ \ Q	simple CQs	CQs	$\mathsf{CQ}^{\neg}$	UCQs
key constraints	✓			
FD				
EGD				
GAV				*

CSP vs. CQA

- The *Constraint Satisfaction Problem* (**CSP**) dichotomy conjecture: Every CSP instance (set of allowed constraints) is either in PTime or NP-complete
- Posed by Feder and Vardi in 1993 [FV93], generalizes Schaefer's 1978 dichotomy theorem [Sch78]
- Officially, **still open!**

CSP vs. CQA

- The *Constraint Satisfaction Problem* (**CSP**) dichotomy conjecture: Every CSP instance (set of allowed constraints) is either in PTime or NP-complete
- Posed by Feder and Vardi in 1993 [FV93], generalizes Schaefer's 1978 dichotomy theorem [Sch78]
- Officially, **still open!**

Theorem [Fon13]

Dichotomy for CQA with GAV constraints and UCQs
$\Rightarrow$ Dichotomy for CSP

CSP vs. CQA

- The *Constraint Satisfaction Problem* (**CSP**) dichotomy conjecture: Every CSP instance (set of allowed constraints) is either in PTime or NP-complete
- Posed by Feder and Vardi in 1993 [FV93], generalizes Schaefer's 1978 dichotomy theorem [Sch78]
- Officially, **still open!**

Theorem [Fon13]

Dichotomy for CQA with GAV constraints and UCQs
$\Rightarrow$ Dichotomy for CSP

- Twist: very recently announced that CSP conjecture has been **proved valid** by Rafiey, Kinne and Feder [RKF17]
- Does it imply CQA dichotomy? Unknown [Fon13]

Counting Repairs

Let Σ be a finite set of constraints.

- how many repairs does D have?
- how many repairs of D satisfy a given query Q?

Counting Repairs

Let Σ be a finite set of constraints.

- how many repairs does D have?
- how many repairs of D satisfy a given query Q?

Closely connected to query evaluation over BID probabilistic databases:

- set the probability of a tuple in a block of size k to $1/k$
- **difference**: in BIDs, tuples in the same block (a) can have non-uniform probabilities (b) their probabilities may not sum to 1

Counting Repairs: Dichotomies

Dichotomy here refers to that *for every Σ and Q, counting is either in PTime (FP) or #P-complete*

Known:

- a dichotomy for counting the number of repairs for FDs [LK17]
- a dichotomy for counting repairs that satisfy CQs with primary keys [MW14]

Counting Repairs: Dichotomies

Dichotomy here refers to that *for every Σ and Q, counting is either in PTime (FP) or #P-complete*

Known:

- a dichotomy for counting the number of repairs for FDs [LK17]
- a dichotomy for counting repairs that satisfy CQs with primary keys [MW14]

Unknown: Everything else: e.g., CQs and FDs

Questions?

References I

- Marcelo Arenas, Leopoldo E. Bertossi, and Jan Chomicki, *Consistent query answers in inconsistent databases*, PODS, ACM Press, 1999, pp. 68–79.
- Ariel Fuxman and Renée J. Miller, *First-order query rewriting for inconsistent databases*, ICDT, 2005, pp. 337–351.
- Gaëlle Fontaine, *Why is it hard to obtain a dichotomy for consistent query answering?*, LICS, 2013, pp. 550–559.
- Tomás Feder and Moshe Y. Vardi, *Monotone monadic SNP and constraint satisfaction*, STOC, ACM, 1993, pp. 612–622.
- Phokion G. Kolaitis and Enela Pema, *A dichotomy in the complexity of consistent query answering for queries with two atoms*, Inf. Process. Lett. **112** (2012), no. 3, 77–85.

References II

- Paraschos Koutris and Dan Suciu, *A dichotomy on the complexity of consistent query answering for atoms with simple keys*, ICDT, 2014, pp. 165–176.
- Paraschos Koutris and Jef Wijsen, *The data complexity of consistent query answering for self-join-free conjunctive queries under primary key constraints*, PODS, ACM, 2015, pp. 17–29.
- Ester Livshits and Benny Kimelfeld, *Counting and enumerating (preferred) database repairs*, To appear in PODS, 2017.
- Dany Maslowski and Jef Wijsen, *Counting database repairs that satisfy conjunctive queries with self-joins*, ICDT, 2014, pp. 155–164.
- Arash Rafiey, Jeff Kinne, and Tomás Feder, *Dichotomy for digraph homomorphism problems*, CoRR **abs/1701.02409** (2017).

References III

Thomas J. Schaefer, *The complexity of satisfiability problems*, STOC, ACM, 1978, pp. 216–226.

Jef Wijsen, *On the first-order expressibility of computing certain answers to conjunctive queries over uncertain databases*, PODS, 2010, pp. 179–190.