Open-World Probabilistic Databases

Guy Van den Broeck

UCLA

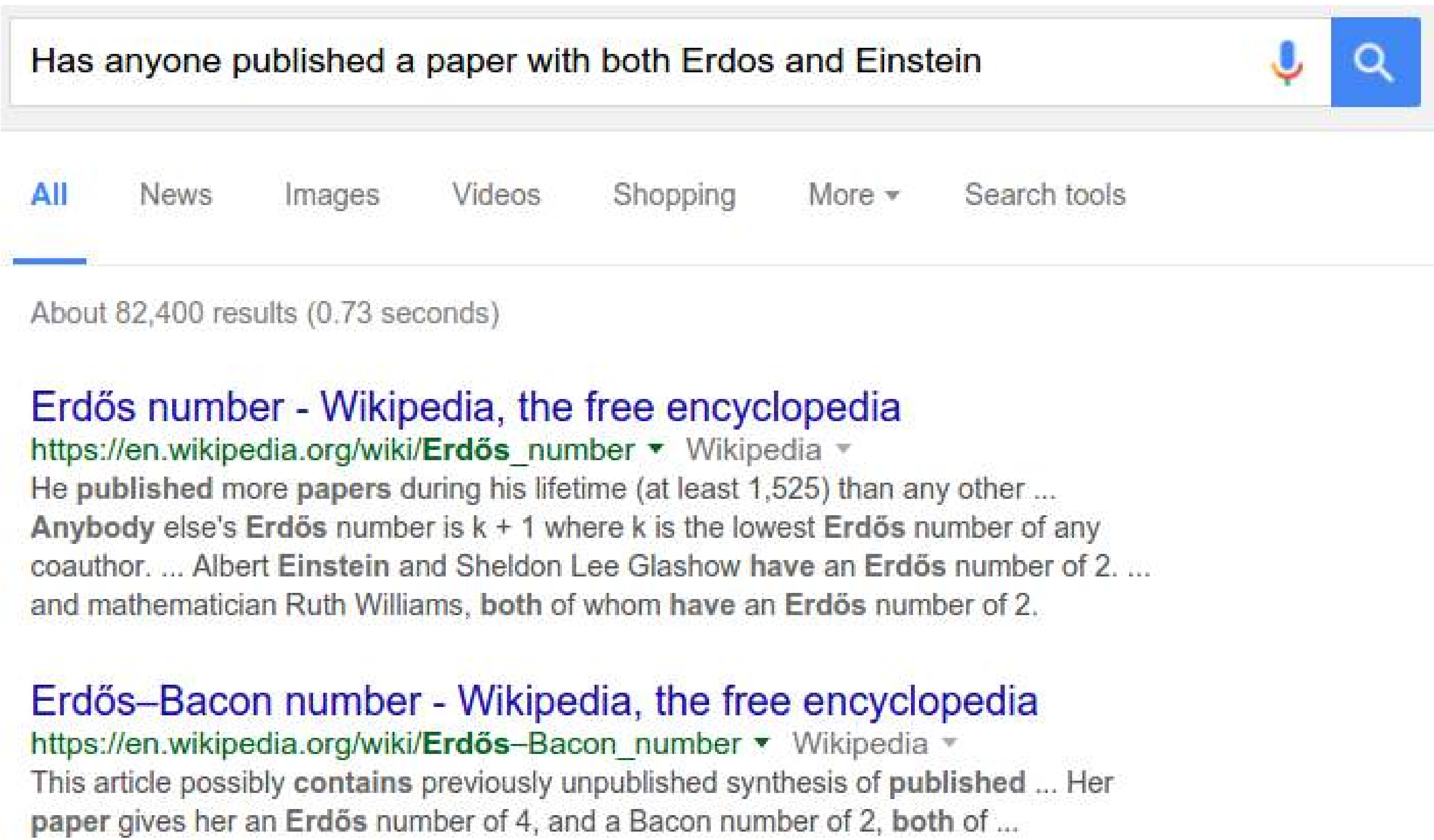
Scalable Uncertainty Management (SUM)
Sep 21, 2016

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…



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)

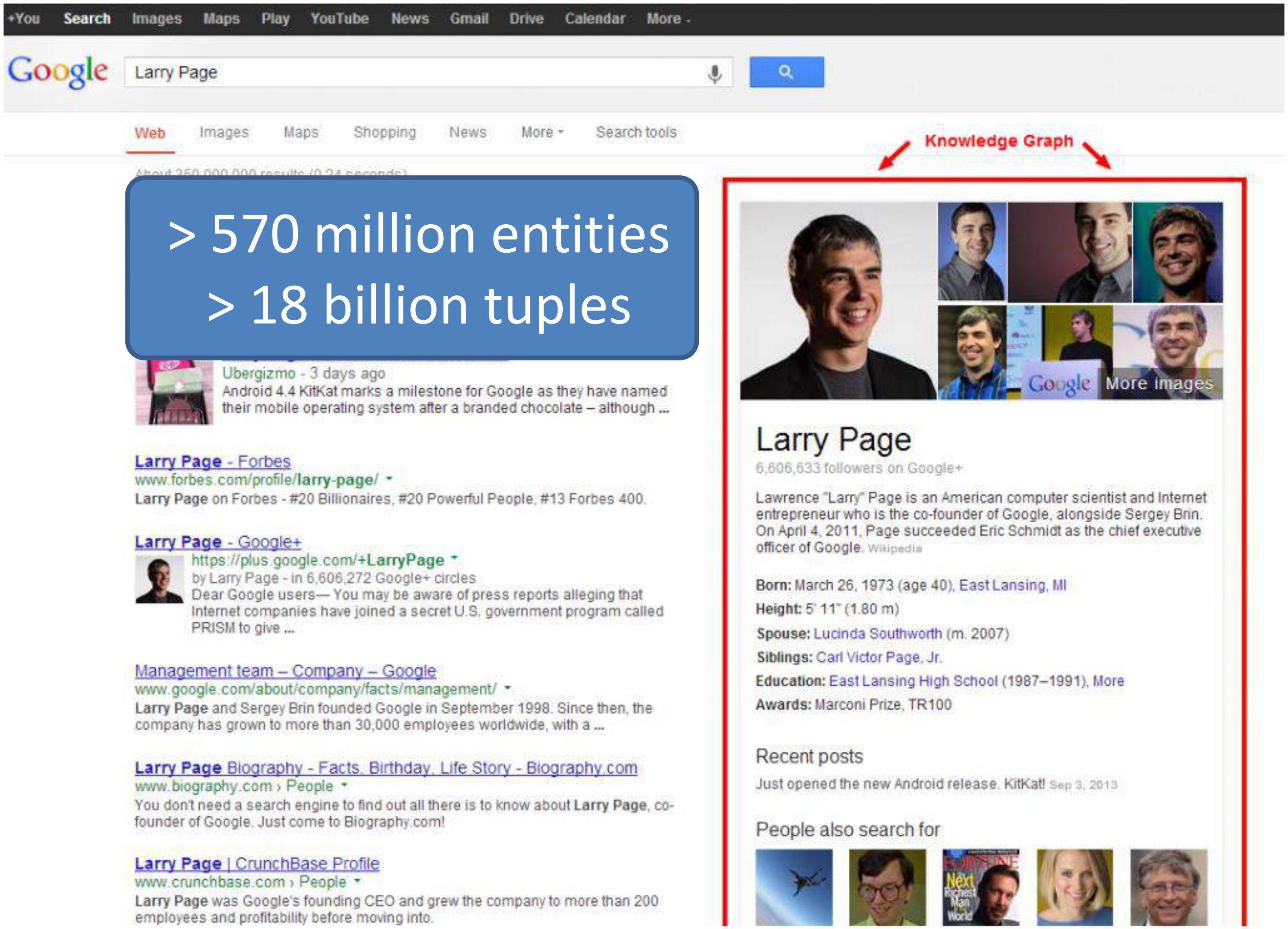
Erdős number - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/**Erdős**_number Wikipedia
He **published** more **papers** during his lifetime (at least 1,525) than any other ...
Anybody else's **Erdős** number is k + 1 where k is the lowest **Erdős** number of any
coauthor. ... Albert **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This article possibly **contains** previously unpublished synthesis of **published** ...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 ...

Google Knowledge Graph

> 570 million entities

> 18 billion tuples



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.

[Suciu'11]

Information Extraction

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

x	y	P
Luc	Laura	0.7
Luc	Hendrik	0.6
Luc	Kathleen	0.3
Luc	Paol	0.3
Luc	Paolo	0.1

Information Extraction

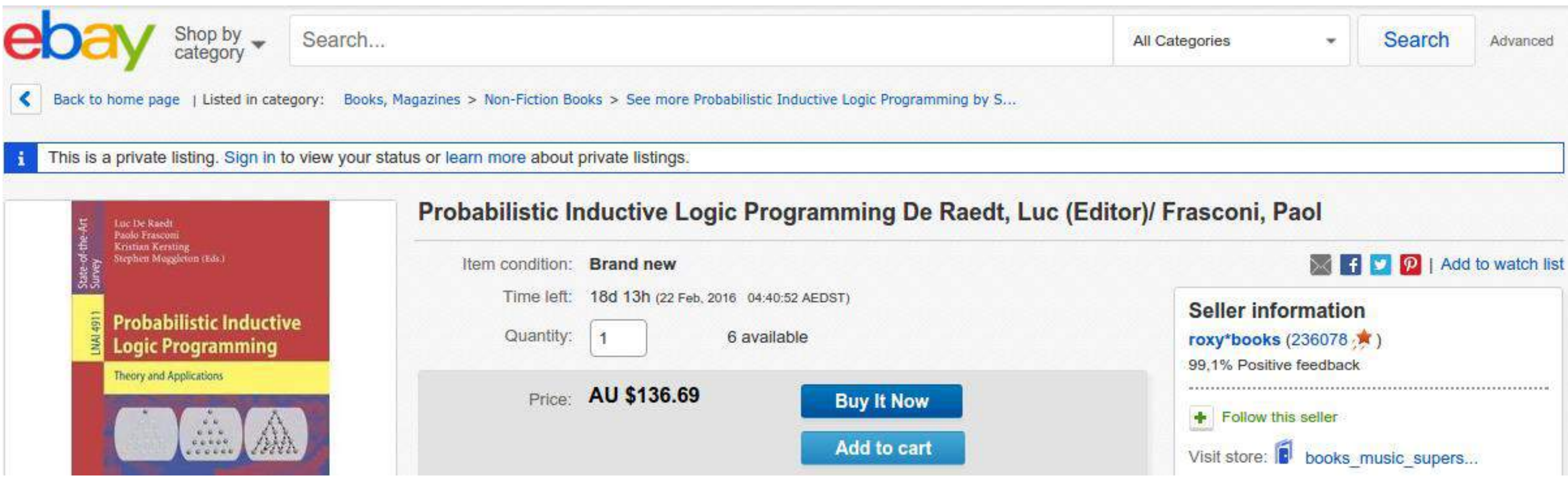
PhD Students Luc De Raedt

- 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

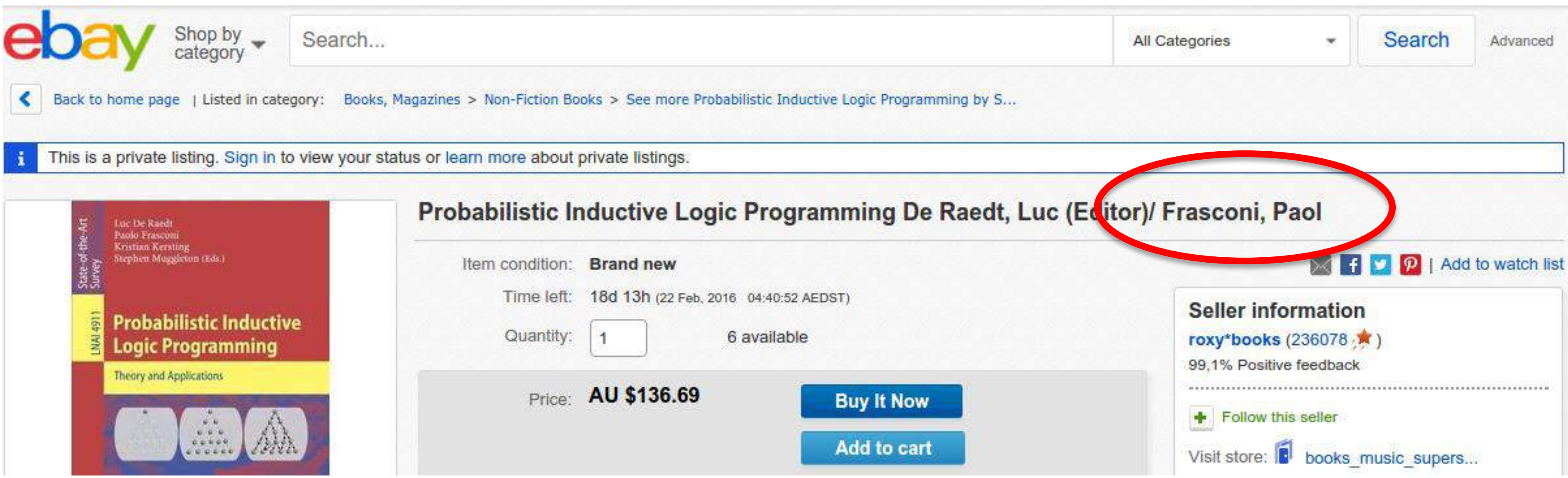
x	y	P
Luc	Laura	0.7
Luc	Hendrik	0.6
Luc	Kathleen	0.3
Luc	Paol	0.3
Luc	Paolo	0.1

Noisy!



ebay
Shop by category
Search...
All Categories
Search
Advanced
Back to home page | Listed in category: Books, Magazines > Non-Fiction Books > See more Probabilistic Inductive Logic Programming by S...
This is a private listing. Sign in to view your status or learn more about private listings.
Probabilistic Inductive Logic Programming De Raedt, Luc (Editor)/ Frasconi, Paol
Item condition: Brand new
Time left: 18d 13h (22 Feb, 2016 04:40:52 AEDST)
Quantity: 1
6 available
Price: AU $136.69
Buy It Now
Add to cart
Add to watch list
Seller information
roxy*books (236078)
99,1% Positive feedback
Follow this seller
Visit store: books_music_supers...
Luc De Raedt
Paolo Frasconi
Kristian Kersting
Stephen Muggleton (Eds.)
Probabilistic Inductive Logic Programming
Theory and Applications
State-of-the-Art Survey
LNAI 4911

Noisy!



Shop by category
Search...
All Categories
Search
Advanced
Back to home page | Listed in category: Books, Magazines > Non-Fiction Books > See more Probabilistic Inductive Logic Programming by S...
This is a private listing. Sign in to view your status or learn more about private listings.
Probabilistic Inductive Logic Programming De Raedt, Luc (Editor)/ Frasconi, Paol
Luc De Raedt
Paolo Frasconi
Kristian Kersting
Stephen Muggleton (Eds.)
State-of-the-Art Survey
LNAI 4911
Probabilistic Inductive Logic Programming
Theory and Applications
Item condition: Brand new
Time left: 18d 13h (22 Feb, 2016 04:40:52 AEDST)
Quantity: 1 6 available
Price: AU $136.69
Buy It Now
Add to cart
Add to watch list
Seller information
roxy*books (236078)
99.1% Positive feedback
Follow this seller
Visit store: books_music_supers...

