Open-World Probabilistic Databases

Guy Van den Broeck

FLAIRS
May 23, 2017
Overview

1. **Why** probabilistic databases?

2. **How** probabilistic query evaluation?

3. **Why** open world?

4. **How** open-world query evaluation?

5. **What** is the broader picture?
Why probabilistic databases?
What we’d like to do…

Has anyone published a paper with both Erdos and Einstein

About 82,400 results (0.73 seconds)

Erdős number - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Erdős_number
He published more papers during his lifetime (at least 1,525) than any other...
Anybody else's Erdős number is k + 1 where k is the lowest Erdős number of any coauthor...

Albert Einstein and Sheldon Lee Glashow have an Erdős number of 2.

Erdős–Bacon number - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Erdős–Bacon_number
This article possibly contains previously unpublished synthesis of published...
Her paper gives her an Erdős number of 4, and a Bacon number of 2, both of...
Google Knowledge Graph

> 570 million entities
> 18 billion tuples
Probabilistic Databases

- Tuple-independent probabilistic database

<table>
<thead>
<tr>
<th>Scientist</th>
<th>x</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erdos</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Einstein</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Pauli</td>
<td>0.6</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coauthor</th>
<th>x</th>
<th>y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erdos</td>
<td>Renyi</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Einstein</td>
<td>Pauli</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Obama</td>
<td>Erdos</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

- Learned from the web, large text corpora, ontologies, etc., using **statistical** machine learning.

[Suciu’11]
Information Extraction is Noisy!

PhD Students Luc De Raedt

- Laura-Andrea Antanas (co-promotor Tinne Tuytelaars)
- Dries Van Daele (co-promotor Kathleen Marchal)
- Thanh Le Van (co-promotor Kathleen Marchal)
- Bogdan Moldovan
- Davide Nitti (co-promotor Tinne De Laet)
- José Antonio Oramas Mogrovejo (key supervisor Tinne Tuytelaars)
- Francesco Orsini (co-supervisor Paol Frasconi)
- Sergey Paramonov
- Joris Renkens
- Mathias Verbeke (with Bettina Berendt)
- Jonas Vlasselaer

Coauthor

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luc</td>
<td>Laura</td>
<td>0.7</td>
</tr>
<tr>
<td>Luc</td>
<td>Hendrik</td>
<td>0.6</td>
</tr>
<tr>
<td>Luc</td>
<td>Kathleen</td>
<td>0.3</td>
</tr>
<tr>
<td>Luc</td>
<td>Paol</td>
<td>0.3</td>
</tr>
<tr>
<td>Luc</td>
<td>Paolo</td>
<td>0.1</td>
</tr>
</tbody>
</table>
What we’d like to do…

∃x Coauthor(Einstein,x) ∧ Coauthor(Erdos,x)
Einstein is in the Knowledge Graph
Erdős is in the Knowledge Graph
This guy is in the Knowledge Graph

Ernst Straus

Ernst G. Straus
Mathematician

Ernst Gabor Straus was a German-American mathematician who helped found the theories of Euclidean Ramsey theory and of the arithmetic properties of analytic functions. Wikipedia

Born: February 25, 1922, Munich, Germany
Died: July 12, 1983, Los Angeles, CA
Residence: United States of America

Straus biography - University of St Andrews

Ernst Straus's mother was Rahel Goitein who had the distinction of being one of the first women medical students officially studying at a German university.

[Search results for Ernst Straus]

... and he published with both Einstein and Erdos!
Has anyone published a paper with both Erdos and Einstein

1. Fuse uncertain information from web
   ⇒ Embrace probability!

2. Cannot come from labeled data
   ⇒ Embrace query eval!

Ernst Straus

Barack Obama, …

Justin Bieber, …
WHEN I asked Alexa earlier this week who was playing in the Super Bowl, she responded, somewhat monotonously, “Super Bowl 49’s winner is New England Patriots.”

“Come on, that’s last year’s Super Bowl,” I said. “Even I can do better than that.”

