# Towards Querying Probabilistic Knowledge Bases

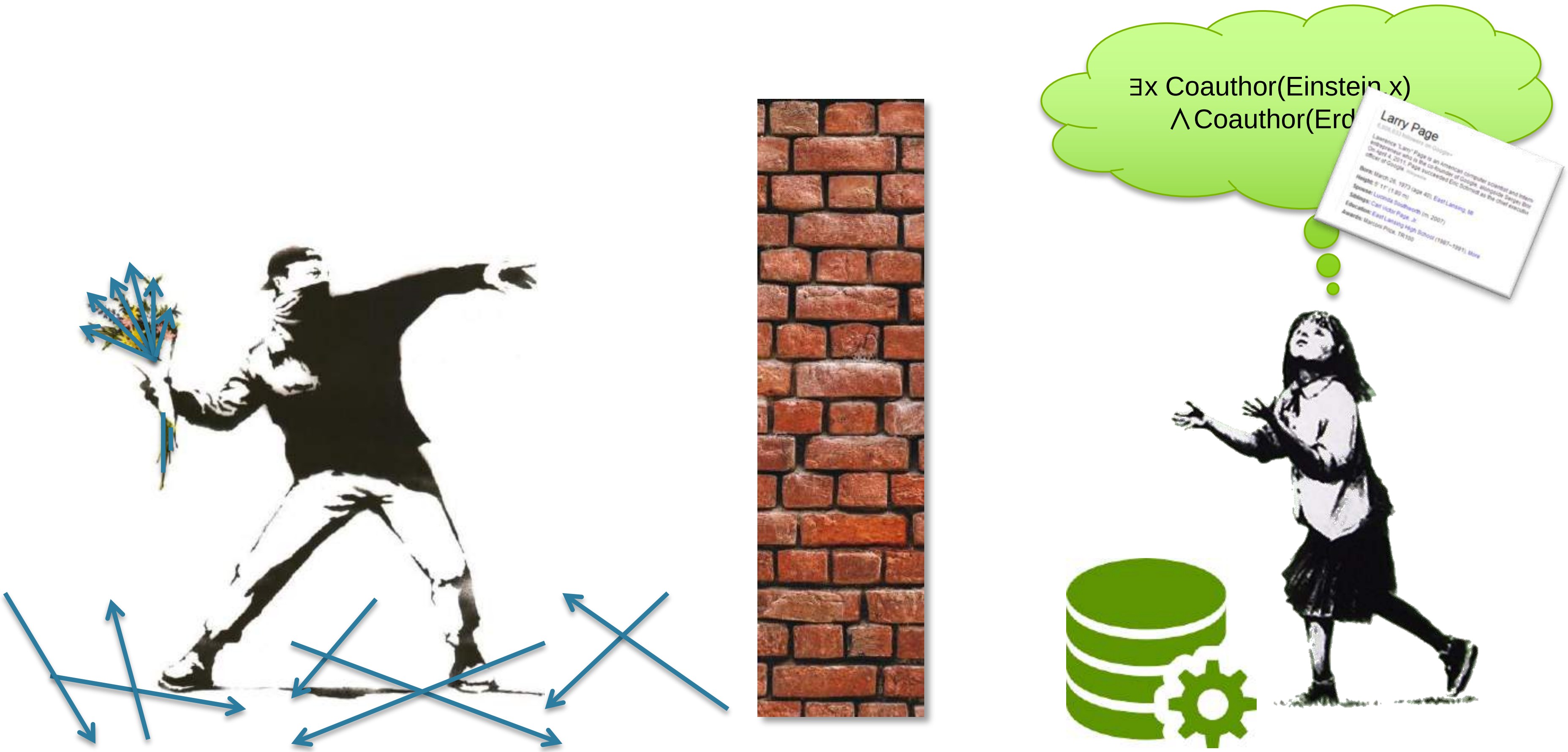
Guy Van den Broeck

First Conference on
Automated Knowledge Base Construction

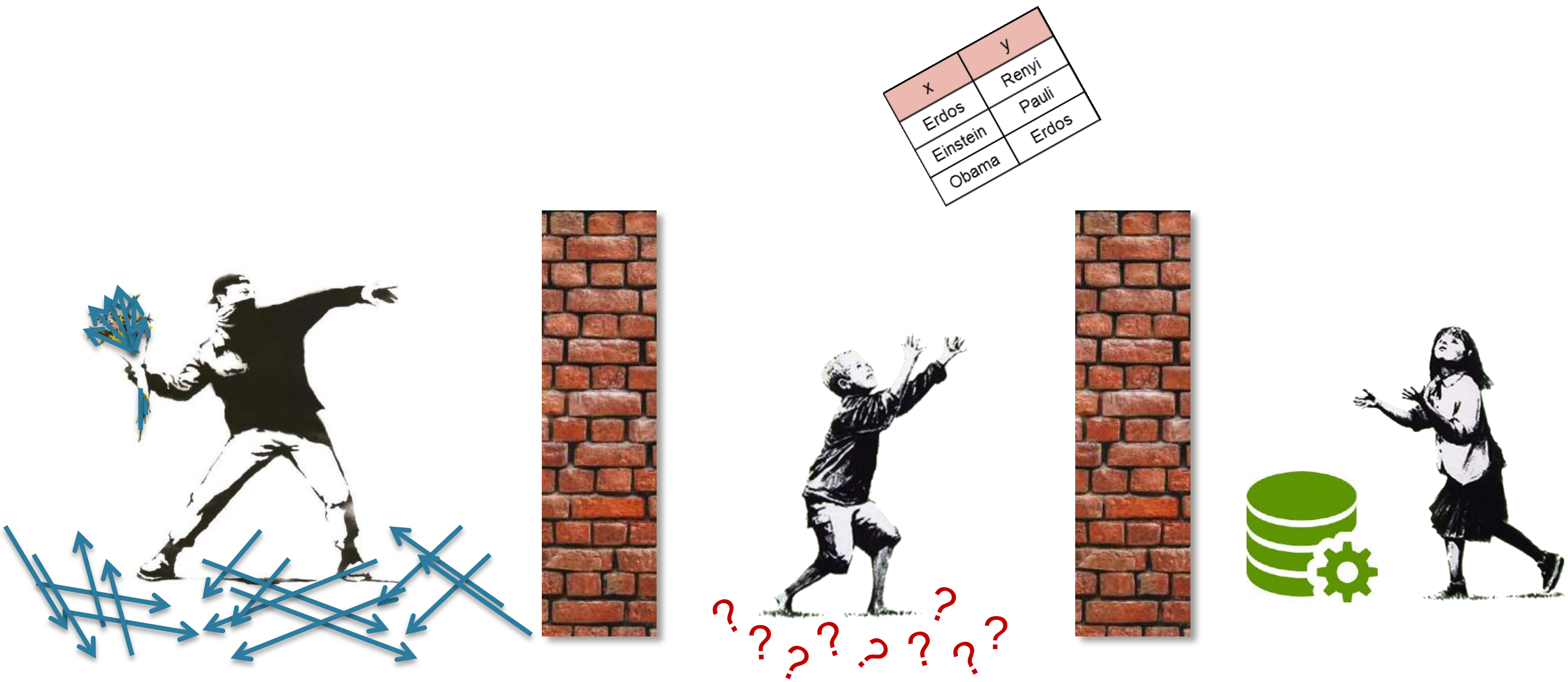
May 20, 2019

*Cartoon Motivation*

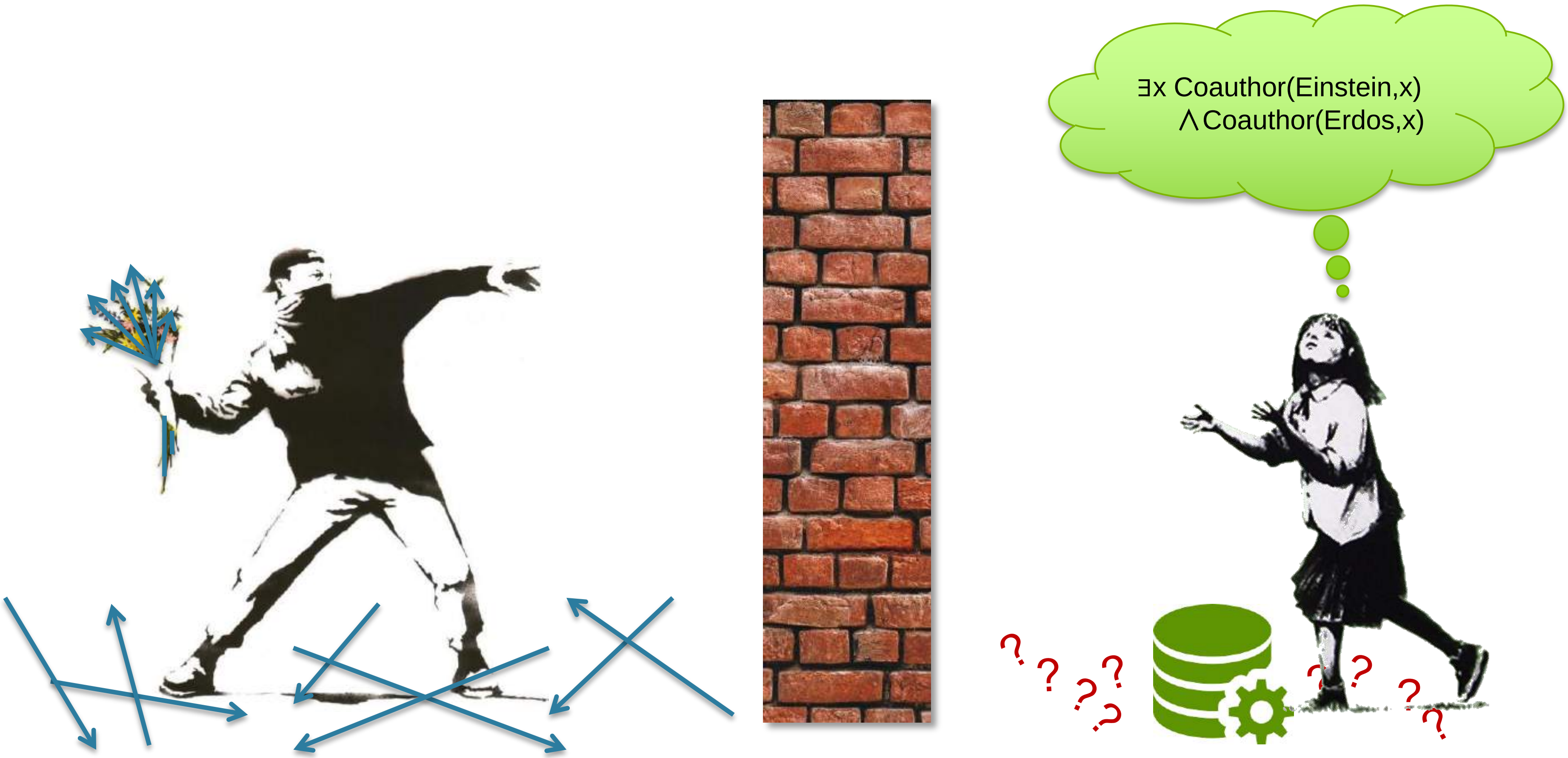
# Cartoon Motivation:
# Relational Embedding Models



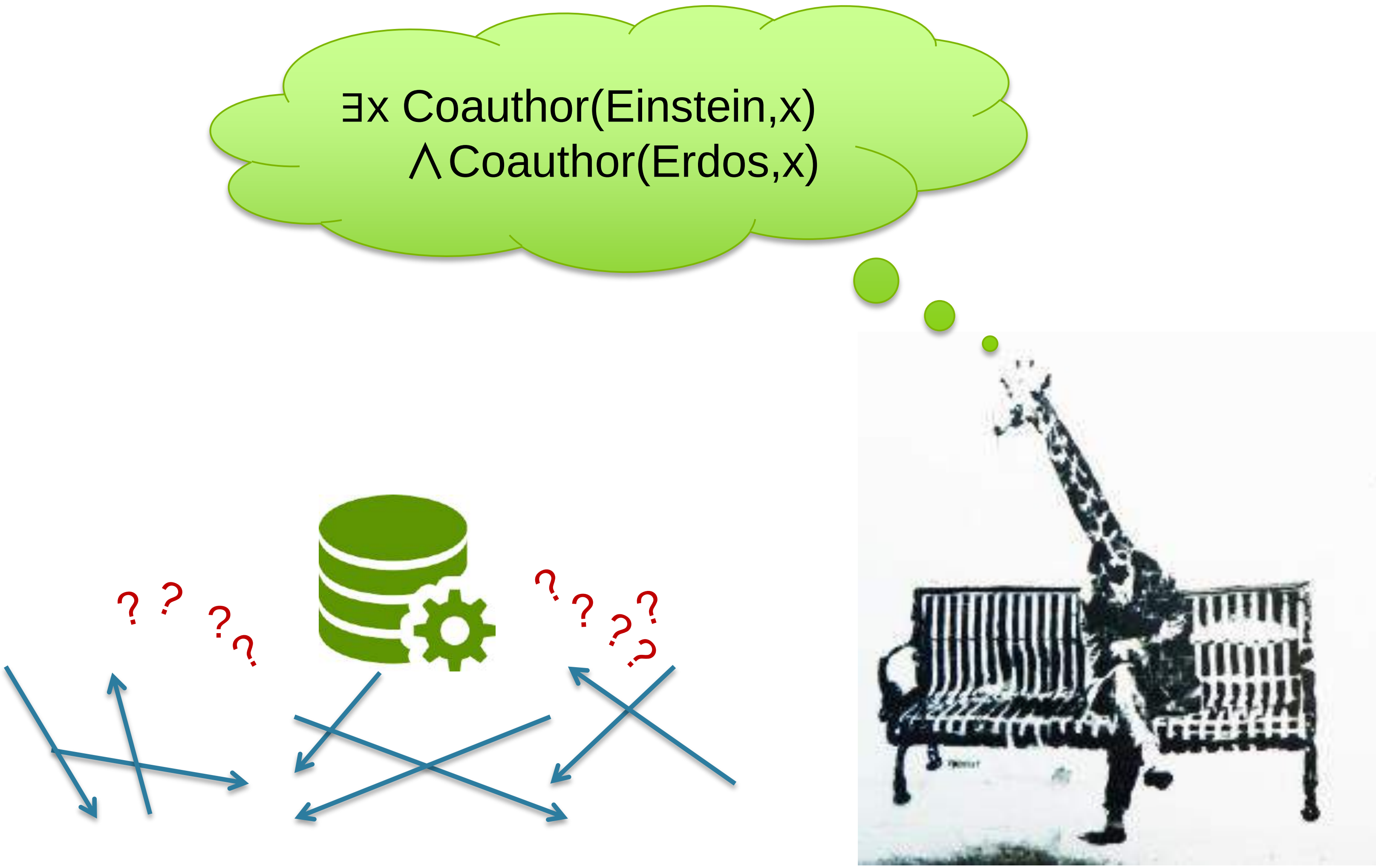
# Cartoon Motivation 2: Relational Embedding Models



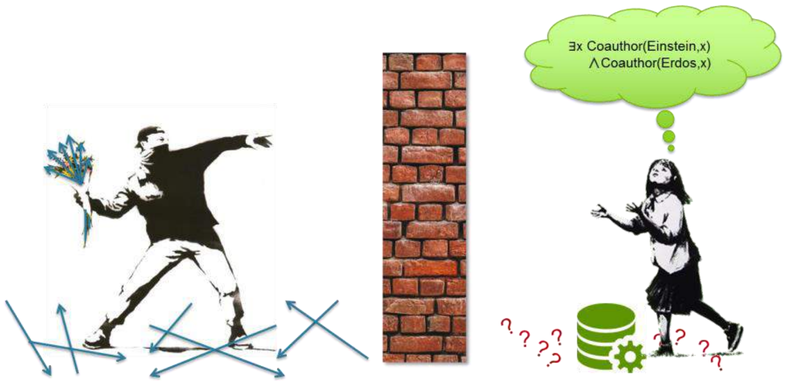
# Goal 1: Probabilistic Query Evaluation



# Goal 2: Querying Relational Embedding Models



# *Probabilistic Query Evaluation*



# What we'd like to do…

Has anyone published a paper with both Erdos and Einstein

All News Images Videos Shopping More Search tools

About 82,400 results (0.73 seconds)

**Erdős number - Wikipedia, the free encyclopedia**
https://en.wikipedia.org/wiki/**Erdős**_number ▾ Wikipedia ▾
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. ... $\boxed{\text{Albert Einstein and Sheldon Lee Glashow have an Erdős number of 2.}}$ ... and mathematician Ruth Williams, **both** of whom **have** an **Erdős** number of 2.

**Erdős–Bacon number - Wikipedia, the free encyclopedia**
https://en.wikipedia.org/wiki/**Erdős**–Bacon_number ▾ Wikipedia ▾
This article possibly **contains** previously unpublished synthesis of **published** ...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 ...

# What we'd like to do…

$\exists x$ Coauthor(Einstein,x) $\wedge$ Coauthor(Erdos,x)

All News Images Videos Shopping More Search tools

About 82,400 results (0.73 seconds)

**Erdős number - Wikipedia, the free encyclopedia**
https://en.wikipedia.org/wiki/**Erdős**_number Wikipedia
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. ... and mathematician Ruth Williams, **both** of whom **have** an **Erdős** number of 2.