Probabilistic Databases

- Relational data is increasingly probabilistic
  - NELL machine reading (>50M tuples)
  - Google Knowledge Vault (>2BN tuples)
  - DeepDive (>7M tuples)
- Next step: **Probabilistic Query Evaluation**

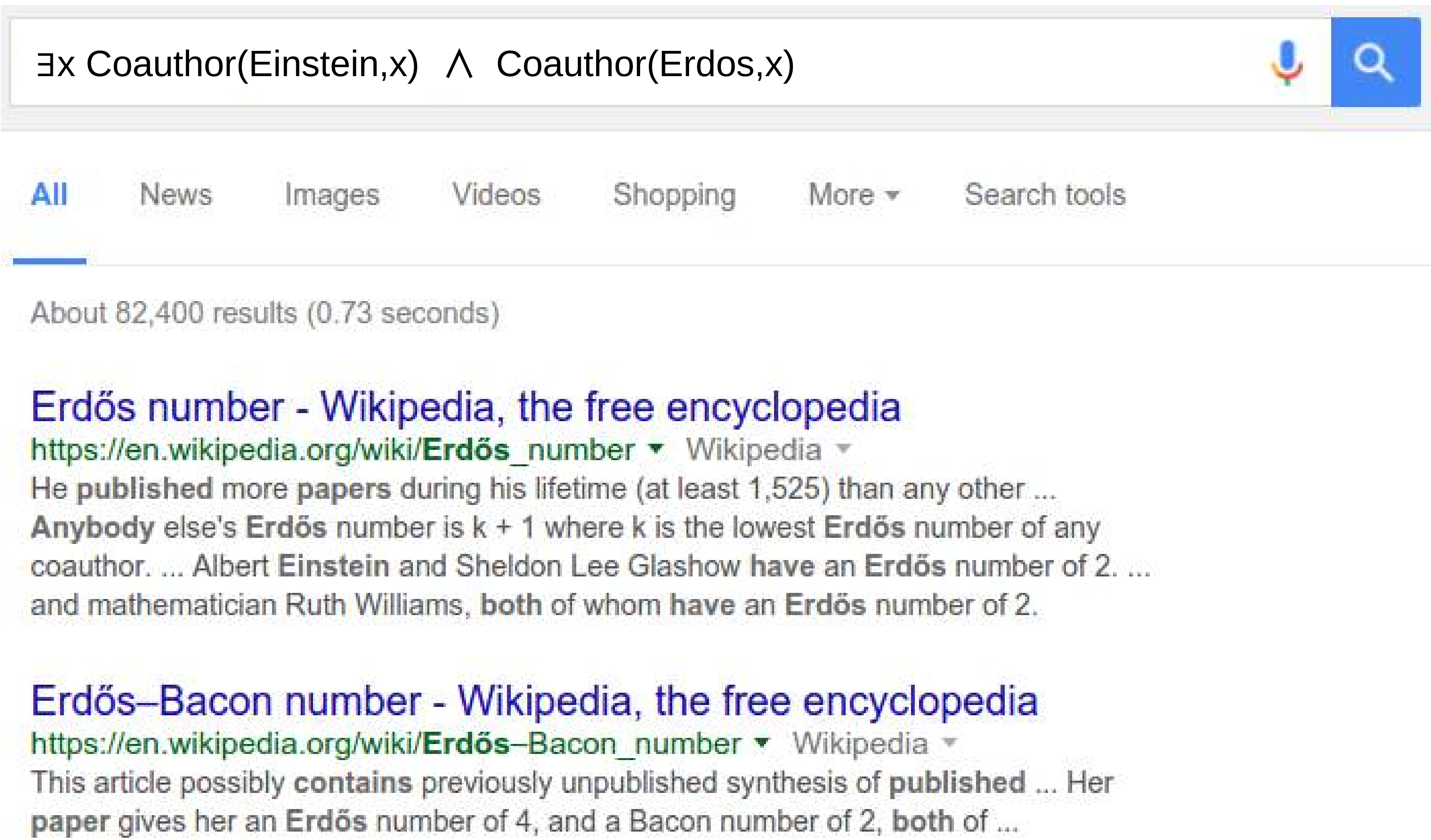
SQL or First-order logic

```
SELECT Scientist.X
FROM Scientist, Coauthor
WHERE Scientist.X = Coauthor.Y
```

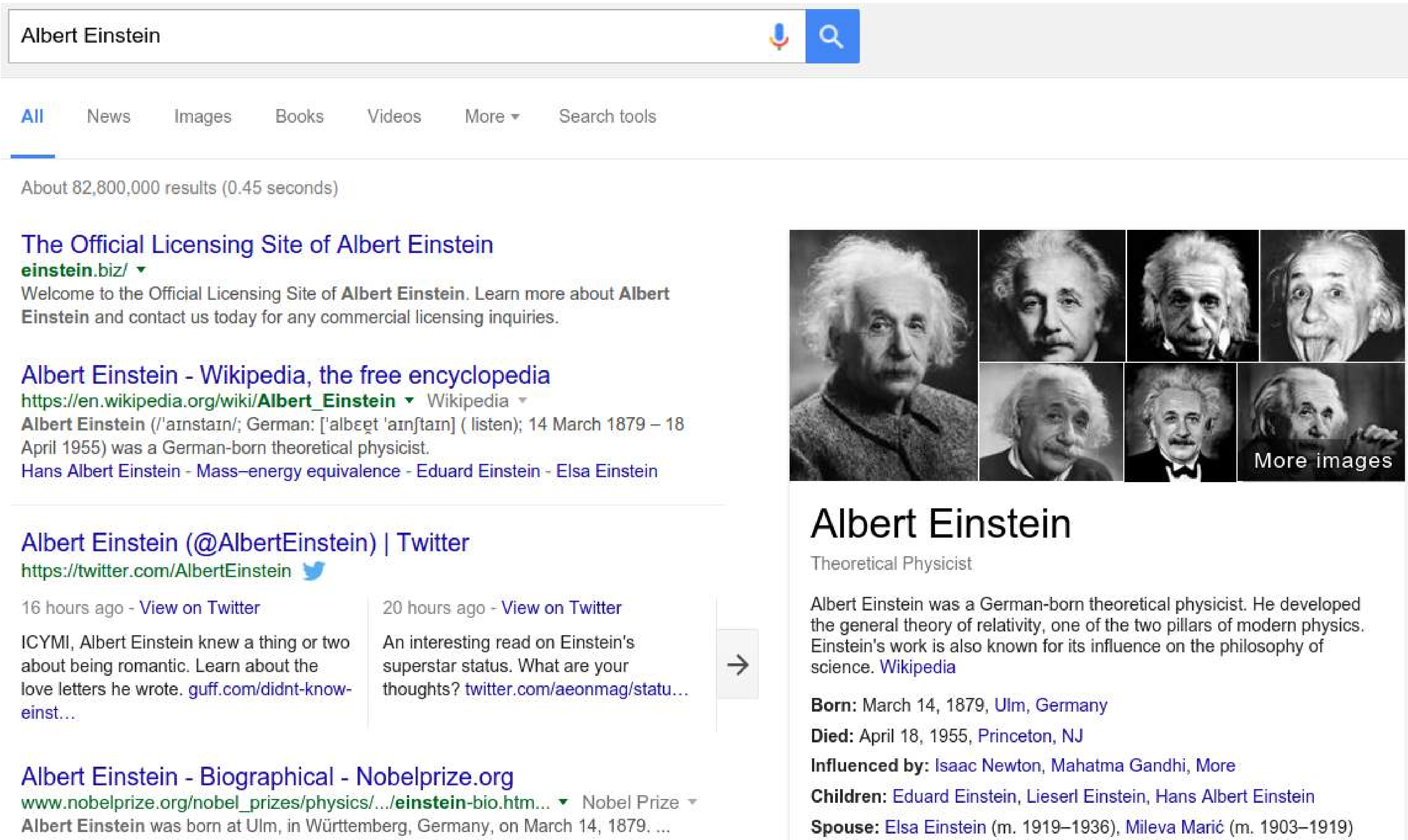
$Q(x) =$
$\exists y\ \mathrm{Scientist}(x) \land \mathrm{Coauthor}(x,y)$

[Carlson'10, Dong'14, Niu'12]

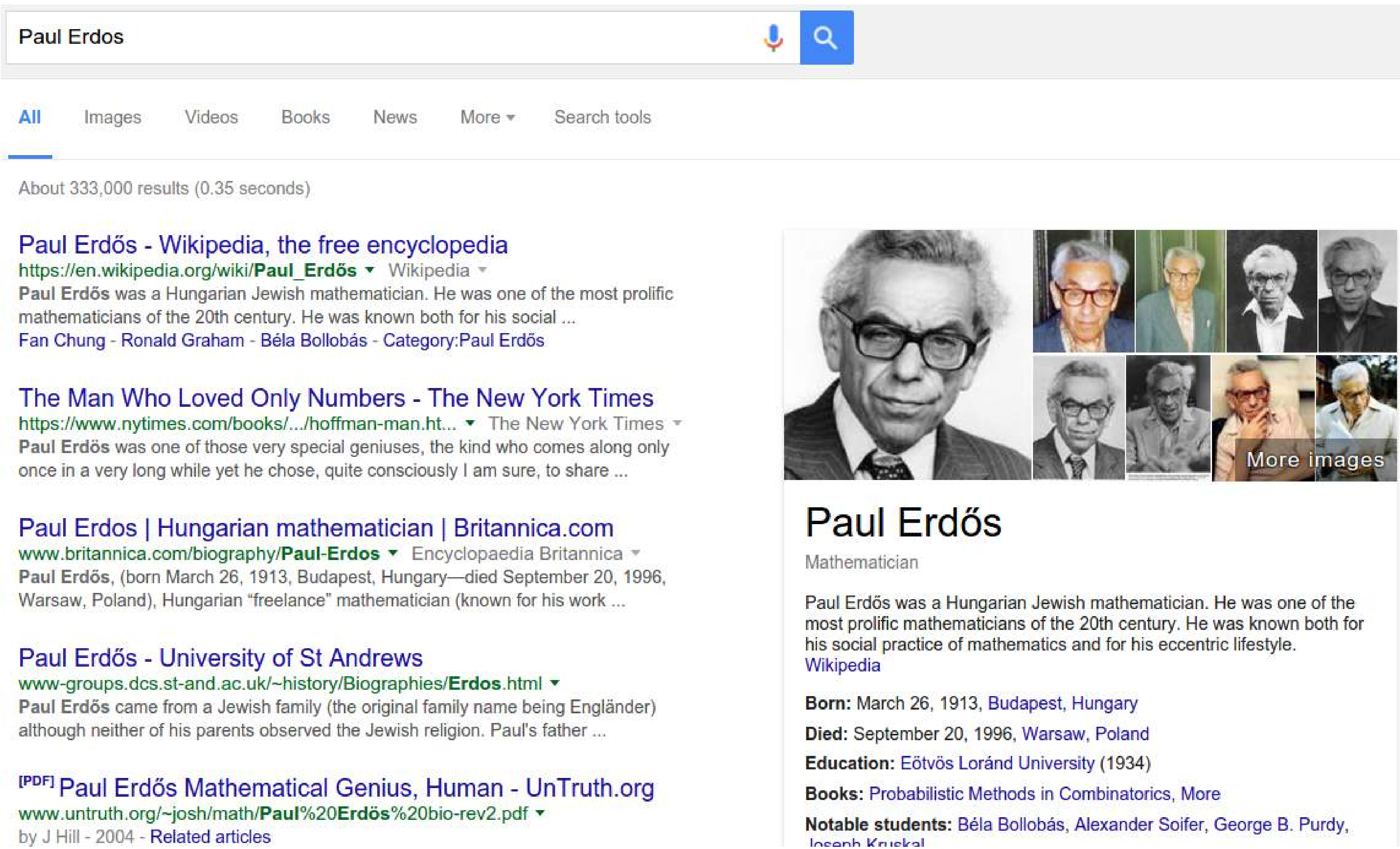
What we’d like to do…



Einstein is in the Knowledge Graph



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 Erdős 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%20Erdös%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

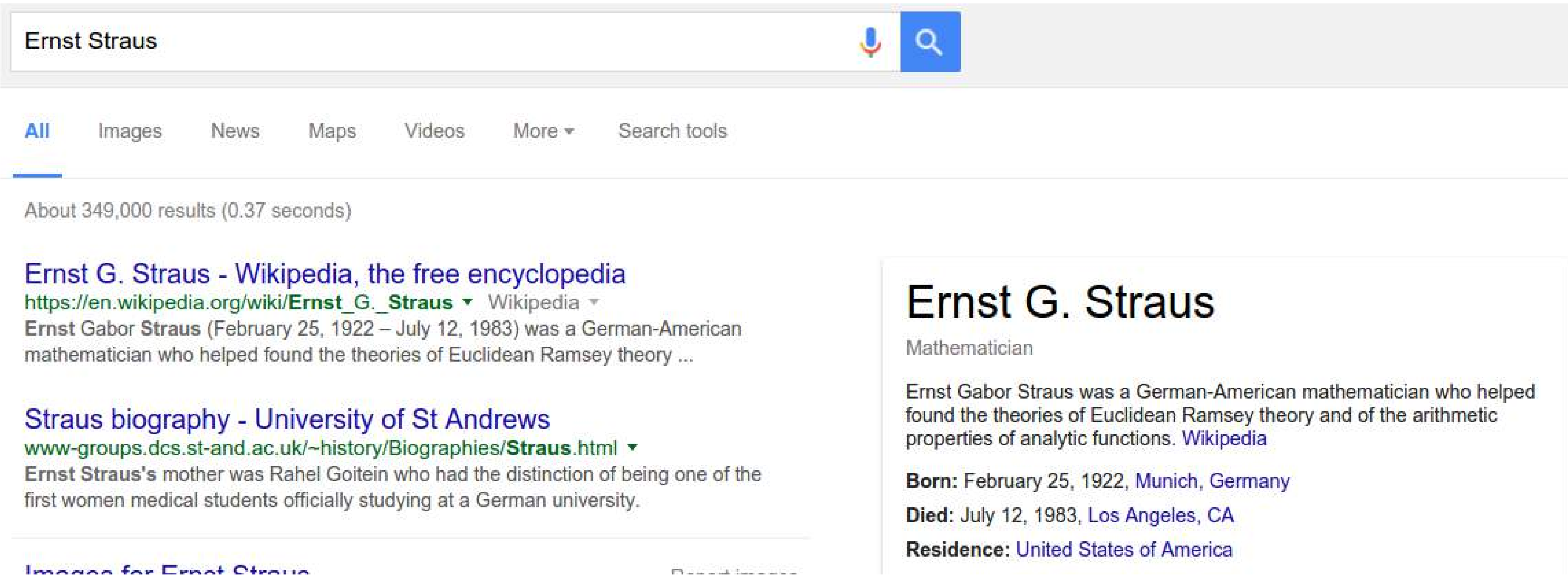
Died: September 20, 1996, Warsaw, Poland