At the time, I was actually alone in my living room. I was talking to the virtual companion inside Amazon’s wireless speaker, Echo, which was released last June. Known as Alexa, she has gained raves from Silicon Valley’s tech-obsessed digerati and has become one of the newest members of the virtual assistants club.

All the so-called Frightful Five tech
How probabilistic query evaluation?
Tuple-Independent Probabilistic DB

Probabilistic database D:

Possible worlds semantics:

\[(1-p_1)(1-p_2)(1-p_3)\]
Probabilistic Query Evaluation

\[ Q = \exists x \exists y \text{Scientist}(x) \land \text{Coauthor}(x,y) \]

\[ P(Q) = 1 - \left\{ 1 - p_1 \left[ 1 - (1-q_1)(1-q_2) \right] \right\} \times \left\{ 1 - p_2 \left[ 1 - (1-q_3)(1-q_4)(1-q_5) \right] \right\} \]

### Scientist

<table>
<thead>
<tr>
<th>x</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>p_1</td>
</tr>
<tr>
<td>B</td>
<td>p_2</td>
</tr>
<tr>
<td>C</td>
<td>p_3</td>
</tr>
</tbody>
</table>

### Coauthor

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>D</td>
<td>q_1</td>
</tr>
<tr>
<td>A</td>
<td>E</td>
<td>q_2</td>
</tr>
<tr>
<td>B</td>
<td>F</td>
<td>q_3</td>
</tr>
<tr>
<td>B</td>
<td>G</td>
<td>q_4</td>
</tr>
<tr>
<td>B</td>
<td>H</td>
<td>q_5</td>
</tr>
</tbody>
</table>
Lifted Inference Rules

Preprocess \( Q \) (omitted),
Then apply rules (some have preconditions)

\[
P(\neg Q) = 1 - P(Q)
\]

Negation

\[
\begin{align*}
P(Q_1 \land Q_2) &= P(Q_1) P(Q_2) \\
P(Q_1 \lor Q_2) &= 1 - (1 - P(Q_1)) (1 - P(Q_2))
\end{align*}
\]

Decomposable \( \land, \lor \)

\[
\begin{align*}
P(\exists z \ Q) &= 1 - \prod_{A \in \text{Domain}} (1 - P(Q[A/z])) \\
P(\forall z \ Q) &= \prod_{A \in \text{Domain}} P(Q[A/z])
\end{align*}
\]

Decomposable \( \exists, \forall \)

\[
\begin{align*}
P(Q_1 \land Q_2) &= P(Q_1) + P(Q_2) - P(Q_1 \lor Q_2) \\
P(Q_1 \lor Q_2) &= P(Q_1) + P(Q_2) - P(Q_1 \land Q_2)
\end{align*}
\]

Inclusion/exclusion

\[\text{[Suciu’11]}\]
Closed-World Lifted Query Eval

\[ Q = \exists x \exists y \text{Scientist}(x) \land \text{Coauthor}(x,y) \]

\[ P(Q) = 1 - \prod_{A \in \text{Domain}} (1 - P(\text{Scientist}(A) \land \exists y \text{Coauthor}(A,y))) \]

= 1 - (1 - P(\text{Scientist}(A) \land \exists y \text{Coauthor}(A,y)))
\times (1 - P(\text{Scientist}(B) \land \exists y \text{Coauthor}(B,y)))
\times (1 - P(\text{Scientist}(C) \land \exists y \text{Coauthor}(C,y)))
\times (1 - P(\text{Scientist}(D) \land \exists y \text{Coauthor}(D,y)))
\times (1 - P(\text{Scientist}(E) \land \exists y \text{Coauthor}(E,y)))
\times (1 - P(\text{Scientist}(F) \land \exists y \text{Coauthor}(F,y)))
\ldots

Complexity PTIME
Limitations

\[ H_0 = \forall x \forall y \text{ Smoker}(x) \lor \text{ Friend}(x,y) \lor \text{ Jogger}(y) \]