**Erdős–Bacon number - Wikipedia, the free encyclopedia**
https://en.wikipedia.org/wiki/**Erdős**–Bacon_number Wikipedia
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 ...

# Einstein is in the Knowledge Graph

Albert Einstein

All News Images Books Videos More ▾ Search tools

About 82,800,000 results (0.45 seconds)

**The Official Licensing Site of Albert Einstein**
**einstein**.biz/ ▾
Welcome to the Official Licensing Site of **Albert Einstein**. Learn more about **Albert Einstein** and contact us today for any commercial licensing inquiries.

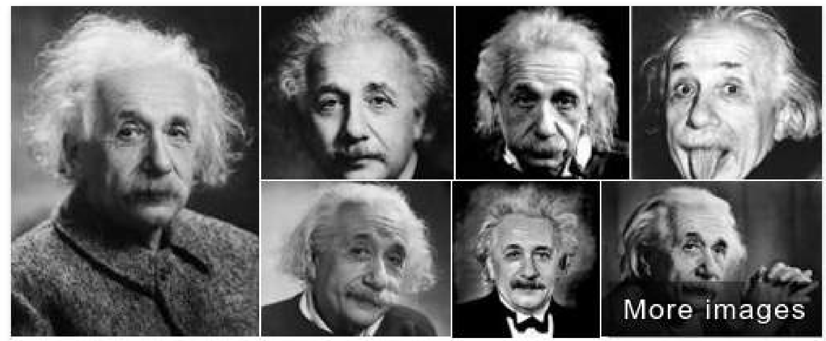
**Albert Einstein - Wikipedia, the free encyclopedia**
https://en.wikipedia.org/wiki/**Albert_Einstein** ▾ Wikipedia ▾
**Albert Einstein** (/ˈaɪnstaɪn/; German: [ˈalbɛɐ̯t ˈaɪnʃtaɪn] ( listen); 14 March 1879 – 18 April 1955) was a German-born theoretical physicist.
Hans Albert Einstein - Mass–energy equivalence - Eduard Einstein - Elsa Einstein

**Albert Einstein (@AlbertEinstein) | Twitter**
https://twitter.com/AlbertEinstein

16 hours ago - View on Twitter
ICYMI, Albert Einstein knew a thing or two about being romantic. Learn about the love letters he wrote. guff.com/didnt-know-einst…

20 hours ago - View on Twitter
An interesting read on Einstein's superstar status. What are your thoughts? twitter.com/aeonmag/statu…

**Albert Einstein - Biographical - Nobelprize.org**
www.nobelprize.org/nobel_prizes/physics/.../**einstein**-bio.htm... ▾ Nobel Prize ▾
**Albert Einstein** was born at Ulm, in Württemberg, Germany, on March 14, 1879. ...
Later, they moved to Italy and Albert continued his education at Aarau




## Albert Einstein

Theoretical Physicist

Albert Einstein was a German-born theoretical physicist. He developed the general theory of relativity, one of the two pillars of modern physics. Einstein's work is also known for its influence on the philosophy of science. Wikipedia

**Born:** March 14, 1879, Ulm, Germany

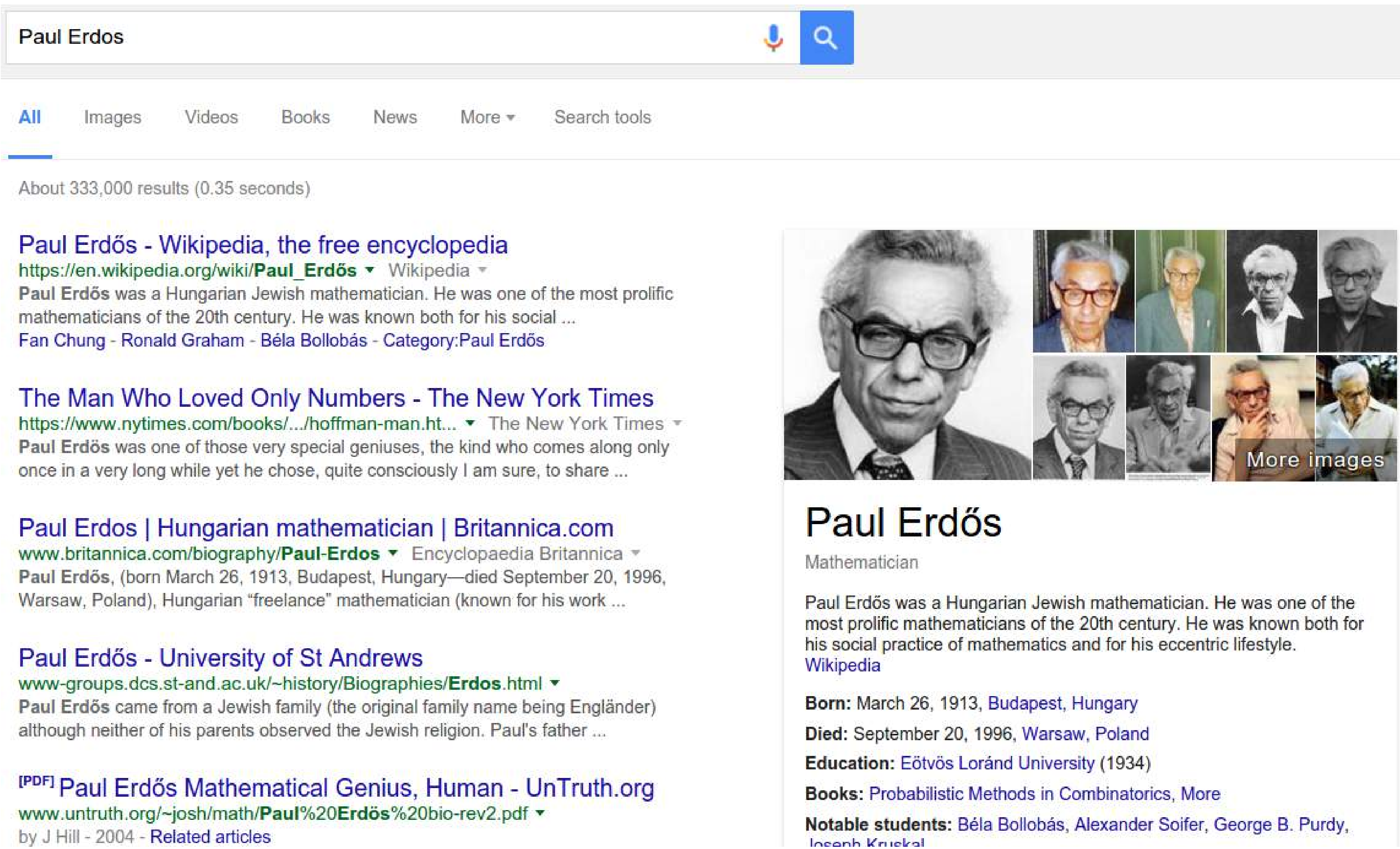
**Died:** April 18, 1955, Princeton, NJ

**Influenced by:** Isaac Newton, Mahatma Gandhi, More

**Children:** Eduard Einstein, Lieserl Einstein, Hans Albert Einstein

**Spouse:** Elsa Einstein (m. 1919–1936), Mileva Marić (m. 1903–1919)

# Erdős is in the Knowledge Graph



Paul Erdos

All Images Videos Books News More Search tools

About 333,000 results (0.35 seconds)

Paul Erdős - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Paul_Erdős Wikipedia
**Paul Erdős** was a Hungarian Jewish mathematician. He was one of the most prolific mathematicians of the 20th century. He was known both for his social ...
Fan Chung - Ronald Graham - Béla Bollobás - Category:Paul Erdős

The Man Who Loved Only Numbers - The New York Times
https://www.nytimes.com/books/.../hoffman-man.ht... The New York Times
**Paul Erdös** was one of those very special geniuses, the kind who comes along only once in a very long while yet he chose, quite consciously I am sure, to share ...

Paul Erdos | Hungarian mathematician | Britannica.com
www.britannica.com/biography/**Paul-Erdos** Encyclopaedia Britannica
**Paul Erdős**, (born March 26, 1913, Budapest, Hungary—died September 20, 1996, Warsaw, Poland), Hungarian “freelance” mathematician (known for his work ...

Paul Erdős - University of St Andrews
www-groups.dcs.st-and.ac.uk/~history/Biographies/**Erdos**.html
**Paul Erdős** came from a Jewish family (the original family name being Engländer) although neither of his parents observed the Jewish religion. Paul's father ...