Education: Eötvös Loránd University (1934)

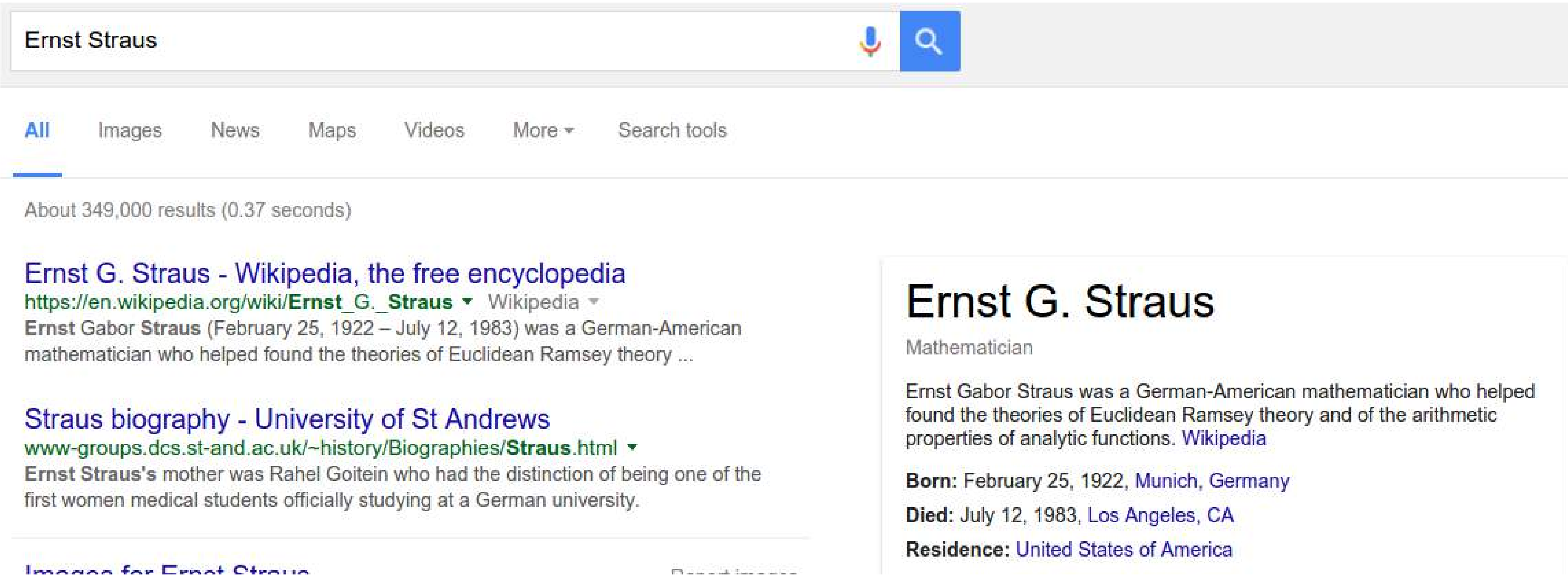
Books: Probabilistic Methods in Combinatorics, More

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

This guy is in the Knowledge Graph



This guy is in the Knowledge Graph



… and he published with both Einstein and Erdos!

Desired Query Answer

Ernst Straus

Barack Obama, …

Justin Bieber, …

Desired Query Answer

Ernst Straus

Barack Obama, …

Justin Bieber, …

1. Fuse uncertain information from web

⇒ **Embrace probability!**

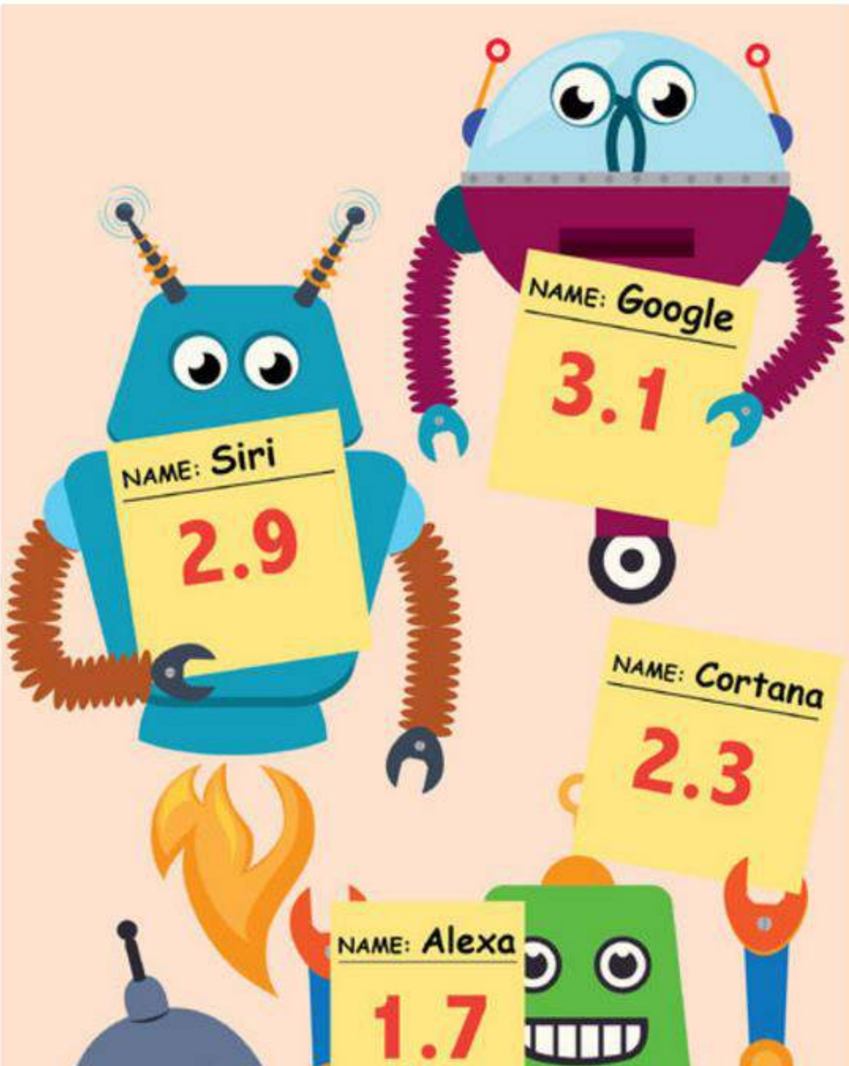
2. Cannot come from labeled data

⇒ **Embrace query eval!**

Siri, Alexa and Other Virtual Assistants Put to the Test

Tech Fix

By BRIAN X. CHEN JAN. 27, 2016



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

[Chen'16] (NYTimes)

How *probabilistic query evaluation?*

Tuple-Independent Probabilistic DB

Probabilistic database D:

Coauthor

x	y	P
A	B	p_1
A	C	p_2
B	C	p_3

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:

x	y
A	B
A	C
B	C

$p_1p_2p_3$

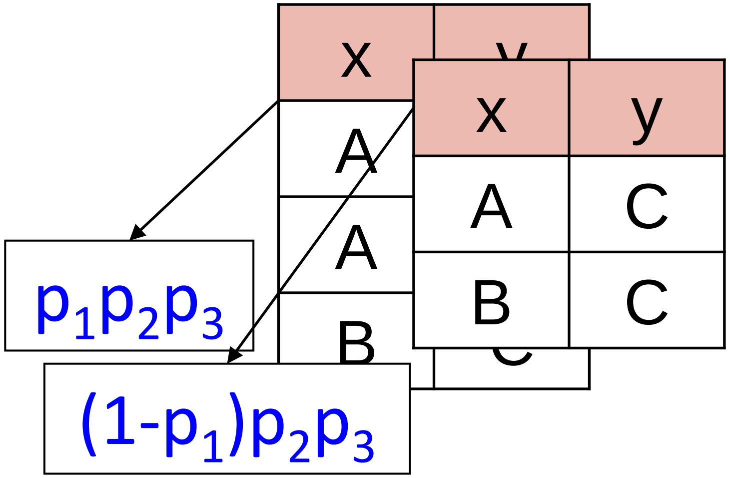
Tuple-Independent Probabilistic DB

Probabilistic database D:

Coauthor

x	y	P
A	B	p_1
A	C	p_2
B	C	p_3

Possible worlds semantics:



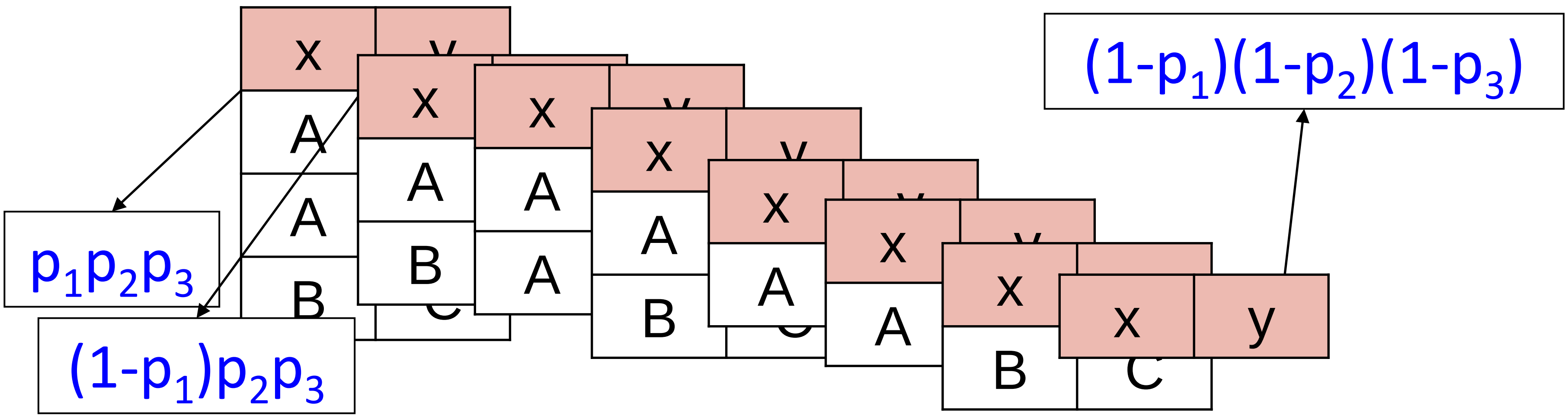
Tuple-Independent Probabilistic DB

Probabilistic database D:

Coauthor

x	y	P
A	B	p_1
A	C	p_2
B	C	p_3

Possible worlds semantics:



Probabilistic Query Evaluation

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

$P(Q) =$

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

Probabilistic Query Evaluation

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

$P(Q) = \quad 1-(1-q_1)*(1-q_2)$

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

Probabilistic Query Evaluation

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

$$P(Q) = \quad p_1 * [\ 1-(1-q_1)*(1-q_2)\ ]$$

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

Probabilistic Query Evaluation

Q = ∃x∃y Scientist(x) ∧ Coauthor(x,y)

$$P(Q) = \quad p_1 * [\ 1-(1-q_1)*(1-q_2)\ ]$$

$$1-(1-q_3)*(1-q_4)*(1-q_5)$$

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

Probabilistic Query Evaluation

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

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

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

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)

[Suciu'11]

Lifted Inference Rules

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

$\boxed{P(\neg Q) = 1 - P(Q)}$ Negation

[Suciu'11]

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 ∧,∨

[Suciu'11]

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 $\wedge$,$\vee$

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

Decomposable $\exists$,$\forall$

[Suciu'11]

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 $\wedge, \vee$

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

Decomposable $\exists, \forall$

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

Inclusion/
exclusion

[Suciu'11]

Limitations

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

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

[Suciu'11]

Limitations

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

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

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

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

[Suciu'11]

Limitations

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

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

… does not apply:

$H_0[\text{Alice}/x]$ and $H_0[\text{Bob}/x]$ 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.
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

[Dalvi and Suciu;JACM'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

[Dalvi and Suciu;JACM'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:

Coauthor

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

Learn:

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

Complete:

x	y	P
Straus	Pauli	0.504
…	…	…

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:

   $P(\texttt{Coauthor(Straus,Pauli)}) = 0$

2. Observe page

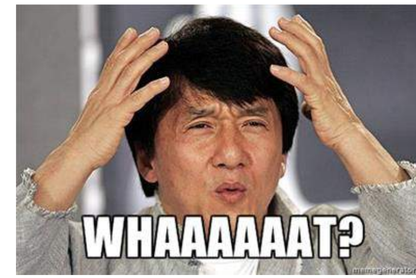
   $P(\texttt{Coauthor(Straus,Pauli} \mid$  $) = 0.2$

3. Observe page

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

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

Problem: Broken Learning Loop



Bayesian view on learning:

1. Prior belief:

   P(Coauthor(Straus,Pauli)) = 0

2. Observe page

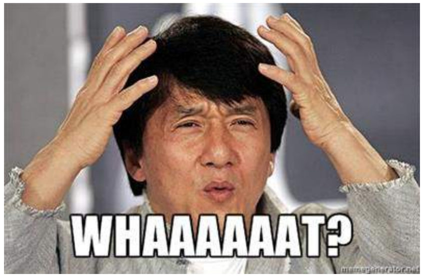
   P(Coauthor(Straus,Pauli| ) = 0.2

3. Observe page

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

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

Problem: Broken Learning Loop



Bayesian view on learning:

1. Prior belief:

   $P(\texttt{Coauthor(Straus,Pauli)}) = 0$

2. Observe page

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

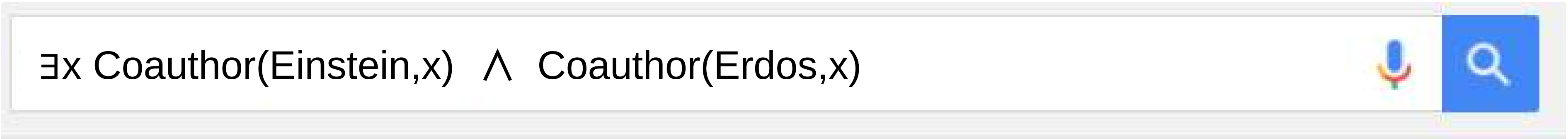
3. Observe page

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

This is mathematical nonsense!