The decomposable \( \forall \)-rule: \[ P(\forall z \ Q) = \prod_{A \in \text{Domain}} P(Q[A/z]) \]

… does not apply: 

\[ H_0[Alice/x] \text{ and } H_0[Bob/x] \text{ are dependent:} \]

\[ \forall y \ (\text{Smoker}(Alice) \lor \text{Friend}(Alice,y) \lor \text{ Jogger}(y)) \]
\[ \forall y \ (\text{Smoker}(Bob) \lor \text{Friend}(Bob,y) \lor \text{ Jogger}(y)) \]

Lifted inference sometimes fails. Computing \( P(H_0) \) is \( \#P \)-hard in size database

[Suciu’11]
Are the Lifted Rules Complete?

You already know:

- Inference rules: PTIME data complexity
- Some queries: #$P$-hard data complexity

**Dichotomy Theorem** for UCQ / Mon. CNF

- If lifted rules succeed, then PTIME query
- If lifted rules fail, then query is #$P$-hard

Lifted rules are complete for UCQ!

[Dalvi and Suciu; JACM'11]
Why open world?
Knowledge Base Completion

**Given:**

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Einstein</td>
<td>Straus</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>Erdos</td>
<td>Straus</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>Einstein</td>
<td>Pauli</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

**Learn:**

0.8::Coauthor(x,y) :- Coauthor(z,x) ∧ Coauthor(z,y).

**Complete:**

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Straus</td>
<td>Pauli</td>
<td>0.504</td>
</tr>
<tr>
<td></td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>
Bayesian Learning Loop

Bayesian view on learning:

1. Prior belief:
   \[ P(\text{Coauthor}(\text{Straus}, \text{Pauli})) = 0.01 \]

2. Observe page
   \[ P(\text{Coauthor}(\text{Straus}, \text{Pauli} | \text{Page} ) = 0.2 \]

3. Observe page
   \[ P(\text{Coauthor}(\text{Straus}, \text{Pauli} | \text{Page}, \text{Page} ) = 0.3 \]

Principled and sound reasoning!
Problem: Broken Learning Loop

Bayesian view on learning:

1. Prior belief:
   \[ P(\text{Coauthor}(\text{Straus}, \text{Pauli})) = 0 \]

2. Observe page
   \[ P(\text{Coauthor}(\text{Straus}, \text{Pauli}) | \text{WHAAAAAT?}) = 0.2 \]

3. Observe page
   \[ P(\text{Coauthor}(\text{Straus}, \text{Pauli}) | \text{Ceylan, Darwiche, Van den Broeck; KR'16}) = 0.3 \]

This is mathematical nonsense!

[Ceylan, Darwiche, Van den Broeck; KR’16]
What we’d like to do...

∃x Coauthor(Einstein,x) ∧ Coauthor(Erdos,x)

Ernst Straus

Kristian Kersting, …

Justin Bieber, …
Open World DB

- What if fact missing?

- Probability 0 for:

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Einstein</td>
<td>Straus</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>Erdos</strong></td>
<td>Straus</td>
<td>0.6</td>
</tr>
<tr>
<td>Einstein</td>
<td>Pauli</td>
<td>0.9</td>
</tr>
<tr>
<td>Erdos</td>
<td>Renyi</td>
<td>0.7</td>
</tr>
<tr>
<td>Kersting</td>
<td>Natarajan</td>
<td>0.8</td>
</tr>
<tr>
<td>Luc</td>
<td>Paol</td>
<td>0.1</td>
</tr>
</tbody>
</table>
| ...        | ...        | ...