[PDF] Paul Erdős Mathematical Genius, Human - UnTruth.org
www.untruth.org/~josh/math/**Paul**%20**Erdös**%20bio-rev2.pdf
by J Hill - 2004 - Related articles

More images

Paul Erdős

Mathematician

Paul Erdős was a Hungarian Jewish mathematician. He was one of the most prolific mathematicians of the 20th century. He was known both for his social practice of mathematics and for his eccentric lifestyle.
Wikipedia

**Born:** March 26, 1913, Budapest, Hungary

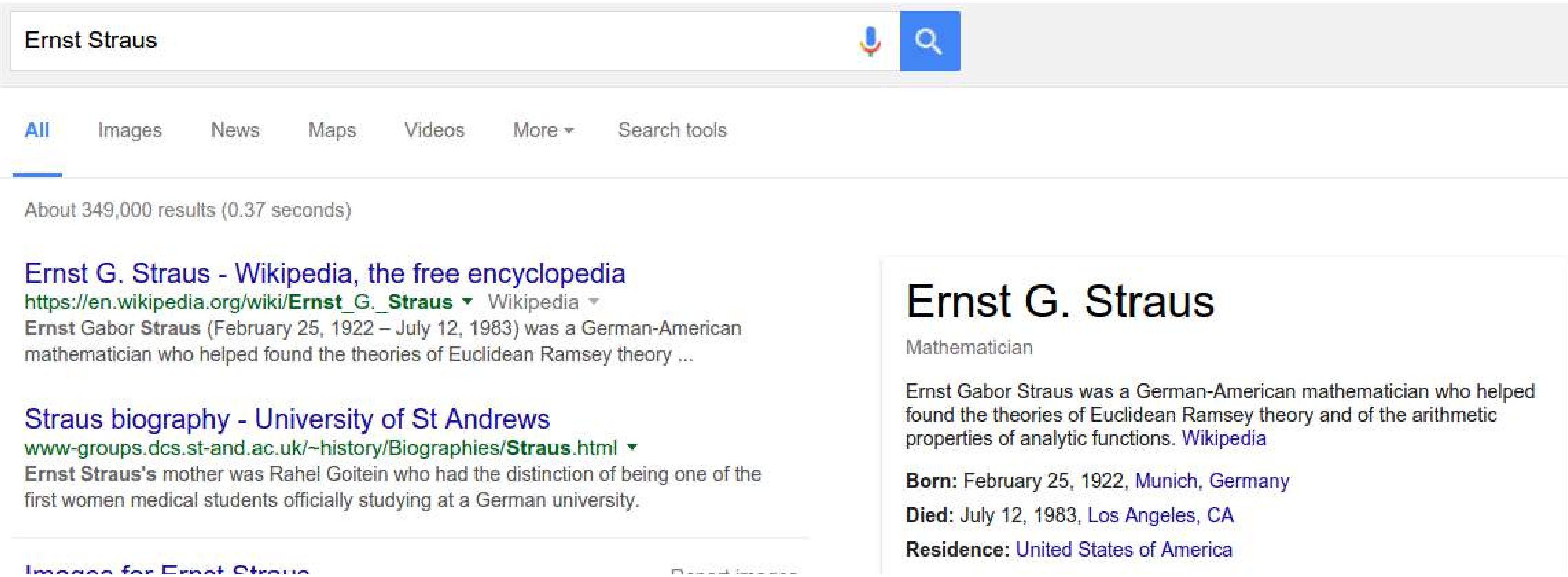
**Died:** September 20, 1996, Warsaw, Poland

**Education:** Eötvös Loránd University (1934)

**Books:** Probabilistic Methods in Combinatorics, More

**Notable students:** Béla Bollobás, Alexander Soifer, George B. Purdy, Joseph Kruskal

# This guy is in the Knowledge Graph




… and he published with both Einstein and Erdos!

# Desired Query Answer


Ernst Straus

Barack Obama, …

Justin Bieber, …

1. Fuse uncertain information from web

   ⇒ **Embrace probability!**

2. Cannot come from labeled data

   ⇒ **Embrace query eval!**

# Probabilistic Databases


- Tuple-independent probabilistic database

**Scientist**

| x | P |
|---|---|
| Erdos | 0.9 |
| Einstein | 0.8 |
| Pauli | 0.6 |

**Coauthor**

| x | y | P |
|---|---|---|
| Erdos | Renyi | 0.6 |
| Einstein | Pauli | 0.7 |
| Obama | Erdos | 0.1 |

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

[VdB&Suciu'17]

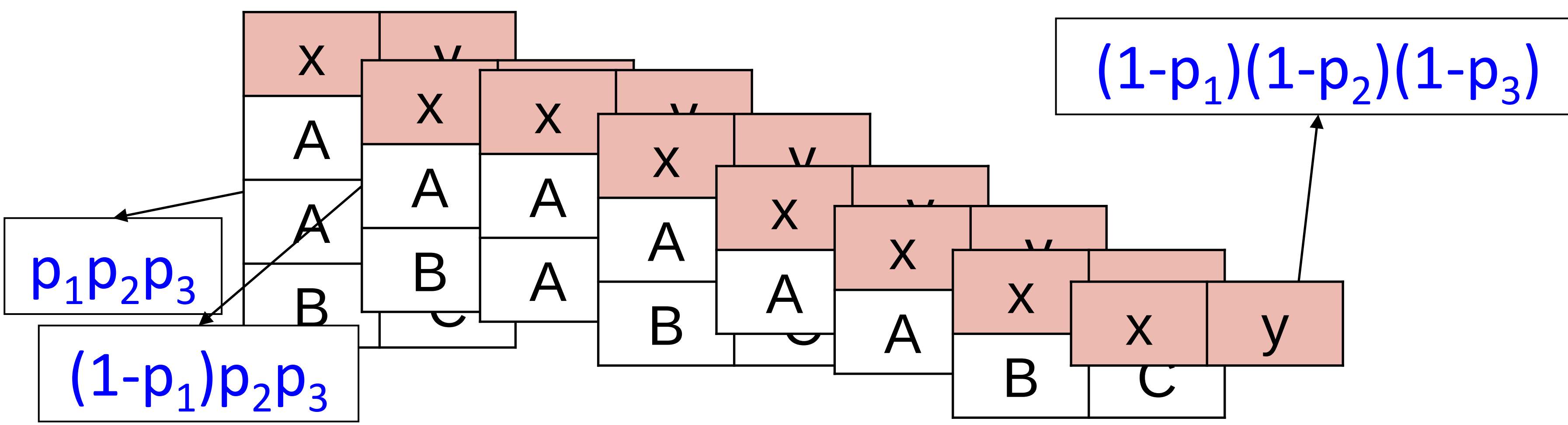
# Tuple-Independent Probabilistic DB

Probabilistic database D:

**Coauthor**

| x | y | P |
|---|---|---|
| A | B | $p_1$ |
| A | C | $p_2$ |
| B | C | $p_3$ |

Possible worlds semantics:




[VdB&Suciu'17]

# Probabilistic Query Evaluation

$Q = \exists x \exists y \; \text{Scientist}(x) \wedge \text{Coauthor}(x,y)$

$$P(Q) = 1 - \{1 - p_1 * [\; 1-(1-q_1)*(1-q_2) \;]\} * \{1 - p_2 * [\; 1-(1-q_3)*(1-q_4)*(1-q_5) \;]\}$$

**Scientist**

| x | P |
|---|---|
| A | $p_1$ |
| B | $p_2$ |
| C | $p_3$ |

**Coauthor**

| x | y | P |
|---|---|---|
| A | D | $q_1$ |
| A | E | $q_2$ |
| B | F | $q_3$ |
| B | G | $q_4$ |
| B | H | $q_5$ |

# Lifted Inference Rules

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

$P(\neg Q) = 1 - P(Q)$

Negation

$P(Q1 \wedge Q2) = P(Q1)\, P(Q2)$
$P(Q1 \vee Q2) = 1 - (1 - P(Q1))\, (1 - P(Q2))$

Decomposable ∧,∨

$P(\forall z\, Q) = \Pi_{A \in Domain}\; P(Q[A/z])$
$P(\exists z\, Q) = 1 - \Pi_{A \in Domain}\; (1 - P(Q[A/z]))$

Decomposable ∃,∀

$P(Q1 \wedge Q2) = P(Q1) + P(Q2) - P(Q1 \vee Q2)$
$P(Q1 \vee Q2) = P(Q1) + P(Q2) - P(Q1 \wedge Q2)$

Inclusion/
exclusion

# Example Query Evaluation

$Q = \exists x\ \exists y\ \text{Scientist}(x) \wedge \text{Coauthor}(x,y)$

Decomposable ∃-Rule

$$P(Q) = 1 - \Pi_{A \in \text{Domain}} (1 - P(\text{Scientist}(A) \wedge \exists y\ \text{Coauthor}(A,y))$$

Check independence:

$\text{Scientist}(A) \wedge \exists y\ \text{Coauthor}(A,y)$

$\text{Scientist}(B) \wedge \exists y\ \text{Coauthor}(B,y)$

$$\begin{aligned}
= 1 - &(1 - P(\text{Scientist}(A) \wedge \exists y\ \text{Coauthor}(A,y)) \\
&\times (1 - P(\text{Scientist}(B) \wedge \exists y\ \text{Coauthor}(B,y)) \\
&\times (1 - P(\text{Scientist}(C) \wedge \exists y\ \text{Coauthor}(C,y)) \\
&\times (1 - P(\text{Scientist}(D) \wedge \exists y\ \text{Coauthor}(D,y)) \\
&\times (1 - P(\text{Scientist}(E) \wedge \exists y\ \text{Coauthor}(E,y)) \\
&\times (1 - P(\text{Scientist}(F) \wedge \exists y\ \text{Coauthor}(F,y)) \\
&\quad \ldots
\end{aligned}$$

Complexity PTIME

# Limitations

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

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

… does not apply:

$H_0$[Alice/x] and $H_0$[Bob/x] are dependent:

Dependent

$\forall y\ (\text{Smoker}(\text{Alice}) \lor \text{Friend}(\text{Alice},y) \lor \text{Jogger}(y))$

$\forall y\ (\text{Smoker}(\text{Bob}) \lor \text{Friend}(\text{Bob},y) \lor \text{Jogger}(y))$

Lifted inference sometimes fails.

# Are the Lifted Rules Complete?

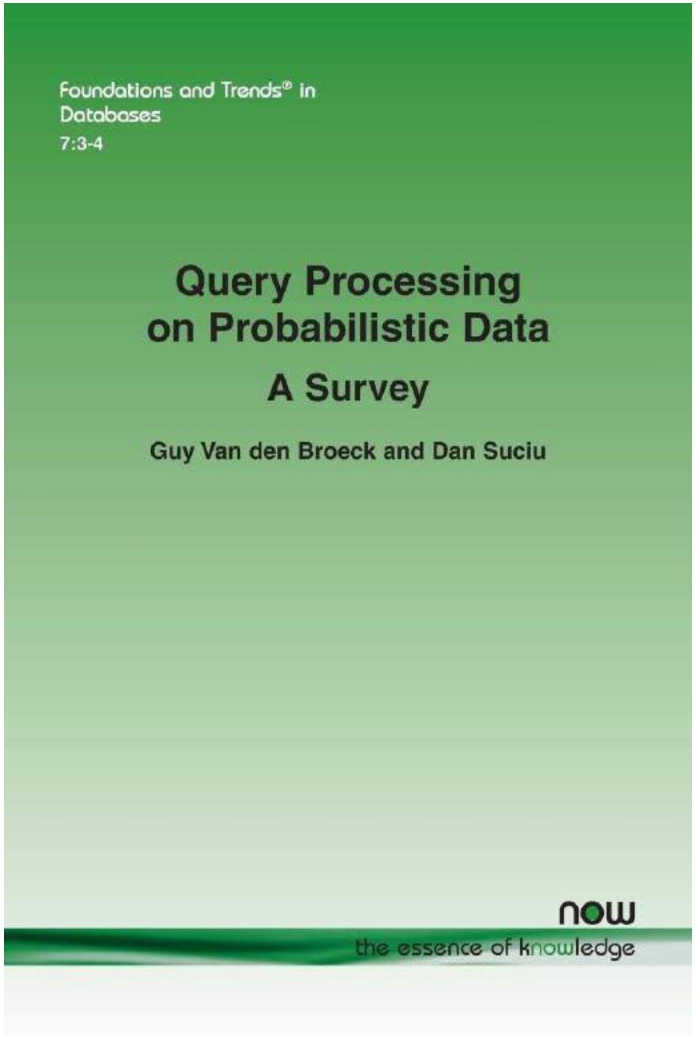
**Dichotomy Theorem** for Unions of Conjunction Queries / Monotone CNF

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

Lifted rules are complete for UCQ!

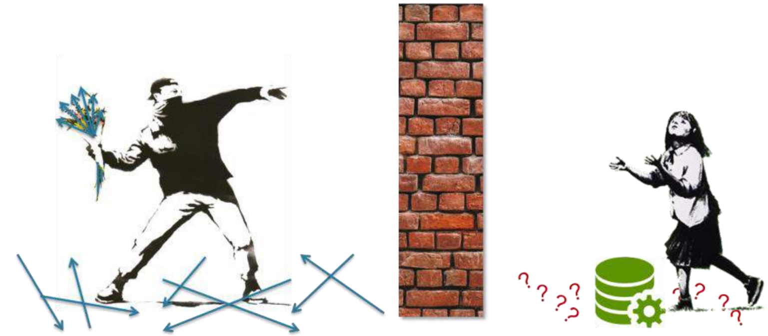
[Dalvi and Suciu;JACM'11]

# Commercial Break




- Survey book
  http://www.nowpublishers.com/article/Details/DBS-052
- IJCAI 2016 tutorial
  http://web.cs.ucla.edu/~guyvdb/talks/IJCAI16-tutorial/

# Throwing Relational Embedding Models Over the Wall



- Notion of distance $d(h, r, t)$ in vector space (Euclidian, 1-cosine, …)
- Probabilistic semantics:
  - Distance $d(h, r, t) = 0$ is certainty
  - Distance $d(h, r, t) > 0$ is uncertainty

$$\mathrm{P}(\mathrm{r}(\mathrm{h},\mathrm{t})) \approx \mathrm{e}^{-\alpha \cdot d(h,r,t)}$$

# What About Tuple-Independence?

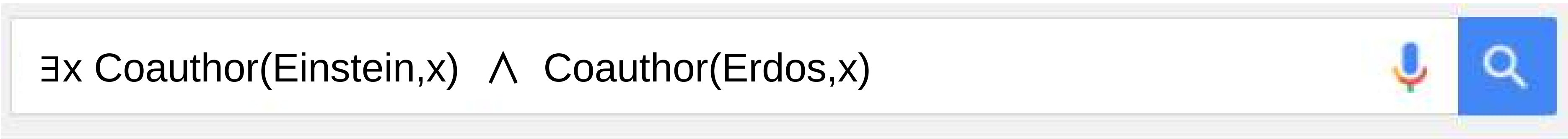
- Deterministic databases
  = tuple-independent
- Relational embedding models
  = tuple-independent

  *At no point do we model*
  *<u>joint</u> uncertainty between tuples*

- We can capture correlations, but query evaluation becomes much harder!
  - See probabilistic database literature
  - See statistical relational learning literature

*So everything is solved?*

# What we’d like to do…




Ernst Straus

Kristian Kersting, …

Justin Bieber, …

# Open World DB

- What if fact missing?
- Probability 0 for:

**Coauthor**

| X | Y | P |
|---|---|---|
| Einstein | Straus | 0.7 |
| ~~**Erdos**~~ | ~~**Straus**~~ | ~~**0.6**~~ |
| Einstein | Pauli | 0.9 |
| Erdos | Renyi | 0.7 |
| Kersting | Natarajan | 0.8 |
| Luc | Paol | 0.1 |
| … | … | … |

Q1 = ∃x Coauthor(Einstein,**x**) ∧ Coauthor(Erdos,**x**)

Q2 = ∃x Coauthor(Bieber,**x**) ∧ Coauthor(Erdos,**x**)

Q3 = Coauthor(Einstein,**Straus**) ∧ Coauthor(Erdos,**Straus**)

Q4 = Coauthor(Einstein,**Bieber**) ∧ Coauthor(Erdos,**Bieber**)

Q5 = Coauthor(Einstein,**Bieber**) ∧ ¬Coauthor(**Einstein**,**Bieber**)

# Intuition

| X | Y | P |
|---|---|---|
| Einstein | Straus | 0.7 |
| **~~Erdos~~** | **~~Straus~~** | **~~0.6~~** |
| Einstein | Pauli | 0.9 |
| Erdos | Renyi | 0.7 |
| Kersting | Natarajan | 0.8 |
| Luc | Paol | 0.1 |
| … | … | … |

Q1 = ∃x Coauthor(Einstein,**x**) ∧ Coauthor(Erdos,**x**)

Q2 = ∃x Coauthor(Bieber,**x**) ∧ Coauthor(Erdos,**x**)

Q3 = Coauthor(Einstein,**Straus**) ∧ Coauthor(Erdos,**Straus**)

Q4 = Coauthor(Einstein,**Bieber**) ∧ Coauthor(Erdos,**Bieber**)

Q5 = Coauthor(Einstein,**Bieber**) ∧ ¬Coauthor(**Einstein**,**Bieber**)

We know for sure that P(Q1) ≥ P(Q3), P(Q1) ≥ P(Q4)

and P(Q3) ≥ P(Q5), P(Q4) ≥ P(Q5) because P(Q5) = 0.