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

What we'd like to do…



Ernst Straus

Kristian Kersting, …

Justin Bieber, …

Open World DB

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
…	…	…

- What if fact missing?
- Probability 0 for:

$$Q1 = \exists x \; \text{Coauthor}(\text{Einstein}, \mathbf{x}) \wedge \text{Coauthor}(\text{Erdos}, \mathbf{x})$$

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
…	…	…

$$Q1 = \exists x \ \text{Coauthor}(\text{Einstein},\mathbf{x}) \wedge \text{Coauthor}(\text{Erdos},\mathbf{x})$$

$$Q2 = \exists x \ \text{Coauthor}(\text{Bieber},\mathbf{x}) \wedge \text{Coauthor}(\text{Erdos},\mathbf{x})$$

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**)

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

Q2 = $\exists$x Coauthor(Bieber,**x**) $\wedge$ Coauthor(Erdos,**x**)

Q3 = Coauthor(Einstein,**Straus**) $\wedge$ Coauthor(Erdos,**Straus**)

Q4 = Coauthor(Einstein,**Bieber**) $\wedge$ Coauthor(Erdos,**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**)

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

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

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

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**)

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

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

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

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

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**)

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)

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

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
…	…	…

$Q1 = \exists x \; \text{Coauthor}(\text{Einstein},\mathbf{x}) \land \text{Coauthor}(\text{Erdos},\mathbf{x})$

$Q3 = \text{Coauthor}(\text{Einstein},\mathbf{Straus}) \land \text{Coauthor}(\text{Erdos},\mathbf{Straus})$

$Q4 = \text{Coauthor}(\text{Einstein},\mathbf{Bieber}) \land \text{Coauthor}(\text{Erdos},\mathbf{Bieber})$

$Q5 = \text{Coauthor}(\text{Einstein},\mathbf{Bieber}) \land \neg\text{Coauthor}(\mathbf{Einstein},\mathbf{Bieber})$

We know for sure that $P(Q1) \geq P(Q3)$, $P(Q1) \geq P(Q4)$

and $P(Q3) \geq P(Q5)$, $P(Q4) \geq P(Q5)$ because $P(Q5) = 0$.

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

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)

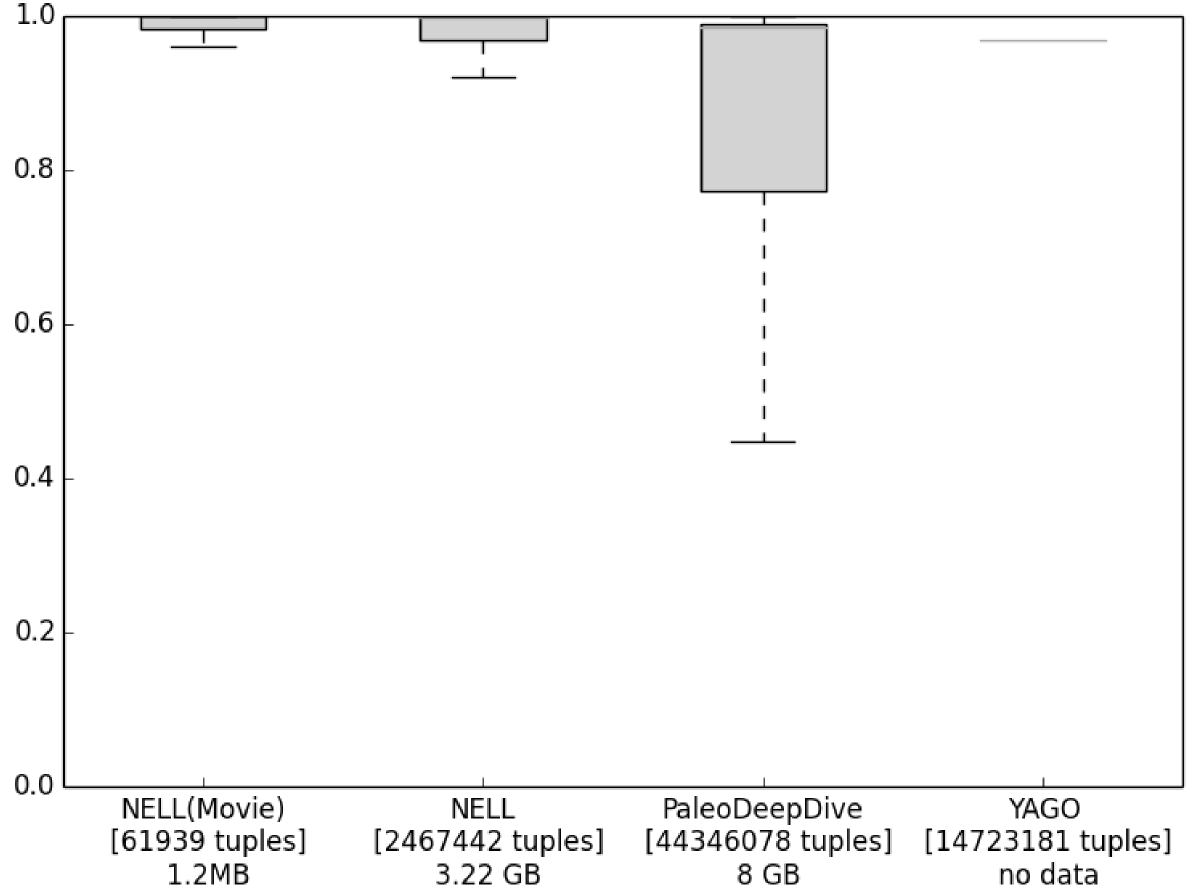
and P(Q3) ≥ P(Q5), P(Q4) ≥ P(Q5) because P(Q5) = 0.

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

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

Problem: Curse of Superlinearity

- Reality is worse!
- Tuples are intentionally missing!
- Every tuple has 99% probability



Problem: Curse of Superlinearity

"This is all true, Guy,
but it's just a temporary issue."

"No
it's not!

- *A single table (Sibling)*
- *Facebook scale (billions of people)*
- *Real (non-zero) Bayesian beliefs*

Sibling

x	y	P
…	…	…

⇒ 200 Exabytes of data"

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

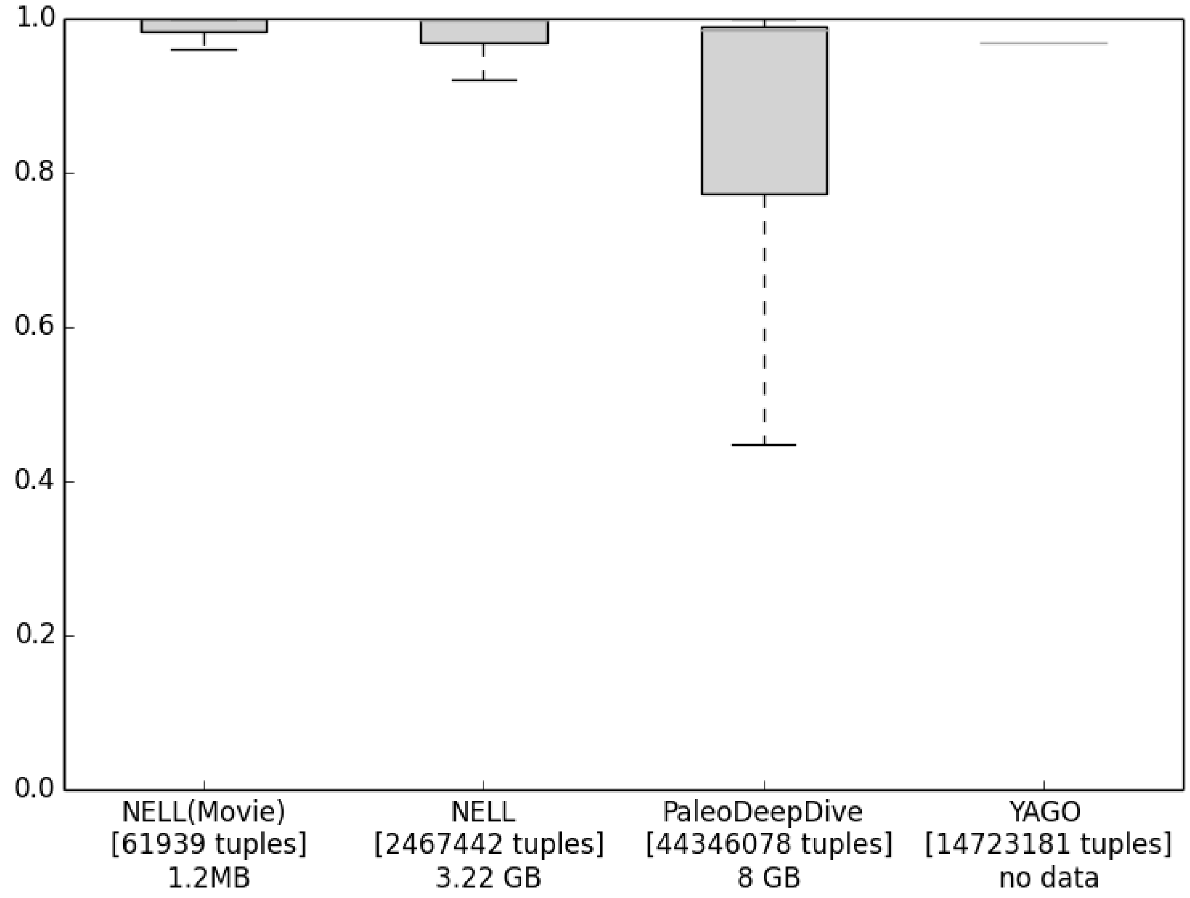
Problem: Curse of Superlinearity

All Google storage is a couple exabytes…

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

Problem: Curse of Superlinearity

We should
be here!



Problem: 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(x,z) ∧ Coauthor(z,y).

OR

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

[De Raedt et al; IJCAI'15]

Problem: 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(x,z) ∧ 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 < λ

Q2 = Coauthor(Einstein,**Straus**) ∧ Coauthor(Erdos,**Straus**)

P(Q2) ≥ 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
…	…	…

Open-World Prob. Databases

Intuition: tuples can be added with P < λ

Q2 = Coauthor(Einstein,**Straus**) ∧ Coauthor(Erdos,**Straus**)

P(Q2) ≥ 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 Prob. Databases

Intuition: tuples can be added with P < λ

Q2 = Coauthor(Einstein,**Straus**) ∧ 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**

Closed-World Prob. Databases

A PDB $\mathcal{P}$ induces a *unique probability distribution* over worlds ω:

$$\mathrm{P}_{\mathcal{P}}(\omega) = \prod_{t \in \omega} \mathrm{P}_{\mathcal{P}}(t) \prod_{t \notin \omega} (1 - \mathrm{P}_{\mathcal{P}}(t)),$$

where for every tuple t, it holds that

$$\mathrm{P}_{\mathcal{P}}(t) = \begin{cases} p & \text{if } \langle t : p \rangle \in \mathcal{P} \\ 0 & \text{otherwise. [Probabilistic CWA]} \end{cases}$$

Open-World Prob. Databases

An *OpenPDB* is a pair $\mathcal{G} = (\mathcal{P}, \lambda)$, where $\mathcal{P}$ is a PDB

$$\mathrm{P}_{\mathcal{G}}(t) = \begin{cases} p & \text{if } \langle t : p \rangle \in \mathcal{P} \\ [0, \lambda] & \text{otherwise.} \end{cases}$$

A *λ-completion* of $\mathcal{G}$ contains a tuple $\langle t : p \rangle$ for some $p \in [0, \lambda]$ for every $t \notin \mathcal{P}$. $\mathcal{G}$ induces *a set of probability distributions* $\mathrm{K}_{\mathcal{G}}$ [3]:

$$\underline{\mathrm{P}}_{\mathcal{G}}(\mathrm{Q}) = \min_{\mathrm{P} \in \mathrm{K}_{\mathcal{G}}} \mathrm{P}(\mathrm{Q}) \qquad \text{and} \qquad \overline{\mathrm{P}}_{\mathcal{G}}(\mathrm{Q}) = \max_{\mathrm{P} \in \mathrm{K}_{\mathcal{G}}} \mathrm{P}(\mathrm{Q}).$$

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

How *open-world query evaluation?*

UCQ / Monotone CNF

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

UCQ / Monotone CNF

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

- Polynomial time ☺

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

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

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

Closed-World Lifted Query Eval

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

Decomposable ∀-Rule

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

Closed-World Lifted Query Eval

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

Decomposable ∀-Rule

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

Check independence:

$Scientist(A) \wedge \exists y\, Coauthor(A,y)$

$Scientist(B) \wedge \exists y\, Coauthor(B,y)$

Closed-World Lifted Query Eval

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

Decomposable ∀-Rule

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

Check independence:

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

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

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

Closed-World Lifted Query Eval

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

Decomposable ∀-Rule

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

Check independence:

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

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

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

Complexity PTIME

Closed-World Lifted Query Eval

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

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

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

Closed-World Lifted Query Eval

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

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

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

⇒ No supporting facts in database!

Closed-World Lifted Query Eval

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

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

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

⇒ No supporting facts in database!

⇒ Probability 0 in closed world

Closed-World Lifted Query Eval

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

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

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

⇨ No supporting facts in database!

⇨ Probability 0 in closed world

⇨ Ignore these queries!

Closed-World Lifted Query Eval

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

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

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

⇨ No supporting facts in database!

⇨ Probability 0 in closed world

⇨ Ignore these queries!

Complexity linear time!

Open-World Lifted Query Eval

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

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

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

No supporting facts in database!

Open-World Lifted Query Eval

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

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

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

⇒ No supporting facts in database!

⇒ Probability p in closed world

Open-World Lifted Query Eval

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

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

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

⇨ No supporting facts in database!

⇨ Probability p in closed world

Complexity PTIME!

Open-World Lifted Query Eval

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

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

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

⇒ No supporting facts in database!

⇒ Probability p in closed world

Open-World Lifted Query Eval

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

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

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

⇒ No supporting facts in database!

⇒ Probability p in closed world

⇒ All together, probability $(1-p)^k$
Do symmetric lifted inference

Open-World Lifted Query Eval

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

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

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

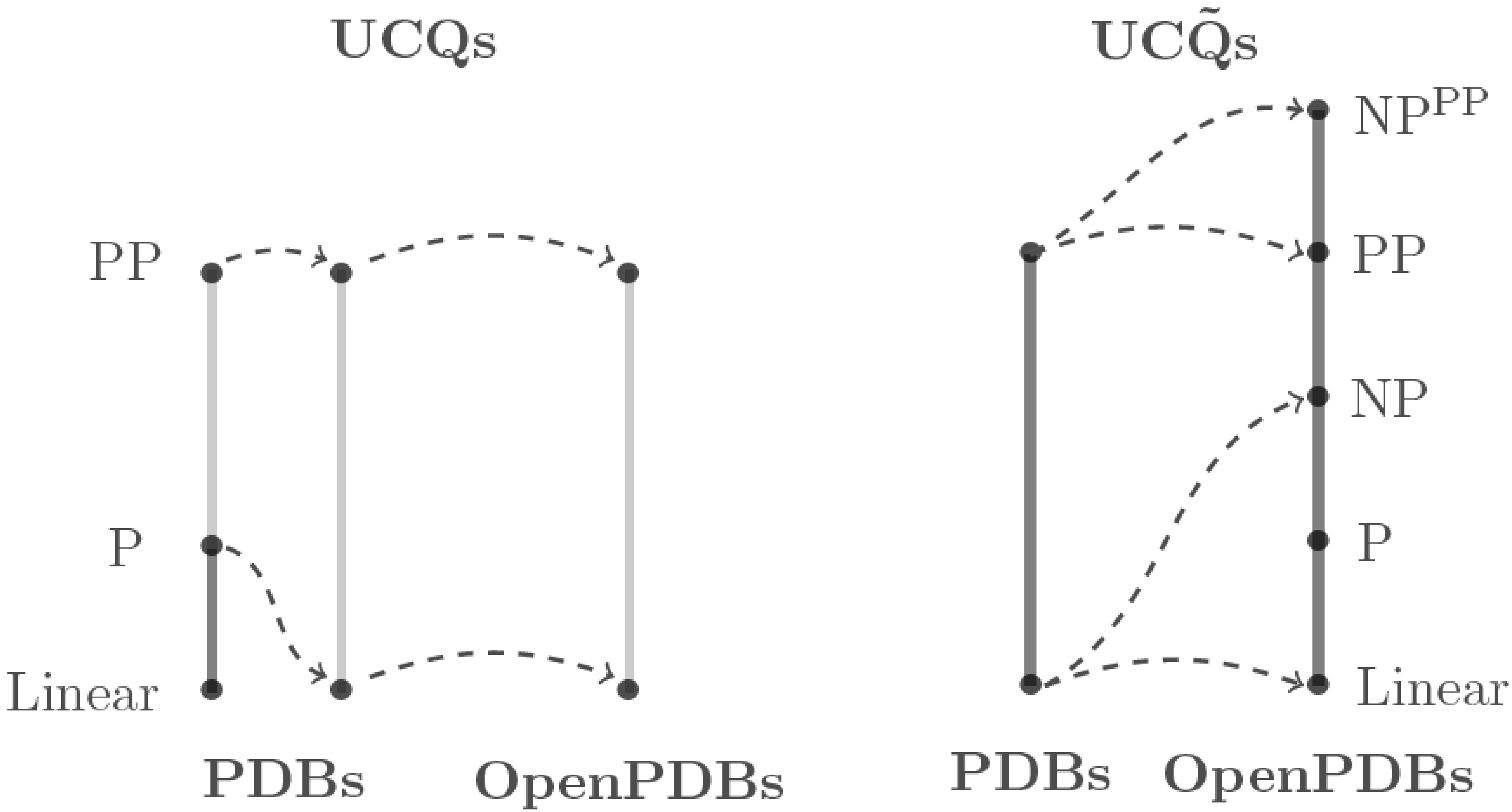
⇨ No supporting facts in database!