\[
Q_1 = \exists x \text{ Coauthor}(Einstein, x) \land \text{ Coauthor}(Erdos, x)
\]

\[
Q_2 = \exists x \text{ Coauthor}(Bieber, x) \land \text{ Coauthor}(Erdos, x)
\]

\[
Q_3 = \text{ Coauthor}(Einstein, \textbf{Straus}) \land \text{ Coauthor}(Erdos, \textbf{Straus})
\]

\[
Q_4 = \text{ Coauthor}(Einstein, \textbf{Bieber}) \land \text{ Coauthor}(Erdos, \textbf{Bieber})
\]

\[
Q_5 = \text{ Coauthor}(Einstein, \textbf{Bieber}) \land \neg \text{ Coauthor}(Einstein, \textbf{Bieber})
\]
Intuition

\[ Q_1 = \exists x \text{Coauthor}(Einstein, x) \land \text{Coauthor}(Erdos, x) \]

\[ Q_2 = \exists x \text{Coauthor}(Bieber, x) \land \text{Coauthor}(Erdos, x) \]

\[ Q_3 = \text{Coauthor}(Einstein, \text{Straus}) \land \text{Coauthor}(Erdos, \text{Straus}) \]

\[ Q_4 = \text{Coauthor}(Einstein, \text{Bieber}) \land \text{Coauthor}(Erdos, \text{Bieber}) \]

\[ Q_5 = \text{Coauthor}(Einstein, \text{Bieber}) \land \lnot \text{Coauthor}(Einstein, \text{Bieber}) \]

We know for sure that \( P(Q_1) \geq P(Q_3) \), \( P(Q_1) \geq P(Q_4) \)

and \( P(Q_3) \geq P(Q_5) \), \( P(Q_4) \geq P(Q_5) \) because \( P(Q_5) = 0 \).

We have strong evidence that \( P(Q_1) \geq P(Q_2) \).

[Ceylan, Darwiche, Van den Broeck; KR’16]
Problem: Curse of Superlinearity

- Reality is worse!
- Tuples are intentionally missing!
Problem: Curse of Superlinearity

Sibling

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Facebook scale
⇒ 200 Exabytes of data”

All Google storage is 2 exabytes…


[Ceylan, Darwiche, Van den Broeck; KR’16]
Problem: Model Evaluation

Given:

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Einstein</td>
<td>Straus</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>Erdos</td>
<td>Straus</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>Einstein</td>
<td>Pauli</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Learn:

0.8::Coauthor(x,y) :- Coauthor(z,x) ∧ Coauthor(z,y).

OR

0.6::Coauthor(x,y) :- Affiliation(x,z) ∧ Affiliation(y,z).

What is the likelihood, precision, accuracy, …?

[De Raedt et al; IJCAI’15]
Open-World Prob. Databases

**Intuition:** tuples can be added with $P < \lambda$

$$Q2 = \text{Coauthor}(\text{Einstein}, \text{Straus}) \land \text{Coauthor}(\text{Erdos}, \text{Straus})$$

$$0.7 \times \lambda \geq P(Q2) \geq 0$$

**Coauthor**

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Einstein</td>
<td>Straus</td>
<td>0.7</td>
</tr>
<tr>
<td>Einstein</td>
<td>Pauli</td>
<td>0.9</td>
</tr>
<tr>
<td>Erdos</td>
<td>Renyi</td>
<td>0.7</td>
</tr>
<tr>
<td>Kersting</td>
<td>Natarajan</td>
<td>0.8</td>
</tr>
<tr>
<td>Luc</td>
<td>Paol</td>
<td>0.1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Coauthor**

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Einstein</td>
<td>Straus</td>
<td>0.7</td>
</tr>
<tr>
<td>Einstein</td>
<td>Pauli</td>
<td>0.9</td>
</tr>
<tr>
<td>Erdos</td>
<td>Renyi</td>
<td>0.7</td>
</tr>
<tr>
<td>Kersting</td>
<td>Natarajan</td>
<td>0.8</td>
</tr>
<tr>
<td>Luc</td>
<td>Paol</td>
<td>0.1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td><strong>Erdos</strong></td>
<td><strong>Straus</strong></td>
<td>$$\lambda$$</td>
</tr>
</tbody>
</table>
How open-world query evaluation?
UCQ / Monotone CNF