We have strong evidence that P(Q1) ≥ P(Q2).

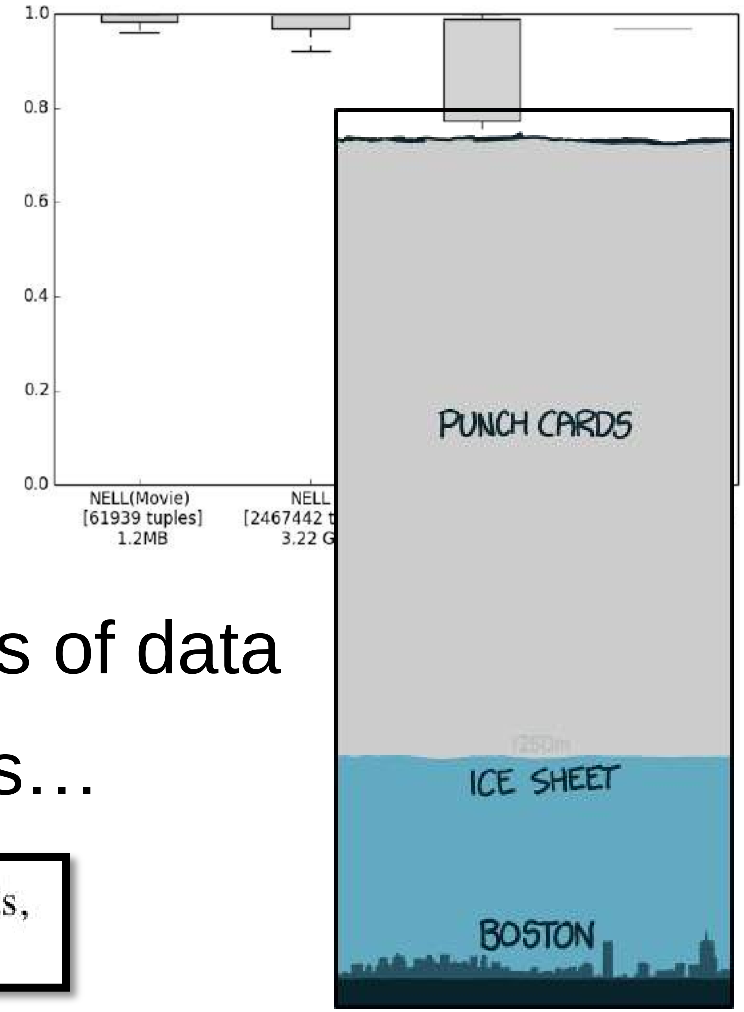
[Ceylan, Darwiche, Van den Broeck; KR'16]

# Problem: Curse of Superlinearity

Reality is worse: tuples **intentionally** missing!

**Sibling**

| x | y | P |
|---|---|---|
| … | … | … |




Facebook scale ⇒ 200 Exabytes of data

All Google storage is 2 exabytes…

Randall Munroe. Google's datacenters on punch cards, 2015.

[Ceylan, Darwiche, Van den Broeck; KR'16]

# Bayesian Learning Loop

Bayesian view on learning:

1. Prior belief:

   $P(\texttt{Coauthor(Straus,Pauli)}) = 0.01$

2. Observe page

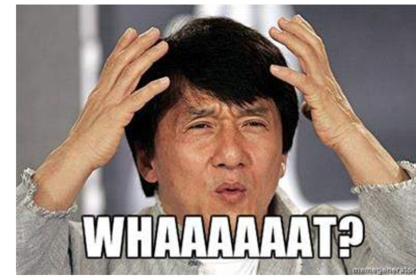
   $P(\texttt{Coauthor(Straus,Pauli} \mid$  $) = 0.2$

3. Observe page

   $P(\texttt{Coauthor(Straus,Pauli)} \mid$  ,  $) = 0.3$

Principled and sound reasoning!

# Problem: Broken Learning Loop




Bayesian view on learning:

1. Prior belief:

   P(Coauthor(Straus,Pauli)) = 0

2. Observe page

   P(Coauthor(Straus,Pauli|  ) = 0.2

3. Observe page

   P(Coauthor(Straus,Pauli)|  ,  ) = 0.3

**This is mathematical nonsense!**

[Ceylan, Darwiche, Van den Broeck; KR'16]

# Problem: Model Evaluation

Given:

**Coauthor**

| x | y | P |
|---|---|---|
| Einstein | Straus | 0.7 |
| Erdos | Straus | 0.6 |
| Einstein | Pauli | 0.9 |
| … | … | … |

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

<u>Intuition</u>: tuples can be added with $P < \lambda$

Q2 = Coauthor(Einstein,**Straus**) $\land$ Coauthor(Erdos,**Straus**)

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

**Coauthor**

| X | Y | P |
|---|---|---|
| Einstein | Straus | 0.7 |
| Einstein | Pauli | 0.9 |
| Erdos | Renyi | 0.7 |
| Kersting | Natarajan | 0.8 |
| Luc | Paol | 0.1 |
| … | … | … |

**Coauthor**

| X | Y | P |
|---|---|---|
| Einstein | Straus | 0.7 |
| Einstein | Pauli | 0.9 |
| Erdos | Renyi | 0.7 |
| Kersting | Natarajan | 0.8 |
| Luc | Paol | 0.1 |
| … | … | … |
| **Erdos** | **Straus** | **λ** |

*Open-world query evaluation*

# UCQ / Monotone CNF

- Lower bound = closed-world probability
- Upper bound = probability after **adding all tuples** with probability λ

- Polynomial time ☺
- Quadratic blow-up ☹
- 200 exabytes … again ☹

# Closed-World Lifted Query Eval

$\boxed{Q = \exists x\, \exists y\, \mathrm{Scientist}(x) \wedge \mathrm{Coauthor}(x,y)}$

Decomposable ∃-Rule

$$P(Q) = 1 - \Pi_{A \in \mathrm{Domain}} (1 - P(\mathrm{Scientist}(A) \wedge \exists y\, \mathrm{Coauthor}(A,y))$$

Check independence:

$\mathrm{Scientist}(A) \wedge \exists y\, \mathrm{Coauthor}(A,y)$

$\mathrm{Scientist}(B) \wedge \exists y\, \mathrm{Coauthor}(B,y)$

$$\begin{aligned}
= 1 - &(1 - P(\mathrm{Scientist}(A) \wedge \exists y\, \mathrm{Coauthor}(A,y)) \\
&\times (1 - P(\mathrm{Scientist}(B) \wedge \exists y\, \mathrm{Coauthor}(B,y)) \\
&\times (1 - P(\mathrm{Scientist}(C) \wedge \exists y\, \mathrm{Coauthor}(C,y)) \\
&\times (1 - P(\mathrm{Scientist}(D) \wedge \exists y\, \mathrm{Coauthor}(D,y)) \\
&\times (1 - P(\mathrm{Scientist}(E) \wedge \exists y\, \mathrm{Coauthor}(E,y)) \\
&\times (1 - P(\mathrm{Scientist}(F) \wedge \exists y\, \mathrm{Coauthor}(F,y)) \\
&\quad \ldots
\end{aligned}$$

Complexity PTIME

# Closed-World Lifted Query Eval

$Q = \exists x\ \exists y\ \text{Scientist}(x) \wedge \text{Coauthor}(x,y)$

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

$$\begin{aligned}
= 1 - &(1 - P(\text{Scientist}(A) \wedge \exists y\ \text{Coauthor}(A,y)) \\
&\times (1 - P(\text{Scientist}(B) \wedge \exists y\ \text{Coauthor}(B,y)) \\
&\times (1 - P(\text{Scientist}(C) \wedge \exists y\ \text{Coauthor}(C,y)) \\
&\times (1 - P(\text{Scientist}(D) \wedge \exists y\ \text{Coauthor}(D,y)) \\
&\times (1 - P(\text{Scientist}(E) \wedge \exists y\ \text{Coauthor}(E,y)) \\
&\times (1 - P(\text{Scientist}(F) \wedge \exists y\ \text{Coauthor}(F,y)) \\
&\qquad \ldots
\end{aligned}$$

⇨ No supporting facts in database!

⇨ Probability 0 in closed world

⇨ Ignore these sub-queries!