⇨ Probability p in closed world

⇨ All together, probability $(1-p)^k$
Do symmetric lifted inference

Complexity linear time!

Complexity Results

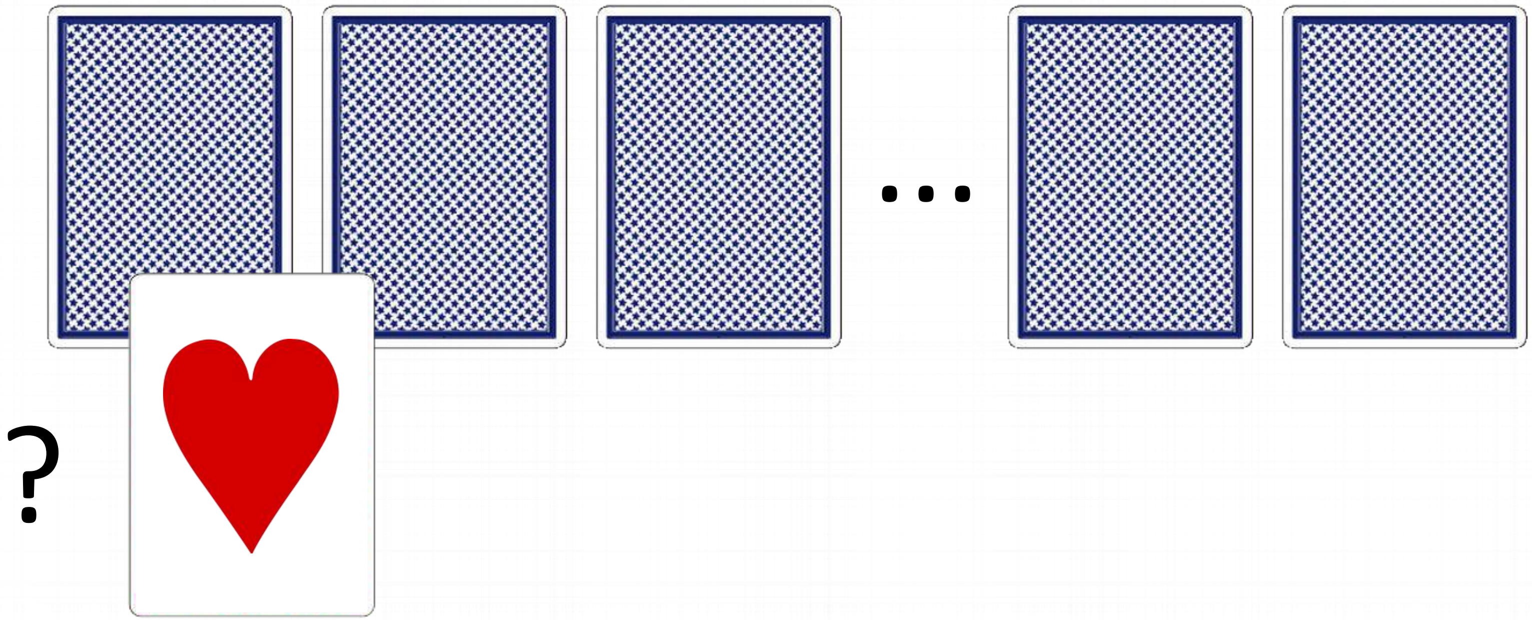


Linear ⊆ P ⊆ NP ⊆ PP ⊆ P^{PP} ⊆ NP^{PP} ⊆ PSpace ⊆ ExpTime

[Ceylan'16]

What *is the broader picture?*

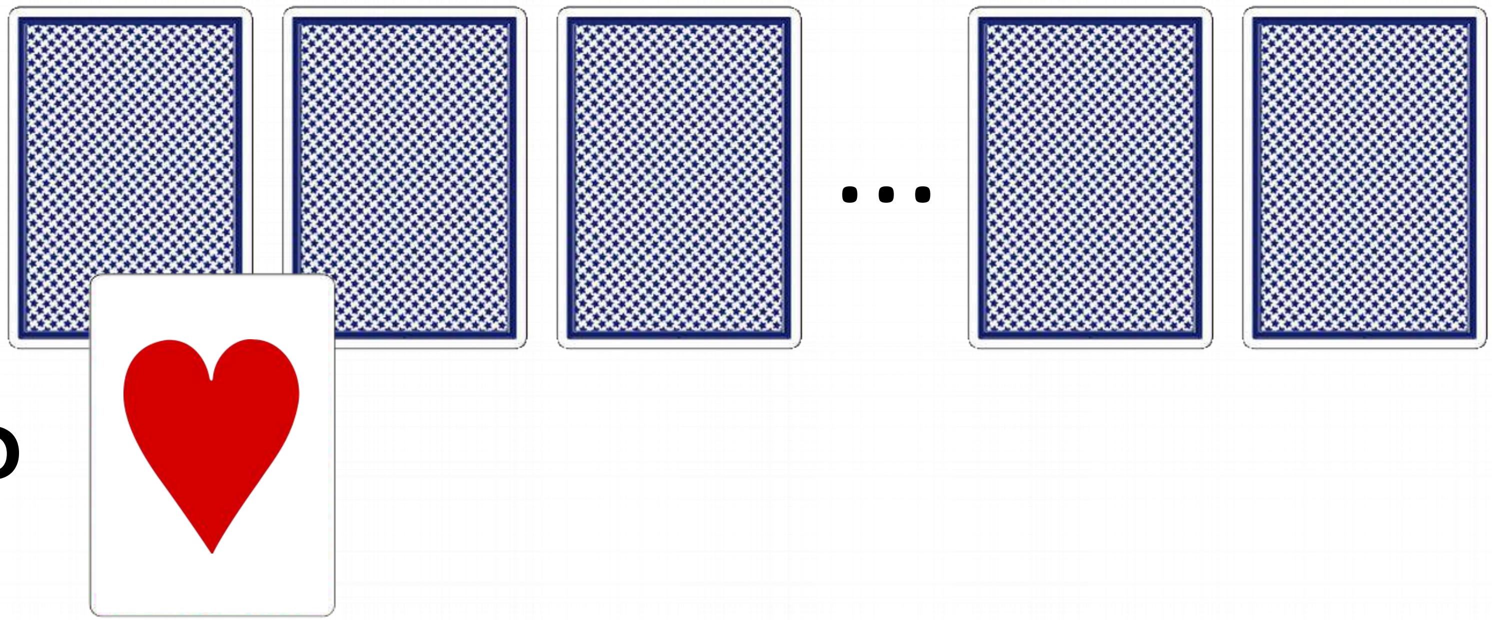
A Simple Reasoning Problem



Probability that Card1 is Hearts?

[Van den Broeck; AAAI-KRR'15]

A Simple Reasoning Problem

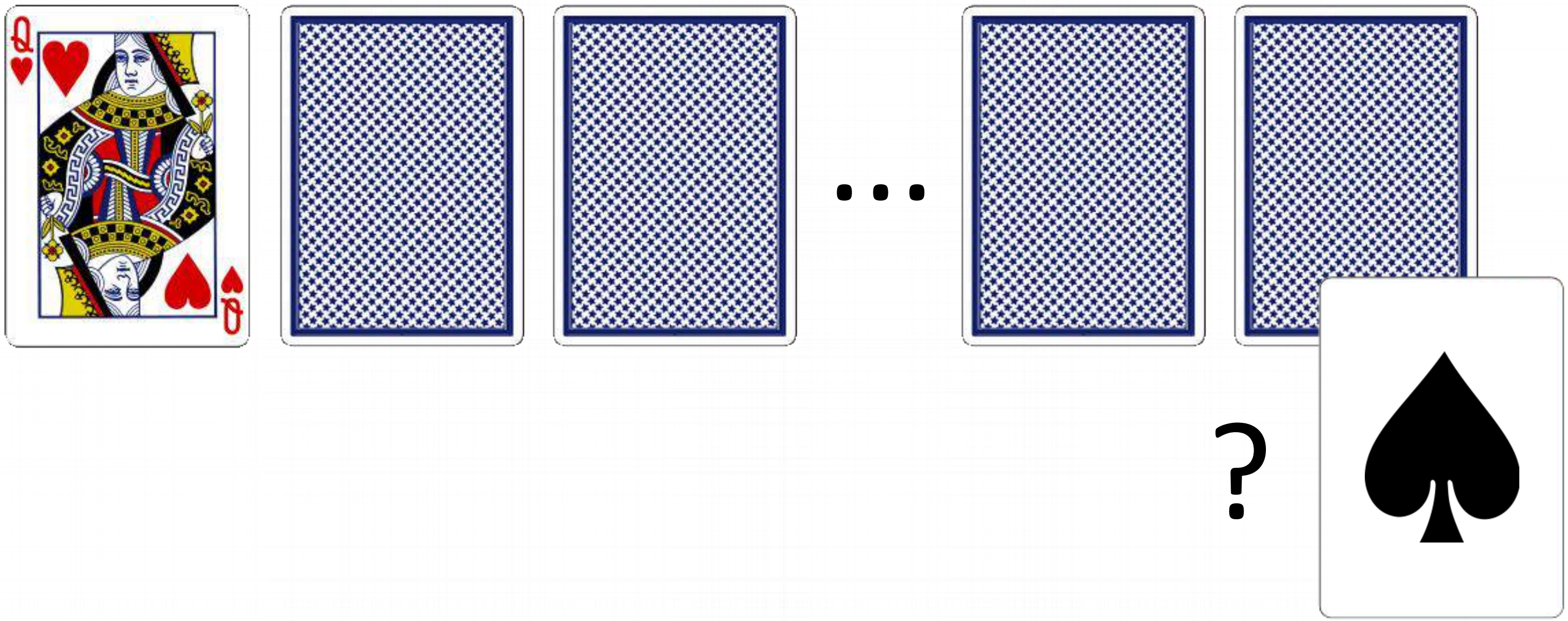


?

Probability that Card1 is Hearts? 1/4

[Van den Broeck; AAAI-KRR'15]

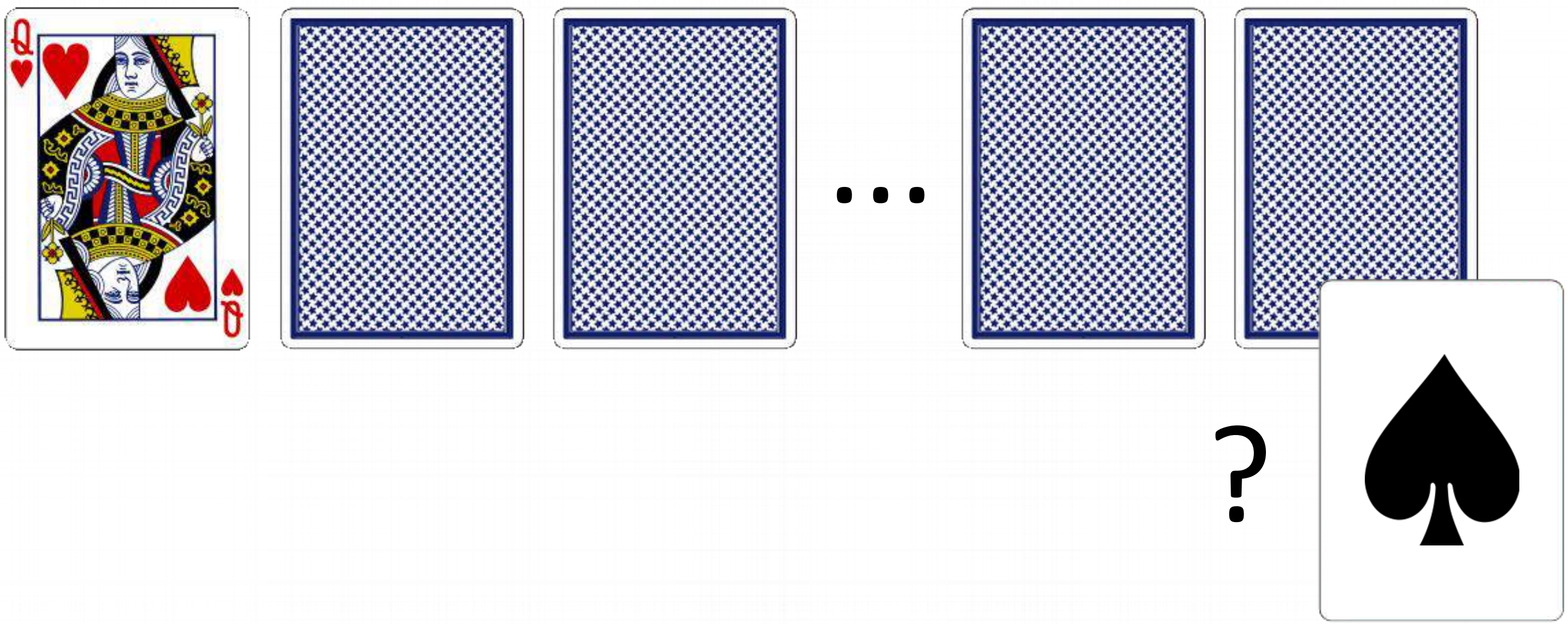
A Simple Reasoning Problem



Probability that Card52 is Spades given that Card1 is QH?