- Lower bound = closed-world probability
- Upper bound = probability after adding all tuples with probability $\lambda$

- Polynomial time 😊
- Quadratic blow-up 😞
- 200 exabytes … again 😞
Closed-World Lifted Query Eval

\[ Q = \exists x \exists y \text{Scientist}(x) \land \text{Coauthor}(x,y) \]

\[ P(Q) = 1 - \prod_{A \in \text{Domain}} (1 - P(\text{Scientist}(A) \land \exists y \text{Coauthor}(A,y))) \]

\[ = 1 - (1 - P(\text{Scientist}(A) \land \exists y \text{Coauthor}(A,y))) \times (1 - P(\text{Scientist}(B) \land \exists y \text{Coauthor}(B,y))) \times (1 - P(\text{Scientist}(C) \land \exists y \text{Coauthor}(C,y))) \times (1 - P(\text{Scientist}(D) \land \exists y \text{Coauthor}(D,y))) \times (1 - P(\text{Scientist}(E) \land \exists y \text{Coauthor}(E,y))) \times (1 - P(\text{Scientist}(F) \land \exists y \text{Coauthor}(F,y))) \times \ldots \]

Check independence:
- \(\text{Scientist}(A) \land \exists y \text{Coauthor}(A,y)\)
- \(\text{Scientist}(B) \land \exists y \text{Coauthor}(B,y)\)

Decomposable \(\lor\)-Rule

Complexity PTIME
Closed-World Lifted Query Eval

\[ Q = \exists x \ \exists y \ \text{Scientist}(x) \land \text{Coauthor}(x,y) \]

\[ P(Q) = 1 - \prod_{A \in \text{Domain}} (1 - P(\text{Scientist}(A) \land \exists y \ \text{Coauthor}(A,y))) \]

\[ = 1 - (1 - P(\text{Scientist}(A) \land \exists y \ \text{Coauthor}(A,y))) \times (1 - P(\text{Scientist}(B) \land \exists y \ \text{Coauthor}(B,y))) \times (1 - P(\text{Scientist}(C) \land \exists y \ \text{Coauthor}(C,y))) \]
\[ \times (1 - P(\text{Scientist}(D) \land \exists y \ \text{Coauthor}(D,y))) \times (1 - P(\text{Scientist}(E) \land \exists y \ \text{Coauthor}(E,y))) \]
\[ \times (1 - P(\text{Scientist}(F) \land \exists y \ \text{Coauthor}(F,y))) \]

\[ \cdots \]

No supporting facts in database!

Probability 0 in closed world

Ignore these queries!

Complexity linear time!
Open-World Lifted Query Eval

\[ Q = \exists x \ \exists y \ \text{Scientist}(x) \land \text{Coauthor}(x,y) \]

\[ P(Q) = 1 - \prod_{A \in \text{Domain}} (1 - P(\text{Scientist}(A) \land \exists y \ \text{Coauthor}(A,y))) \]

\[ = 1 - (1 - P(\text{Scientist}(A) \land \exists y \ \text{Coauthor}(A,y))) \times (1 - P(\text{Scientist}(B) \land \exists y \ \text{Coauthor}(B,y))) \times (1 - P(\text{Scientist}(C) \land \exists y \ \text{Coauthor}(C,y))) \times (1 - P(\text{Scientist}(D) \land \exists y \ \text{Coauthor}(D,y))) \times (1 - P(\text{Scientist}(E) \land \exists y \ \text{Coauthor}(E,y))) \times (1 - P(\text{Scientist}(F) \land \exists y \ \text{Coauthor}(F,y))) \times \ldots \]

\[ \text{No supporting facts in database!} \]
\[ \text{Probability } p \text{ in closed world} \]