**Complexity linear time!**

# Open-World Lifted Query Eval

$$\boxed{Q = \exists x\, \exists y\, \text{Scientist}(x) \wedge \text{Coauthor}(x,y)}$$

$$P(Q) = 1 - \Pi_{A \in \text{Domain}} (1 - P(\text{Scientist}(A) \wedge \exists y\, \text{Coauthor}(A,y))$$

$$\begin{aligned}
= 1 - (1 - P(&\text{Scientist}(A) \wedge \exists y\, \text{Coauthor}(A,y)) \\
\times (1 - P(&\text{Scientist}(B) \wedge \exists y\, \text{Coauthor}(B,y)) \\
\times (1 - P(&\text{Scientist}(C) \wedge \exists y\, \text{Coauthor}(C,y)) \\
\times (1 - P(&\text{Scientist}(D) \wedge \exists y\, \text{Coauthor}(D,y)) \\
\times (1 - P(&\text{Scientist}(E) \wedge \exists y\, \text{Coauthor}(E,y)) \\
\times (1 - P(&\text{Scientist}(F) \wedge \exists y\, \text{Coauthor}(F,y)) \\
&\ldots
\end{aligned}$$

⇒ No supporting facts in database!

⇒ Probability $\lambda$ in open world

**Complexity PTIME!**

# Open-World Lifted Query Eval

$Q = \exists x\, \exists y\, \text{Scientist}(x) \wedge \text{Coauthor}(x,y)$

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

$$\begin{aligned}
= 1 - &(1 - P(\text{Scientist}(A) \wedge \exists y\, \text{Coauthor}(A,y)) \\
&\times (1 - P(\text{Scientist}(B) \wedge \exists y\, \text{Coauthor}(B,y)) \\
&\times (1 - P(\text{Scientist}(C) \wedge \exists y\, \text{Coauthor}(C,y)) \\
&\times (1 - P(\text{Scientist}(D) \wedge \exists y\, \text{Coauthor}(D,y)) \\
&\times (1 - P(\text{Scientist}(E) \wedge \exists y\, \text{Coauthor}(E,y)) \\
&\times (1 - P(\text{Scientist}(F) \wedge \exists y\, \text{Coauthor}(F,y)) \\
&\quad \ldots
\end{aligned}$$

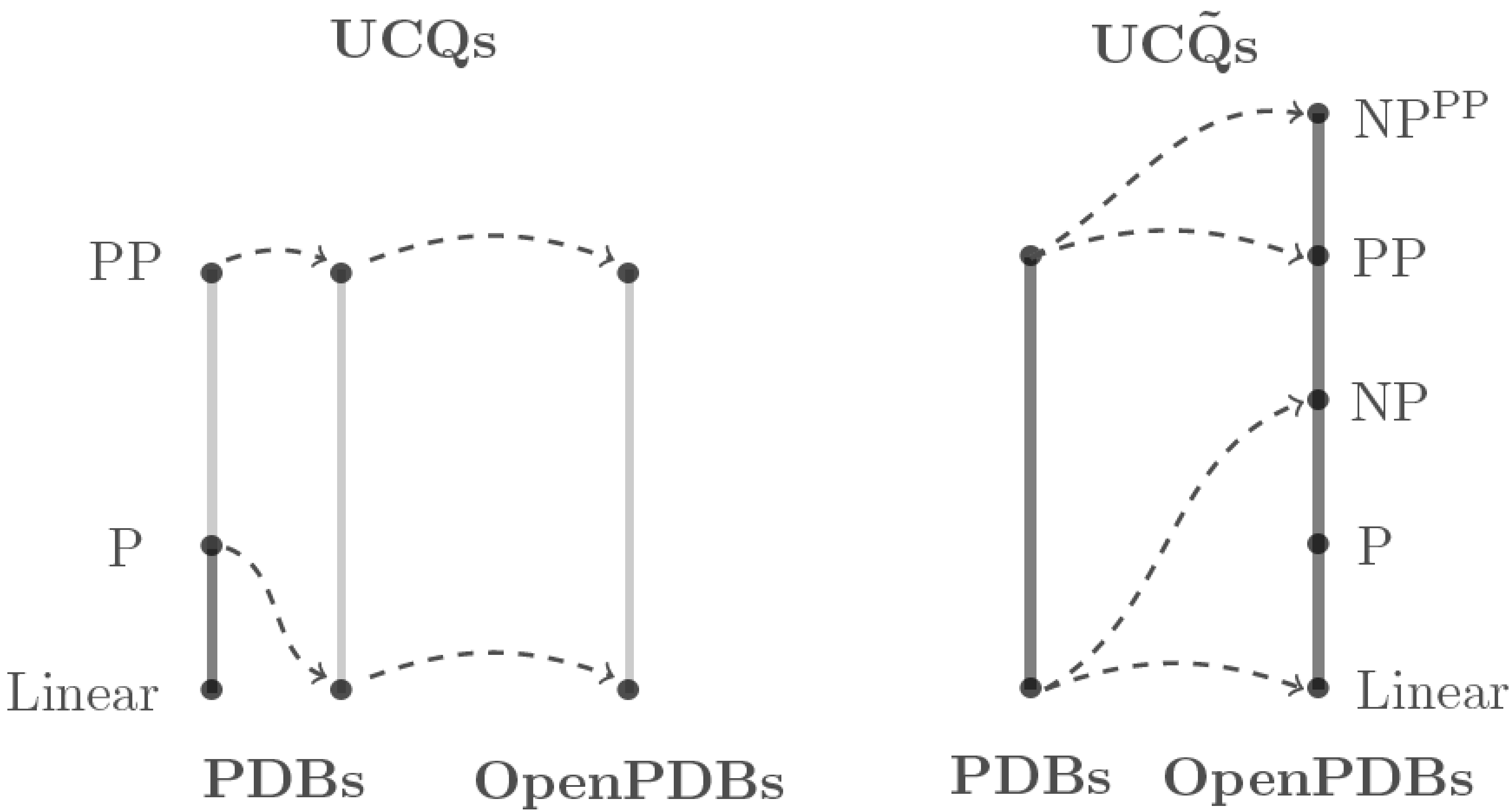
⇨ No supporting facts in database!

⇨ Probability $p$ in closed world

⇨ All together, probability $(1-p)^k$
Exploit symmetry
Lifted inference

**Complexity linear time!**

# Complexity Results




$\mathrm{Linear} \subseteq \mathrm{P} \subseteq \mathrm{NP} \subseteq \mathrm{PP} \subseteq \mathrm{P}^{\mathrm{PP}} \subseteq \mathrm{NP}^{\mathrm{PP}} \subseteq \mathrm{PSpace} \subseteq \mathrm{ExpTime}$

[Ceylan'16]

# Implement PDB Query in SQL

$Q = \exists x\ P(x) \wedge Q(x)$

- 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

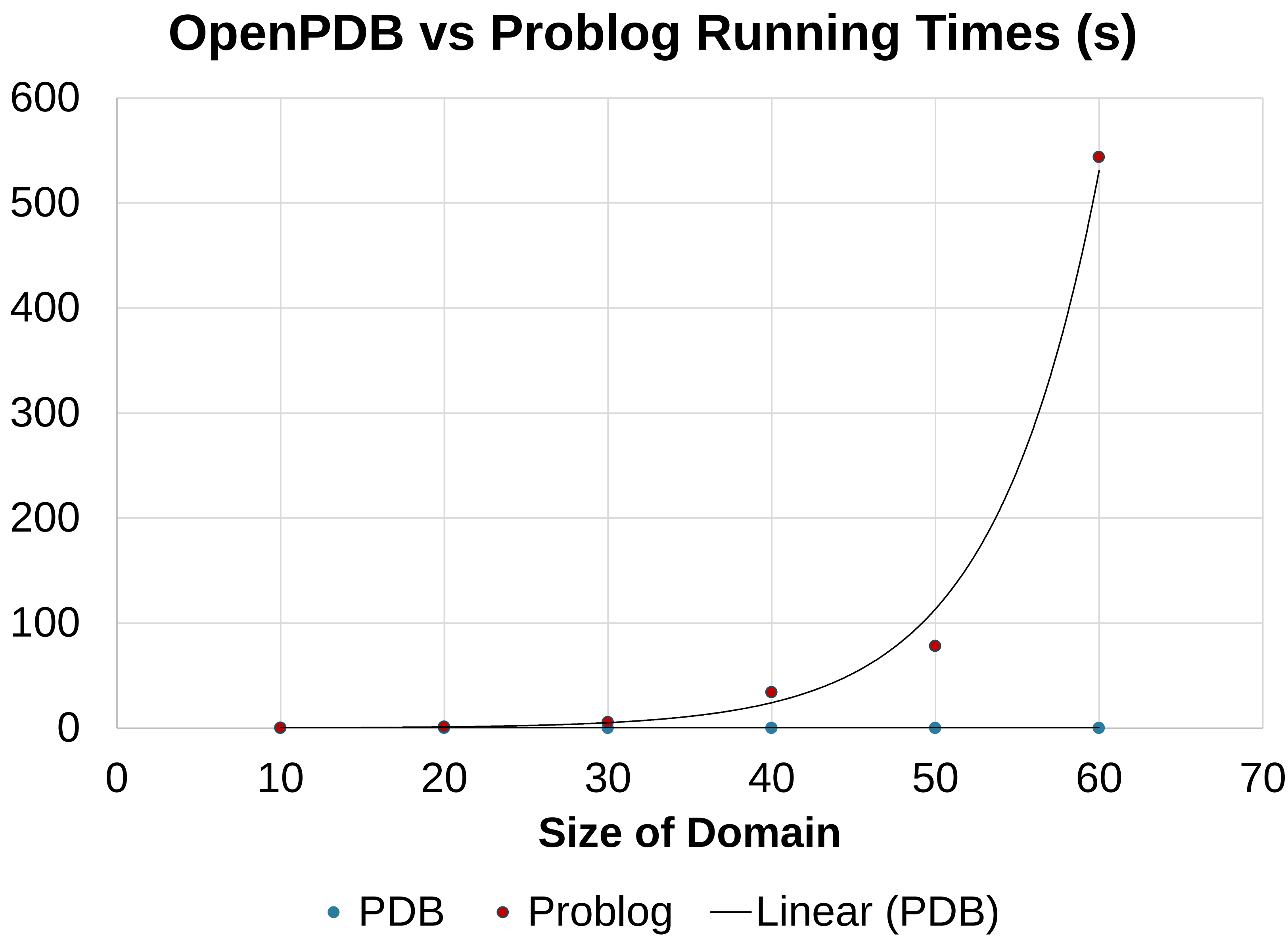
```
SELECT q9.c5,
  COALESCE(q9.pUse,λ)*COALESCE(q10.pUse,λ) AS pUse
FROM
  SQL(Q(X)) OUTER JOIN SQL(P(X))
```

```
SELECT Q.v0 AS c5,
       p AS pUse
        FROM Q
```