[Van den Broeck; AAAI-KRR'15]

A Simple Reasoning Problem



Probability that Card52 is Spades given that Card1 is QH?

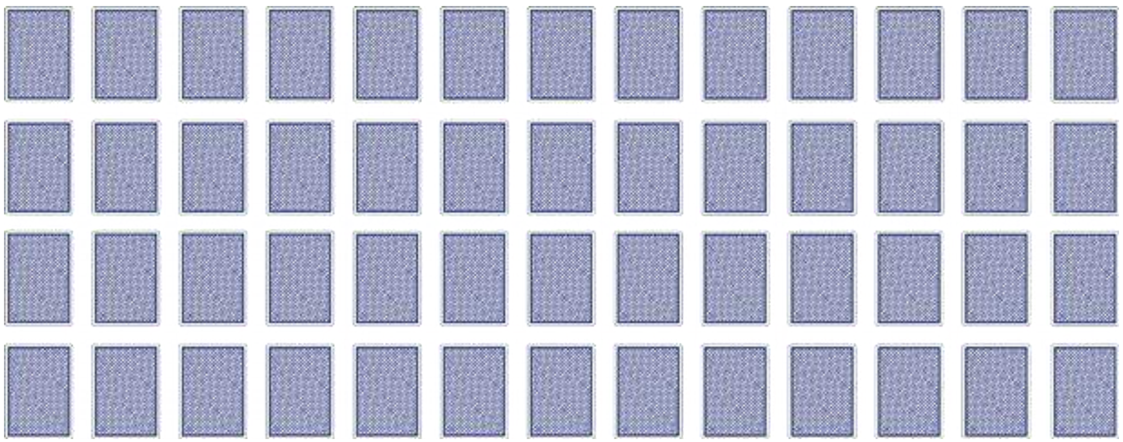
13/51

[Van den Broeck; AAAI-KRR'15]

Automated Reasoning

Let us automate this:

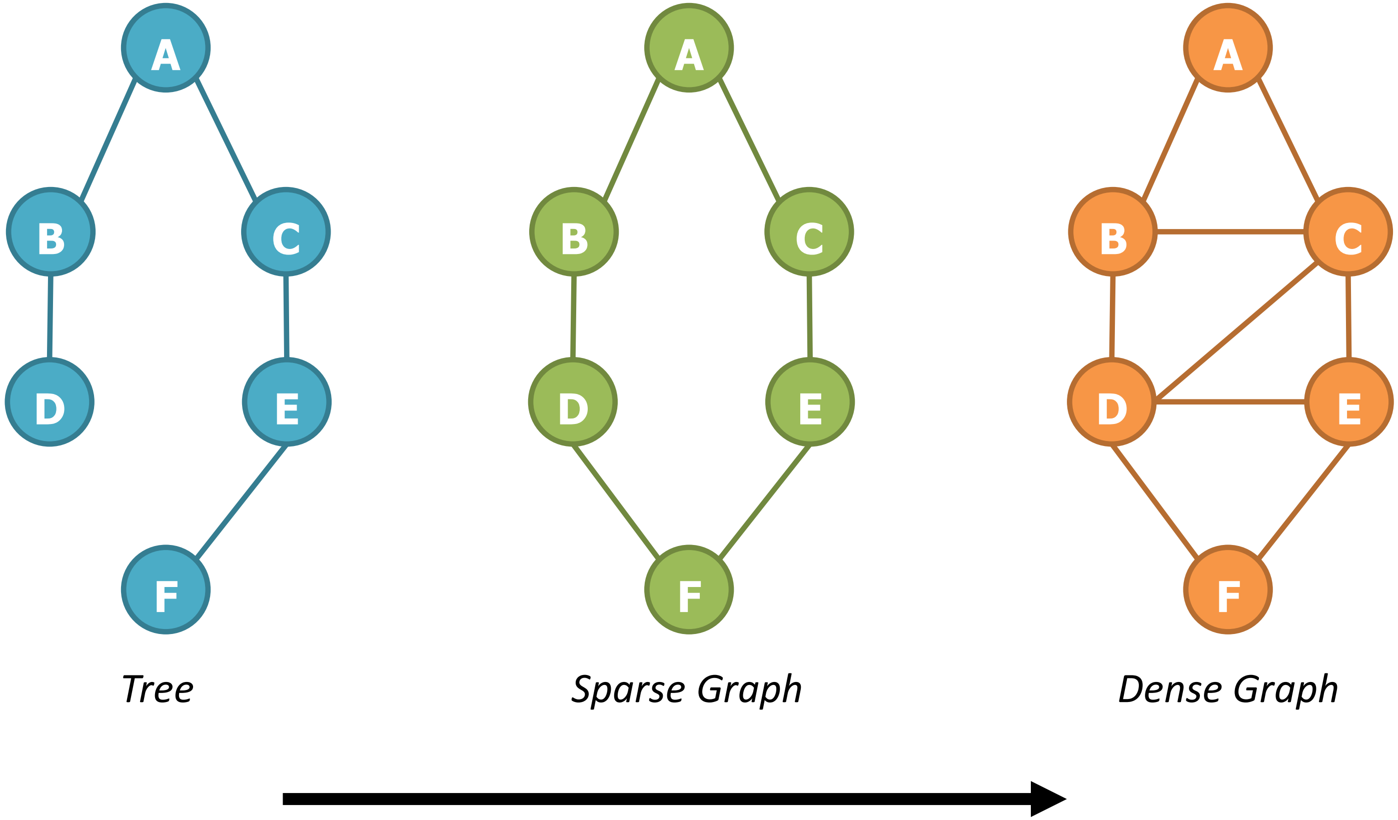
1. Probabilistic graphical model (e.g., factor graph)



2. Probabilistic inference algorithm
   (e.g., variable elimination or junction tree)

[Van den Broeck; AAAI-KRR'15]

Classical Reasoning

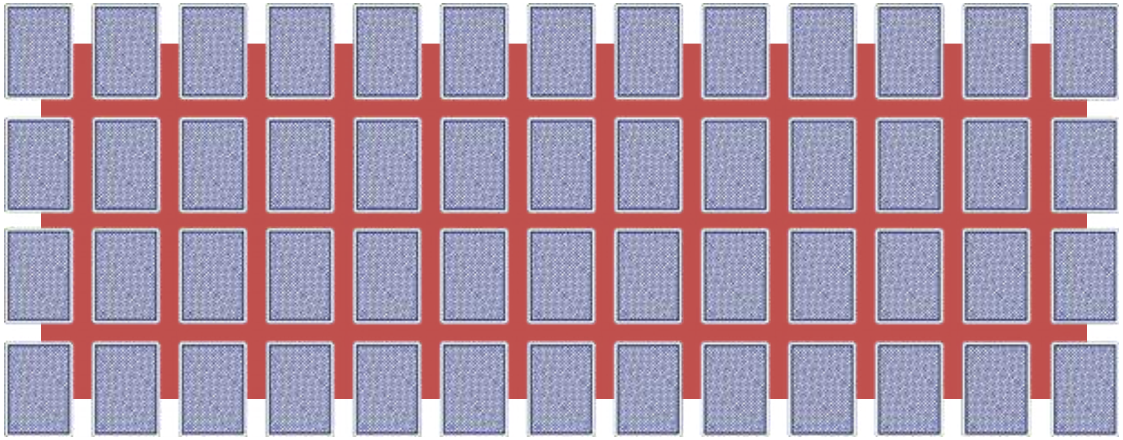


- Higher treewidth
- Fewer conditional independencies
- Slower inference

Automated Reasoning

Let us automate this:

1. Probabilistic graphical model (e.g., factor graph) is fully connected!



(artist's impression)

2. Probabilistic inference algorithm (e.g., variable elimination or junction tree) builds a table with **52^{52}** rows

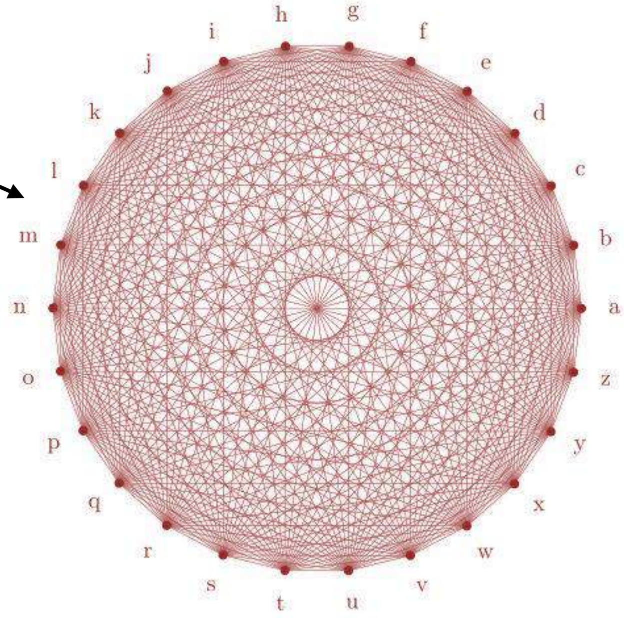
[Van den Broeck; AAAI-KRR'15]

Lifted Inference in SRL

- Statistical relational model (e.g., MLN)

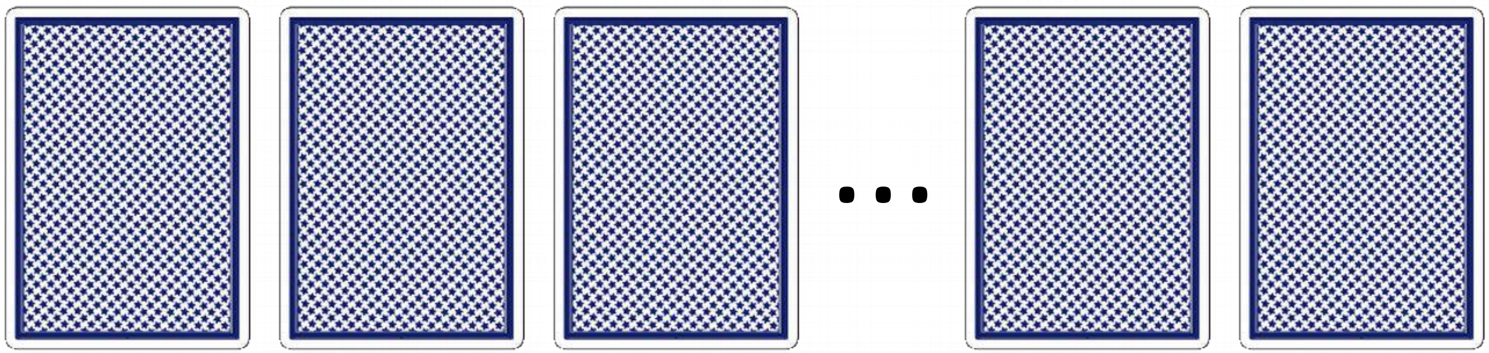
3.14 FacultyPage(x) ∧ Linked(x,y) ⇒ CoursePage(y)

- As a probabilistic graphical model:
  - 26 pages; 728 variables; 676 factors
  - 1000 pages; 1,002,000 variables; 1,000,000 factors



- Highly intractable?
  - **Lifted inference** in milliseconds!

Tractable Reasoning

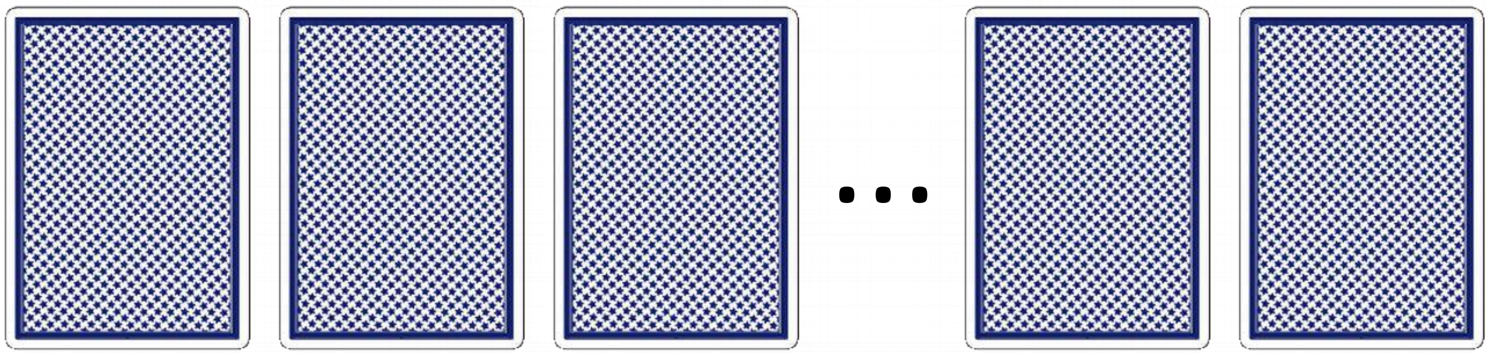


What's going on here?

Which property makes reasoning tractable?

[Niepert and Van den Broeck, AAAI'14], [Van den Broeck, AAAI-KRR'15]

Tractable Reasoning



What's going on here?

Which property makes reasoning tractable?

- High-level (first-order) reasoning
- Symmetry
- Exchangeability

⇒ **Lifted Inference**

[Niepert and Van den Broeck, AAAI'14], [Van den Broeck, AAAI-KRR'15]

Model Counting

- Model = solution to a propositional logic formula Δ
- Model counting = #SAT

$\Delta = (\text{Rain} \Rightarrow \text{Cloudy})$

Rain	Cloudy	Model?
T	T	Yes
T	F	No
F	T	Yes
F	F	Yes

+

#SAT = 3

First-Order Model Counting

Model = solution to first-order logic formula Δ

Δ = ∀d (Rain(d)
⇒ Cloudy(d))

Days = {Monday}

First-Order Model Counting

Model = solution to first-order logic formula Δ

Δ = ∀d (Rain(d) ⇒ Cloudy(d))

Days = {Monday}

Rain(M)	Cloudy(M)	Model?
T	T	Yes
T	F	No
F	T	Yes
F	F	Yes

+

FOMC = 3

First-Order Model Counting

Model = solution to first-order logic formula Δ

$\Delta = \forall d \; (\mathrm{Rain}(d) \Rightarrow \mathrm{Cloudy}(d))$

Days = {Monday **Tuesday**}

First-Order Model Counting

Model = solution to first-order logic formula Δ

Δ = ∀d (Rain(d) ⇒ Cloudy(d))

Days = {Monday **Tuesday**}

Rain(M)	Cloudy(M)	Rain(T)	Cloudy(T)	Model?
T	T	T	T	Yes
T	F	T	T	No
F	T	T	T	Yes
F	F	T	T	Yes
T	T	T	F	No
T	F	T	F	No
F	T	T	F	No
F	F	T	F	No
T	T	F	T	Yes
T	F	F	T	No
F	T	F	T	Yes
F	F	F	T	Yes
T	T	F	F	Yes
T	F	F	F	No
F	T	F	F	Yes
F	F	F	F	Yes

+

#SAT = 9

FOMC Inference

Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

Domain = {n people}

FOMC Inference

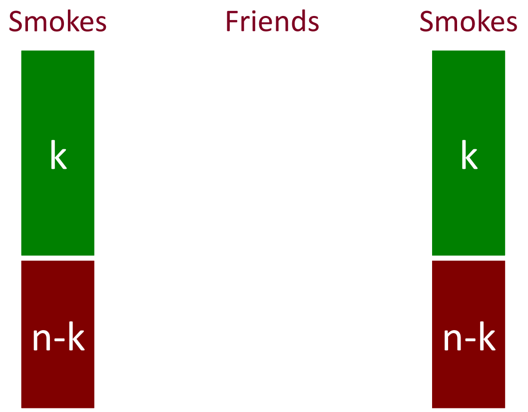
Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...