Complexity PTIME!
Open-World Lifted Query Eval

\[ Q = \exists x \exists y \text{Scientist}(x) \land \text{Coauthor}(x,y) \]

\[ P(Q) = 1 - \prod_{A \in \text{Domain}} (1 - P(\text{Scientist}(A) \land \exists y \text{Coauthor}(A,y)) \]

\[ = 1 - (1 - P(\text{Scientist}(A) \land \exists y \text{Coauthor}(A,y)) \times (1 - P(\text{Scientist}(B) \land \exists y \text{Coauthor}(B,y)) \times (1 - P(\text{Scientist}(C) \land \exists y \text{Coauthor}(C,y)) \times (1 - P(\text{Scientist}(D) \land \exists y \text{Coauthor}(D,y)) \times (1 - P(\text{Scientist}(E) \land \exists y \text{Coauthor}(E,y)) \times (1 - P(\text{Scientist}(F) \land \exists y \text{Coauthor}(F,y)) \]

No supporting facts in database!

Probability \( p \) in closed world

All together, probability \((1-p)^k\)

Do symmetric lifted inference

Complexity linear time!
Complexity Results

Linear ⊆ P ⊆ NP ⊆ PP ⊆ P^{PP} ⊆ NP^{PP} ⊆ PSpace ⊆ ExpTime

[Ceylan’16]
Implement PDB Query in SQL

- Convert to nested SQL recursively
- Open-world existential quantification

```
SELECT (1.0-(1.0-pUse)*power(1.0-0.0001,(4-ct))) AS pUse
FROM
  (SELECT ior(COALESCE(pUse,0)) AS pUse,
  count(*) AS ct
  FROM SQL(conjunction)

0.0001 = open-world probability; 4 = # open-world query instances
ior = Independent OR aggregate function
```

- Conjunction

```
SELECT q9.c5,
  COALESCE(q9.pUse,λ)*COALESCE(q10.pUse,λ) AS pUse
FROM
  SQL(Q(X)) OUTER JOIN SQL(P(X))
```

- Run as single PostgreSQL query!

Q = ∃x P(x) ∧ Q(x)

[Tal Friedman, Eric Gribkoff]
Out of memory trying to run the ProbLog query with 70 constants in domain
OpenPDB vs Problog Running Times (s)

Size of Domain

- PDB
- Problog

Linear (PDB)

12.5 million random variables!
What is the broader picture?
The Broader Picture

- Statistical relational learning (e.g., Markov logic)
  Open-domain models (BLOG)
- Probabilistic description logics
- Certain query answers in databases
- Open information extraction
- Learning from positive-only examples
- Imprecise probabilities
  Credal sets, interval probability, qualitative uncertainty
- Credal Bayesian networks
Related Work: Lifted Probabilistic Inference

Probability that Card1 is Hearts? 1/4

[Van den Broeck; AAAI-KRR'15]
Open-World Lifted Query Eval

\[ Q = \exists x \exists y \text{Smoker}(x) \land \text{Friend}(x,y) \]

\[ P(Q) = 1 - \prod_{A \in \text{Domain}} (1 - P(\text{Scientist}(A) \land \exists y \text{Coauthor}(A,y))) \]

\[ = 1 - (1 - P(\text{Scientist}(A) \land \exists y \text{Coauthor}(A,y))) \times (1 - P(\text{Scientist}(B) \land \exists y \text{Coauthor}(B,y))) \times (1 - P(\text{Scientist}(C) \land \exists y \text{Coauthor}(C,y))) \times (1 - P(\text{Scientist}(D) \land \exists y \text{Coauthor}(D,y))) \times (1 - P(\text{Scientist}(E) \land \exists y \text{Coauthor}(E,y))) \times (1 - P(\text{Scientist}(F) \land \exists y \text{Coauthor}(F,y))) \ldots \]

All together, probability \((1-p)^k\)

Open-world query evaluation on empty db = Lifted inference in AI
Conclusions

- Relational probabilistic reasoning is **frontier** and **integration** of AI, KR, ML, DB, TH, etc.
- We need
  - relational models and logic
  - probabilistic models and statistical learning
  - algorithms that scale
- Open-world data model
  - semantics makes sense
  - FREE for UCQs
  - expensive otherwise
References


References


References


References


References