- Run as single PostgreSQL query!

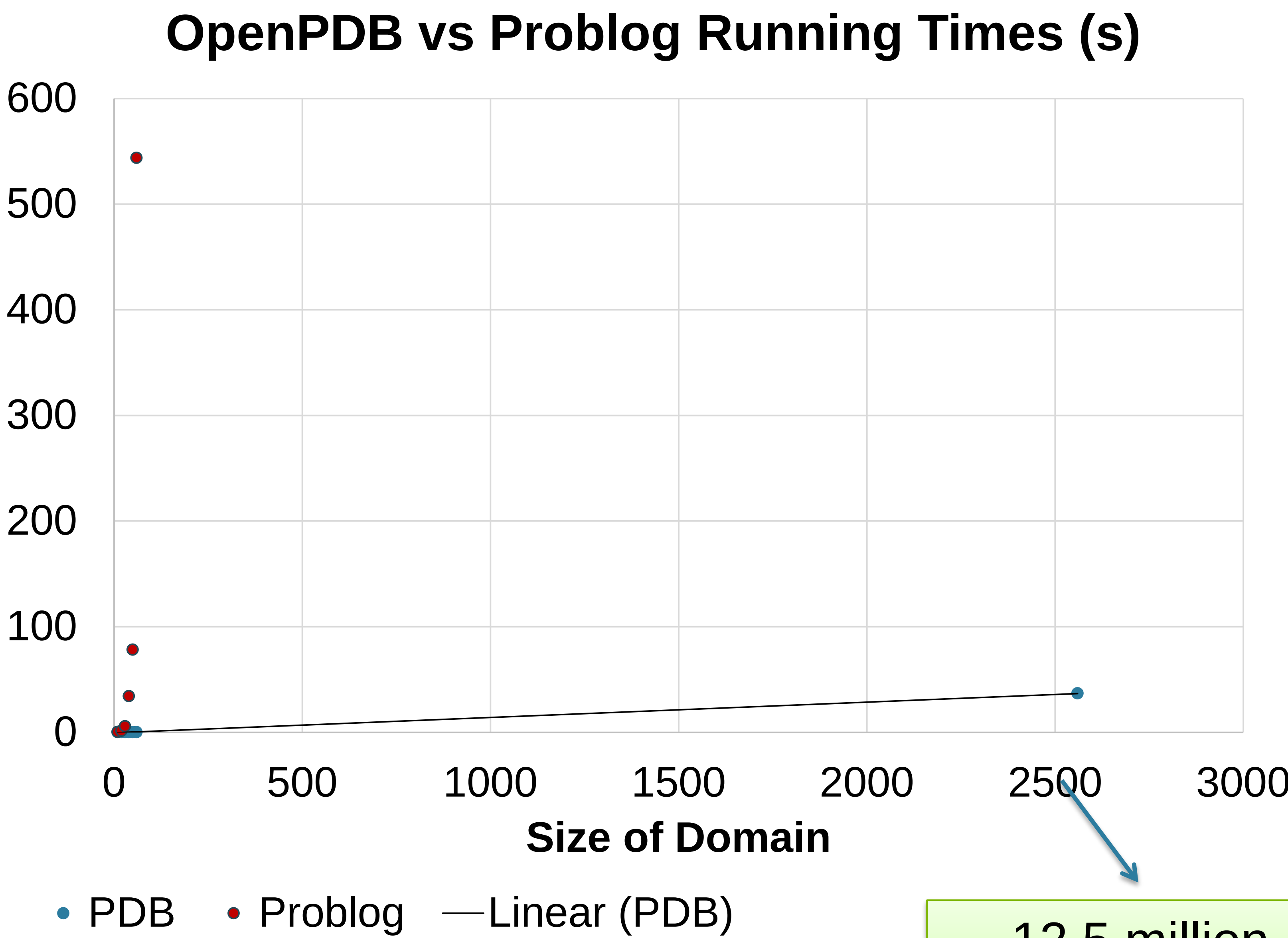
[Tal Friedman, Eric Gribkoff]

[Tal Friedman]



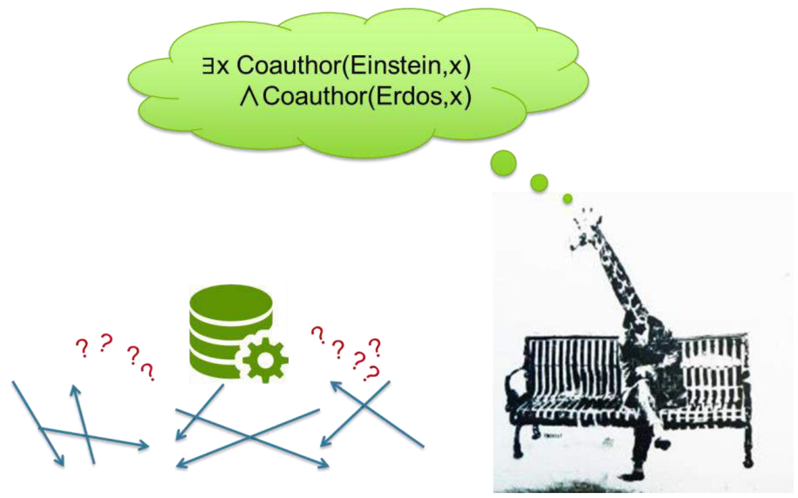

*Out of memory trying to run the ProbLog query with 70 constants in domain*

[Tal Friedman]




12.5 million random variables!

*Querying*
*Relational Embedding Models*



# Ongoing Work

- Run query evaluation *in vector space*?
  - DistMult model on Wordnet18RR
  - Query most likely answers to:

$Q(h) = R(h,\text{'c'})$ $Q(t) = R(\text{'c'},t)$

  - Solver: gradient descent on query probability

| Prediction Method | MRR | hits@5 | hits@1 |
|---|---|---|---|
| Ranking Entities | 0.305 | 0.383 | 0.234 |
| Vector Search | 0.301 | 0.377 | 0.232 |

- Fast approximate query answering

[Tal Friedman, Pasquale Minervini]

# Ongoing Work

- Join algorithms *in vector space*?
  - DistMult model on Wordnet18RR
  - Query most likely answers to:

$$Q(x) = \text{Hypernym}(\text{'OliveTree'}, x) \wedge \text{Hypernym}(x, \text{'FloweringTree'})$$

  - Answer: x = ‘SpiceTree’
- Projection (∃x)?
  - Skolemization in vector space
- More to come!

[Tal Friedman, Pasquale Minervini]

# Conclusions

- You can do much more with knowledge bases/graphs than just completing missing tuples
- Let's tear down the wall(s) between
  - statistical models for knowledge base completion and
  - query evaluation systems
- Relational probabilistic reasoning is **frontier** and **integration** of AI, KR, ML, DB, TH, etc.
- Forward pointers @AKBC:
  - Arcchit Jain, Tal Friedman, Ondrej Kuzelka, Guy Van den Broeck and Luc De Raedt. Scalable Rule Learning in Probabilistic Knowledge Bases
  - Tal Friedman and Guy Van den Broeck. On Constrained Open-World Probabilistic Databases

# QUESTIONS?