FOMC Inference

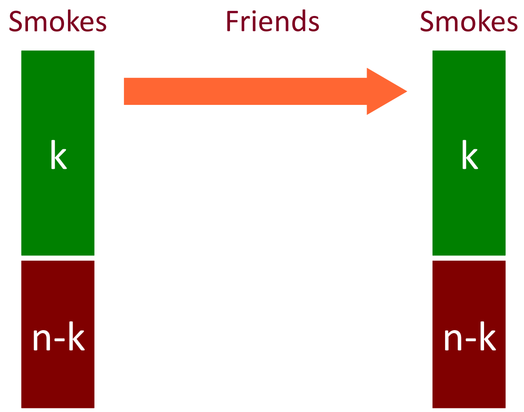
$\Delta = \forall x,y, (\text{Smokes}(x) \wedge \text{Friends}(x,y) \Rightarrow \text{Smokes}(y))$

Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...



FOMC Inference

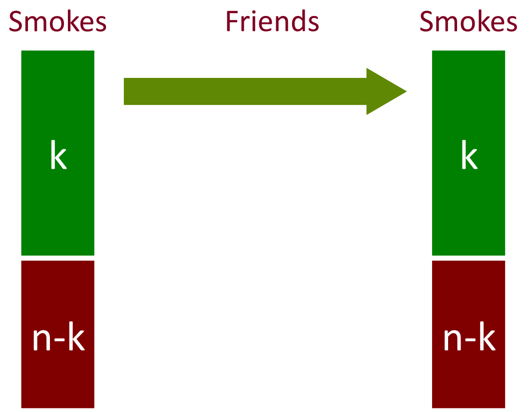
$\Delta = \forall x,y, (Smokes(x) \wedge Friends(x,y) \Rightarrow Smokes(y))$

Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...



FOMC Inference

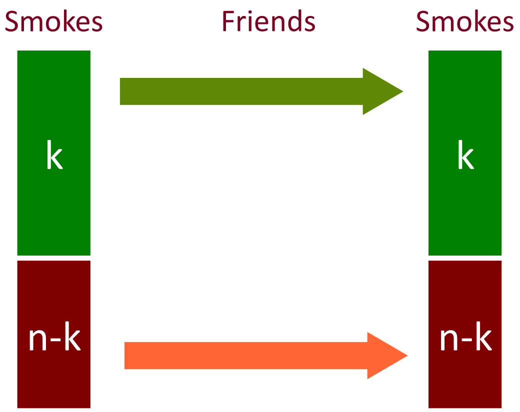
$\Delta = \forall x,y, (Smokes(x) \land Friends(x,y) \Rightarrow Smokes(y))$

Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...



FOMC Inference

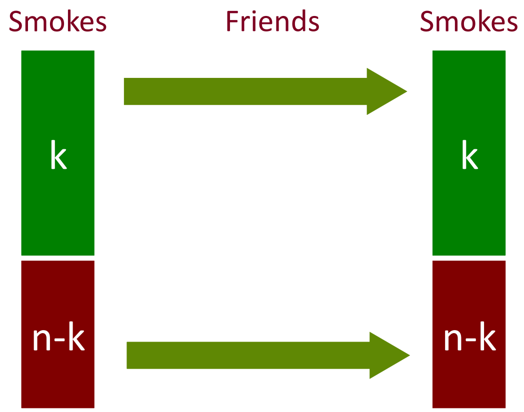
Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...



FOMC Inference

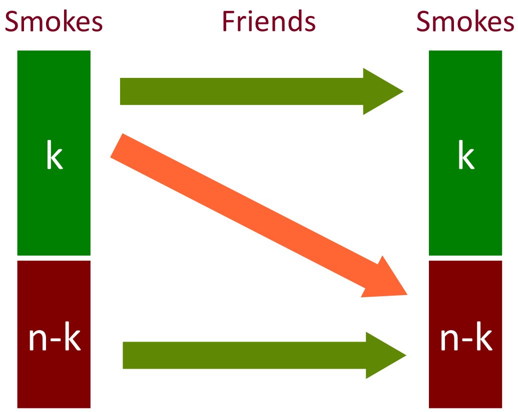
Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...



FOMC Inference

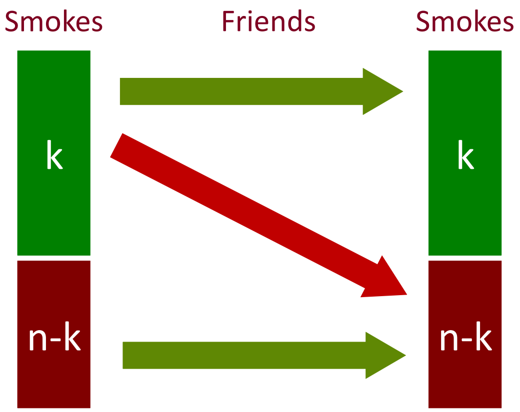
Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...



FOMC Inference

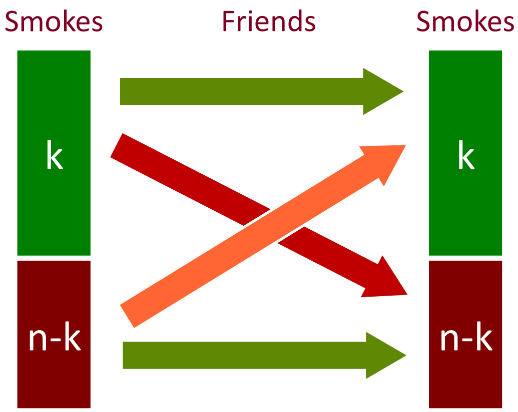
Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...



FOMC Inference

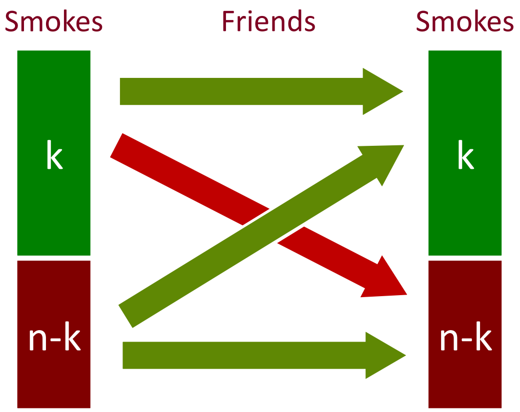
Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

<u>**Database:**</u>

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...



FOMC Inference

Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

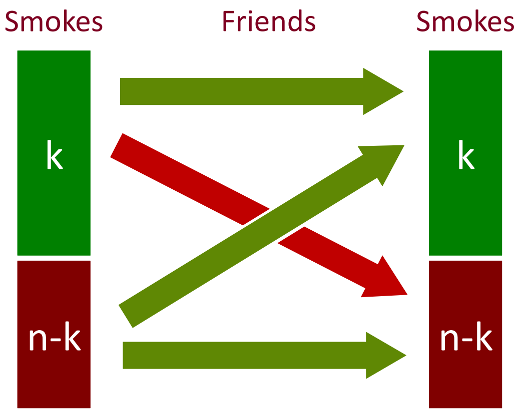
Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...

$\rightarrow 2^{n^2-k(n-k)}$ models



FOMC Inference

$\Delta = \forall x,y, (\text{Smokes}(x) \wedge \text{Friends}(x,y) \Rightarrow \text{Smokes}(y))$

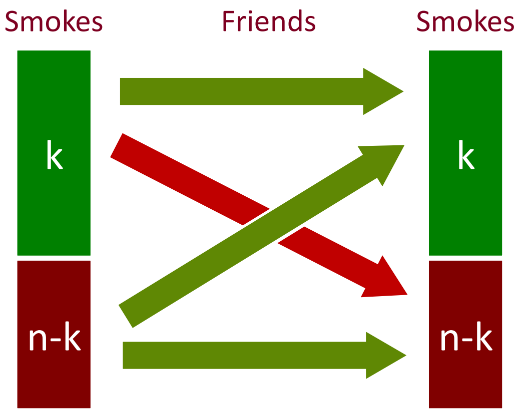
Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...

$\rightarrow 2^{n^2 - k(n-k)}$ models



- If we know that there are *k* smokers?

FOMC Inference

Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

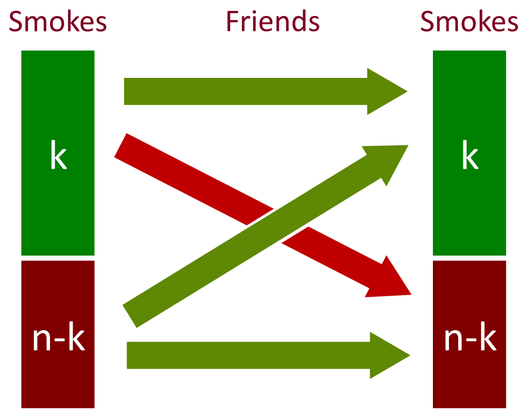
Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...

→ $2^{n^2-k(n-k)}$ models



- If we know that there are *k* smokers?

→ $\binom{n}{k} 2^{n^2-k(n-k)}$ models

FOMC Inference

Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

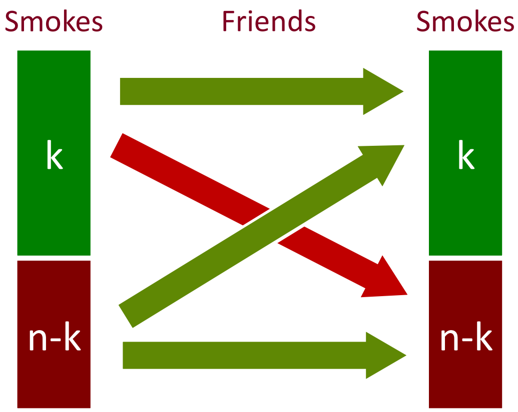
Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...

→ $2^{n^2-k(n-k)}$ models



- If we know that there are *k* smokers? → $\binom{n}{k}2^{n^2-k(n-k)}$ models

- In total…

FOMC Inference

Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

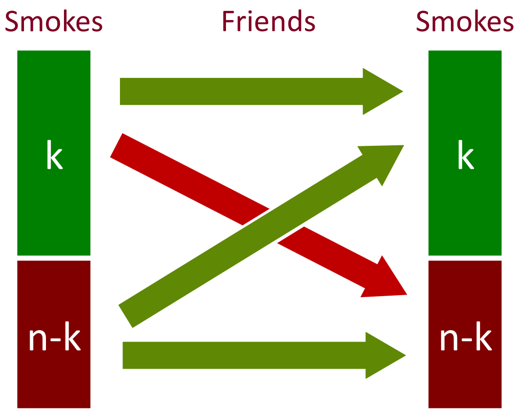
Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

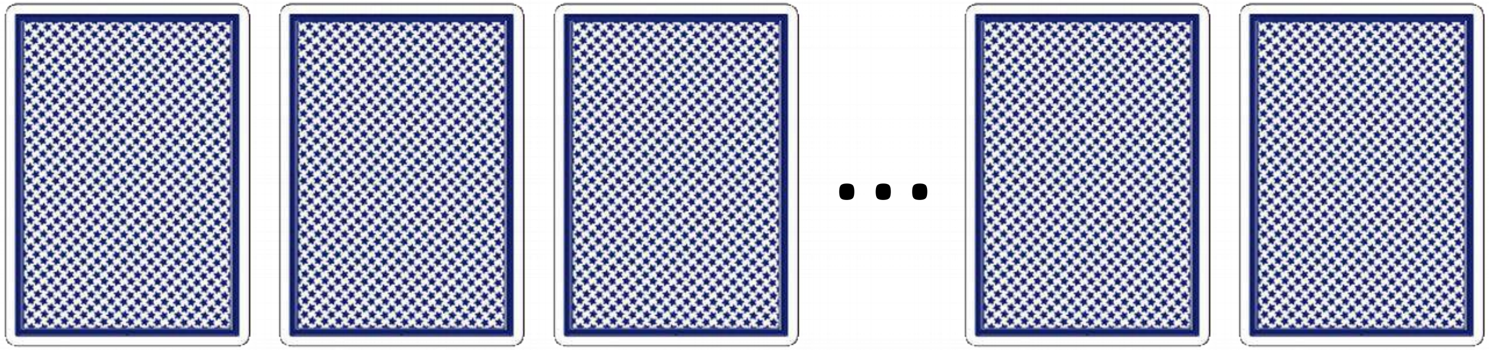
Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...

