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 $\Sigma$ of constraints
Database Repairs [ABC99]

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

Let $\Sigma$ 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 $\Sigma$ 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 $\Sigma$ (e.g., GAV-tgds, egds) and $Q$ (e.g., UCQs), CQA is in coNP
Consistent Query Answering

**Consistent Query Answering (CQA)**

Let $\Sigma$ 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 $\Sigma$ (e.g., GAV-tgds, egds) and $Q$ (e.g., UCQs), CQA is in coNP
- **Dichotomy** for such classes of $\Sigma$ and $Q$ refers to the conjecture that for every $\Sigma$ 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 $\Sigma$ and $Q$
Consistent Query Answering

Consistent Query Answering (CQA)

Let $\Sigma$ 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 $\Sigma$ (e.g., GAV-tgds, egds) and $Q$ (e.g., UCQs), CQA is in coNP
- **Dichotomy** for such classes of $\Sigma$ and $Q$ refers to the conjecture that for every $\Sigma$ 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 $\Sigma$ and $Q$
- **FO rewritability**: Can CQA for $Q$ and $\Sigma$ 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\FO, coNP-complete
Open Problems

Are there dichotomies/trichotomies for broader classes of $Q$ and $\Sigma$?

<table>
<thead>
<tr>
<th>$\Sigma$</th>
<th>$Q$</th>
<th>simple CQs</th>
<th>key constraints</th>
<th>✓</th>
</tr>
</thead>
</table>
Are there dichotomies/trichotomies for broader classes of $Q$ and $Σ$?

<table>
<thead>
<tr>
<th>$Σ$</th>
<th>$Q$</th>
<th>key constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>simple CQs</th>
<th>CQs</th>
</tr>
</thead>
</table>

Are there dichotomies/trichotomies for broader classes of $Q$ and $\Sigma$?

<table>
<thead>
<tr>
<th></th>
<th>$\Sigma$</th>
<th>$Q$</th>
<th>simple CQs</th>
<th>CQs</th>
<th>CQ$\neg$</th>
</tr>
</thead>
<tbody>
<tr>
<td>key constraints</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Are there dichotomies/trichotomies for broader classes of $Q$ and $\Sigma$?

<table>
<thead>
<tr>
<th>$\Sigma$</th>
<th>$Q$</th>
<th>simple CQs</th>
<th>CQs</th>
<th>$\text{CQ}^-$</th>
<th>UCQs</th>
</tr>
</thead>
<tbody>
<tr>
<td>key constraints</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Open Problems

Are there dichotomies/trichotomies for broader classes of $Q$ and $\Sigma$?

<table>
<thead>
<tr>
<th>$\Sigma$</th>
<th>$Q$</th>
<th>simple CQs</th>
<th>CQs</th>
<th>CQ$^-$</th>
<th>UCQs</th>
</tr>
</thead>
<tbody>
<tr>
<td>key constraints</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Benny Kimelfeld, Paris Koutris
ICDT 2017 Open Problems
Open Problems

Are there dichotomies/trichotomies for broader classes of $Q$ and $\Sigma$?

<table>
<thead>
<tr>
<th>$\Sigma$ \ $Q$</th>
<th>simple CQs</th>
<th>CQs</th>
<th>$CQ^-$</th>
<th>UCQs</th>
</tr>
</thead>
<tbody>
<tr>
<td>key constraints</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Are there dichotomies/trichotomies for broader classes of $Q$ and $\Sigma$?

<table>
<thead>
<tr>
<th>$\Sigma$</th>
<th>$Q$</th>
<th>simple CQs</th>
<th>CQs</th>
<th>$CQ^\neg$</th>
<th>UCQs</th>
</tr>
</thead>
<tbody>
<tr>
<td>key constraints</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Open Problems

Are there dichotomies/trichotomies for broader classes of $Q$ and $\Sigma$?

<table>
<thead>
<tr>
<th>$\Sigma$</th>
<th>$Q$</th>
<th>simple CQs</th>
<th>CQs</th>
<th>CQ$^-$</th>
<th>UCQs</th>
</tr>
</thead>
<tbody>
<tr>
<td>key constraints</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
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!

Twist: very recently announced that CSP conjecture has been proved valid by Rafiey, Kinne and Feder [RKF17]. Does it imply CQA dichotomy? Unknown [Fon13].
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
⇒ 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 \( \Sigma \) 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 \( \frac{1}{k} \)
- difference: in BIDs, tuples in the same block (a) can have non-uniform probabilities (b) their probabilities may not sum to 1
Counting Repairs

Let $\Sigma$ 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 $\Sigma$ and $Q$, counting is either in $\text{PTime (FP)}$ or $\text{#P-complete}$

**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 $\Sigma$ 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


References II