→ $2^{n^2-k(n-k)}$ models



- If we know that there are *k* smokers? → $\binom{n}{k}2^{n^2-k(n-k)}$ models

- In total… → $\sum_{k=0}^{n}\binom{n}{k}2^{n^2-k(n-k)}$ models



Let us automate this:

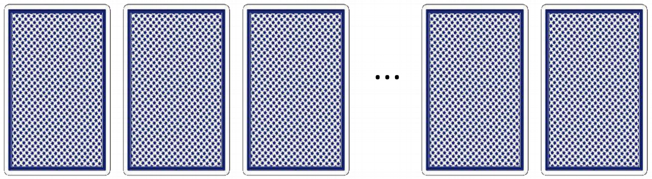
- **Relational** model

$$\forall p, \exists c, \mathrm{Card}(p,c)$$
$$\forall c, \exists p, \mathrm{Card}(p,c)$$
$$\forall p, \forall c, \forall c', \mathrm{Card}(p,c) \wedge \mathrm{Card}(p,c') \Rightarrow c = c'$$

- **Lifted** probabilistic inference algorithm

[Van den Broeck; AAAI-KRR'15]

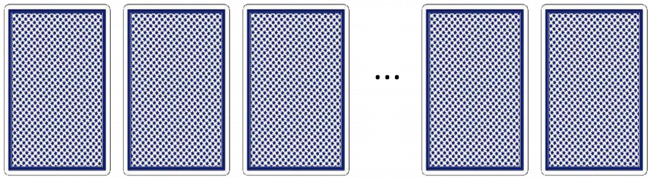
Playing Cards Revisited



$$\forall p, \exists c, \text{Card}(p,c)$$
$$\forall c, \exists p, \text{Card}(p,c)$$
$$\forall p, \forall c, \forall c', \text{Card}(p,c) \land \text{Card}(p,c') \Rightarrow c = c'$$

[Van den Broeck.; AAAI-KR'15]

Playing Cards Revisited

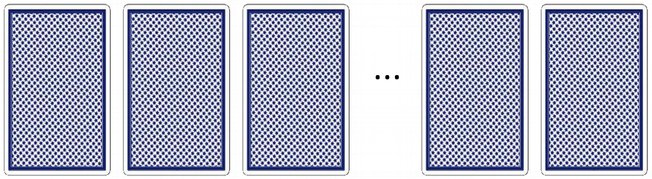


∀p, ∃c, Card(p,c)
∀c, ∃p, Card(p,c)
∀p, ∀c, ∀c', Card(p,c) ∧ Card(p,c') ⇒ c = c'

$$\#SAT = \sum_{k=0}^{n} \binom{n}{k} \sum_{l=0}^{n} \binom{n}{l} (l+1)^k (-1)^{2n-k-l} = n!$$

[Van den Broeck.; AAAI-KR'15]

Playing Cards Revisited



∀p, ∃c, Card(p,c)
∀c, ∃p, Card(p,c)
∀p, ∀c, ∀c', Card(p,c) ∧ Card(p,c') ⇒ c = c'

$$\#SAT = \sum_{k=0}^{n} \binom{n}{k} \sum_{l=0}^{n} \binom{n}{l} (l+1)^k (-1)^{2n-k-l} = n!$$

Computed in time polynomial in n

[Van den Broeck.; AAAI-KR'15]

Open-World Lifted Query Eval

$Q = \exists x\, \exists y\, Smoker(x) \wedge Friend(x,y)$

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

$= 1 - (1 - P(Scientist(A) \wedge \exists y\, Coauthor(A,y))$
$\times (1 - P(Scientist(B) \wedge \exists y\, Coauthor(B,y))$
$\times (1 - P(Scientist(C) \wedge \exists y\, Coauthor(C,y))$
$\times (1 - P(Scientist(D) \wedge \exists y\, Coauthor(D,y))$
$\times (1 - P(Scientist(E) \wedge \exists y\, Coauthor(E,y))$
$\times (1 - P(Scientist(F) \wedge \exists y\, Coauthor(F,y))$
…

⇨ All together, probability $(1-p)^k$

Open-World Lifted Query Eval

$\boxed{Q = \exists x\ \exists y\ \text{Smoker}(x) \land \text{Friend}(x,y)}$

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

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

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

Open-world query evaluation on empty db = Symmetric lifted inference

Even on #P-hard queries!

Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

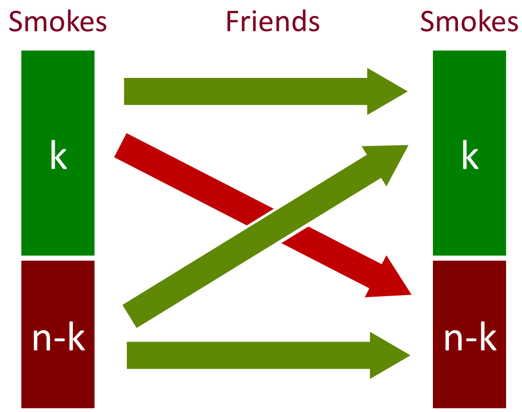
Domain = {n people}

- If we know precisely who smokes, and there are *k* smokers?

Database:

Smokes(Alice) = 1
Smokes(Bob) = 0
Smokes(Charlie) = 0
Smokes(Dave) = 1
Smokes(Eve) = 0
...

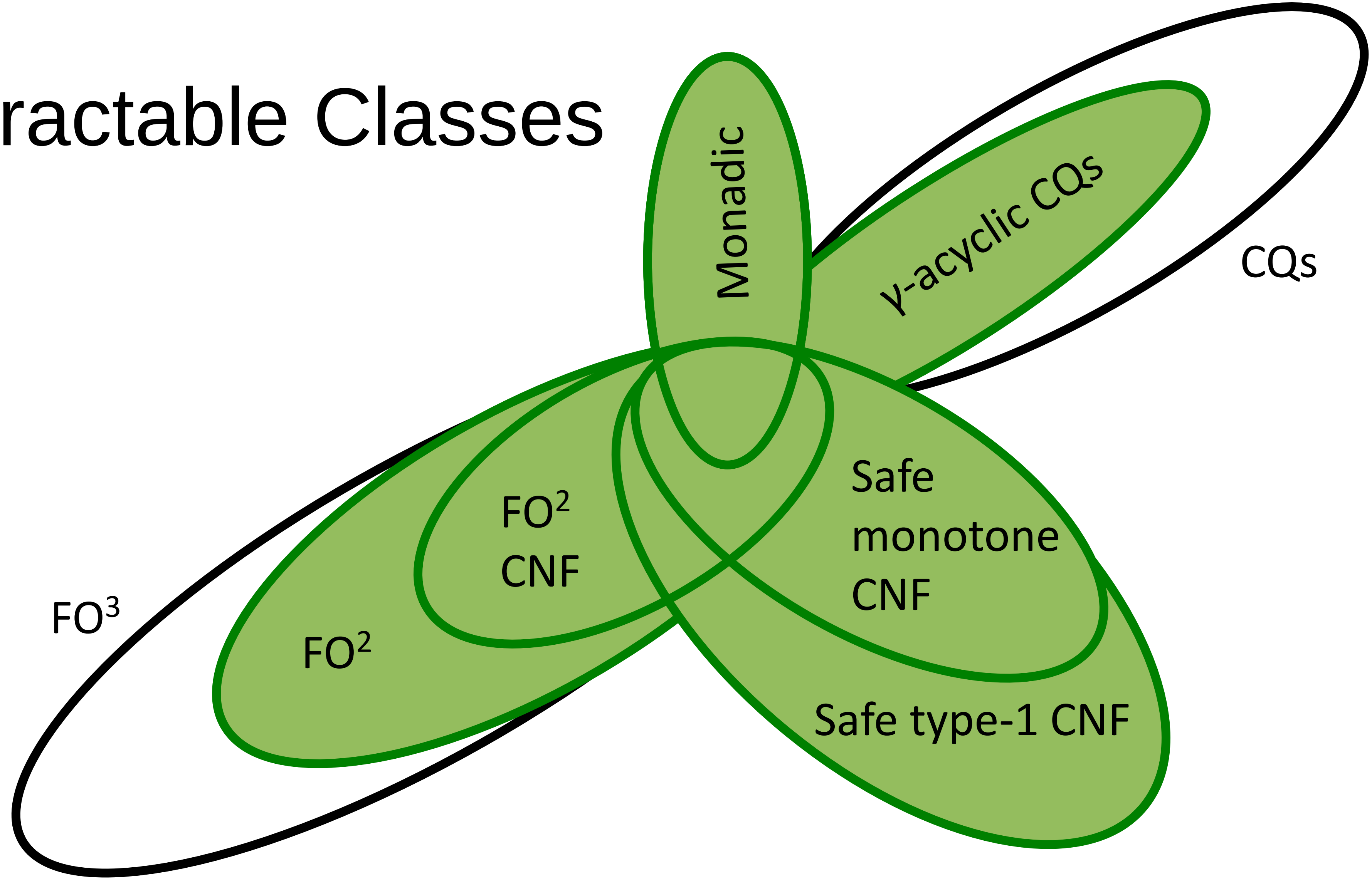
$\rightarrow 2^{n^2 - k(n-k)}$ models



- If we know that there are *k* smokers? $\rightarrow \binom{n}{k} 2^{n^2 - k(n-k)}$ models

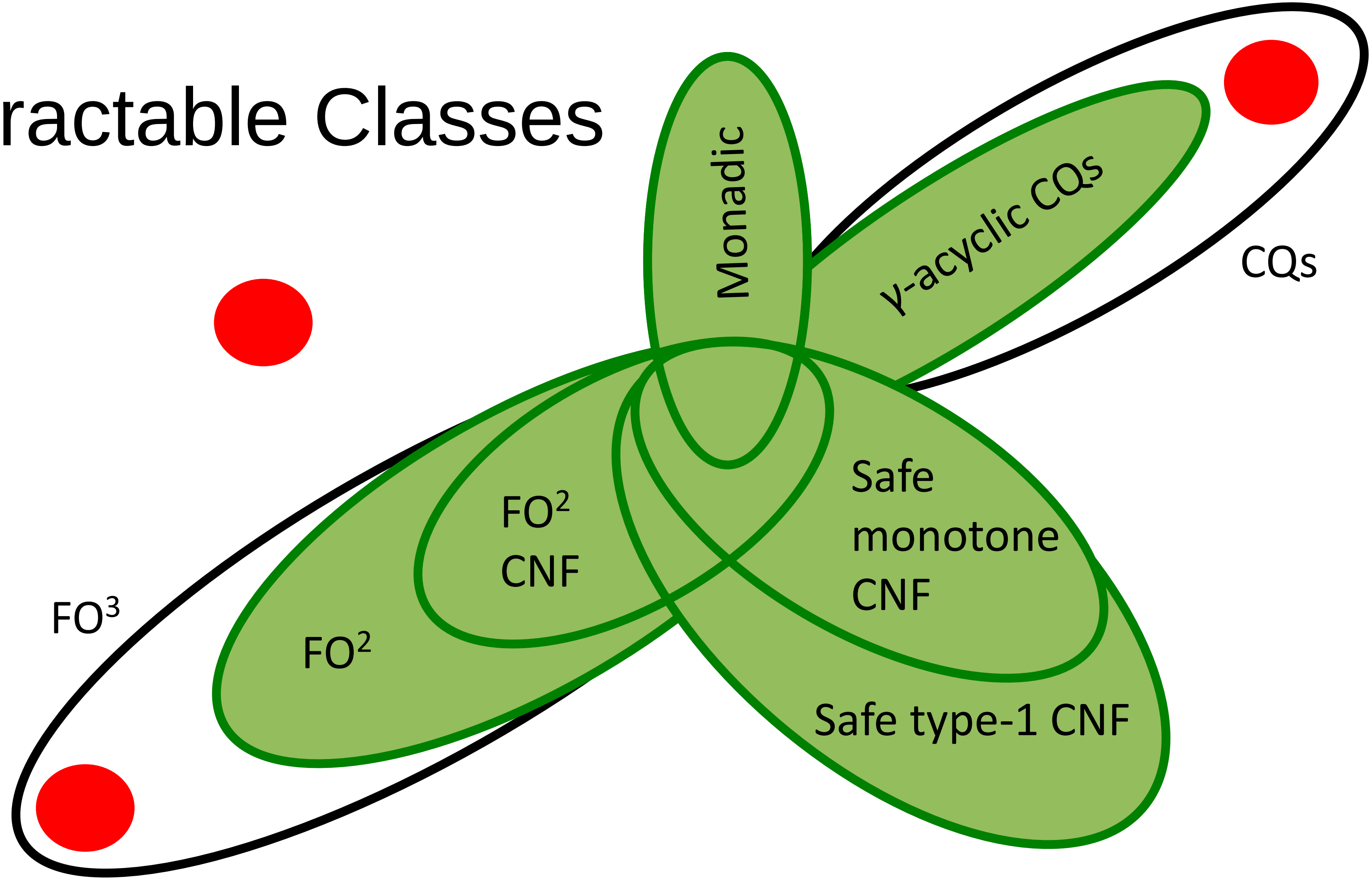
- In total... $\rightarrow \sum_{k=0}^{n} \binom{n}{k} 2^{n^2 - k(n-k)}$ models

Tractable Classes



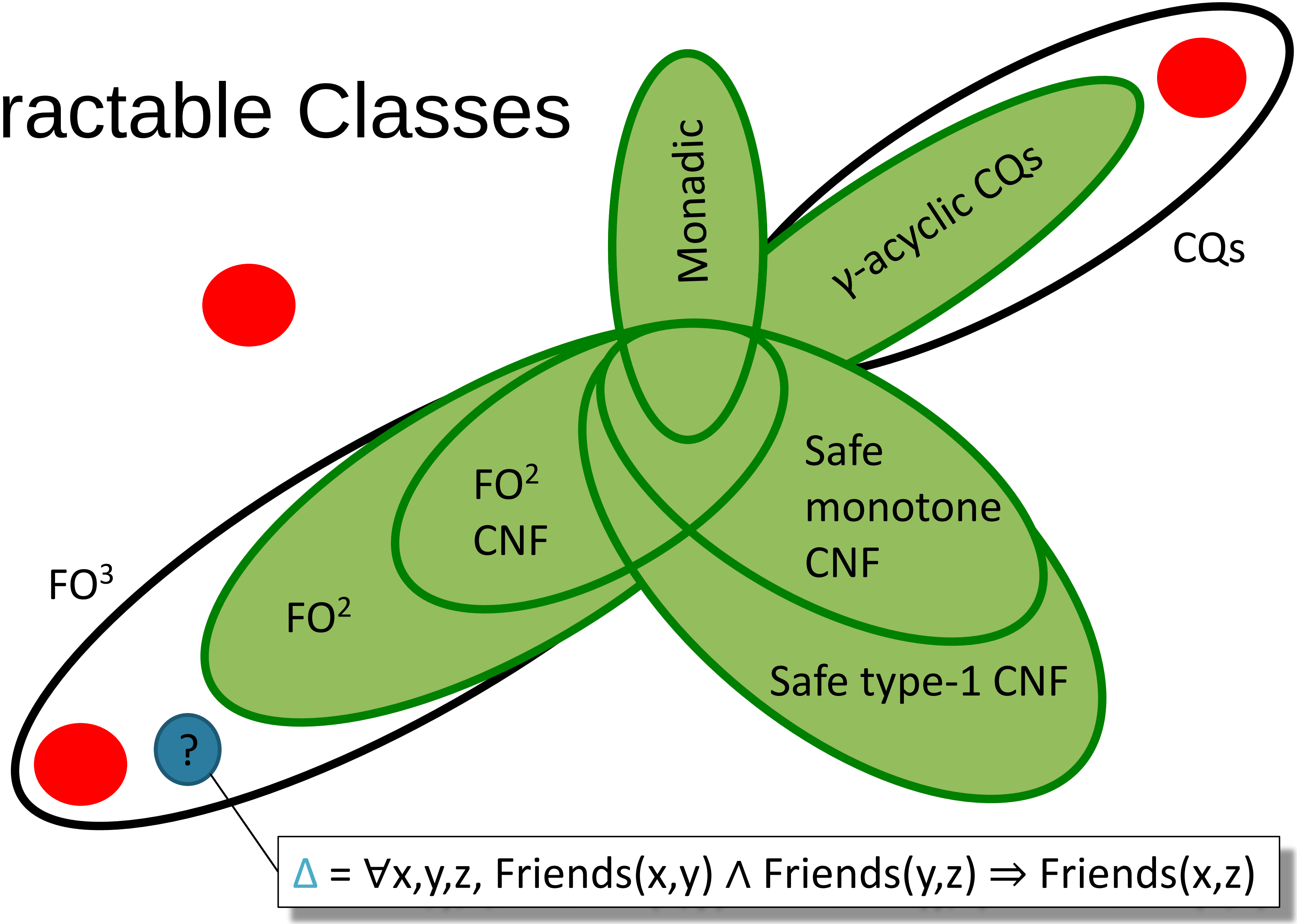
[VdB; NIPS'11], [VdB et al.; KR'14], [Gribkoff, VdB, Suciu; UAI'15], [Beame, VdB, Gribkoff, Suciu; PODS'15], etc.

Tractable Classes



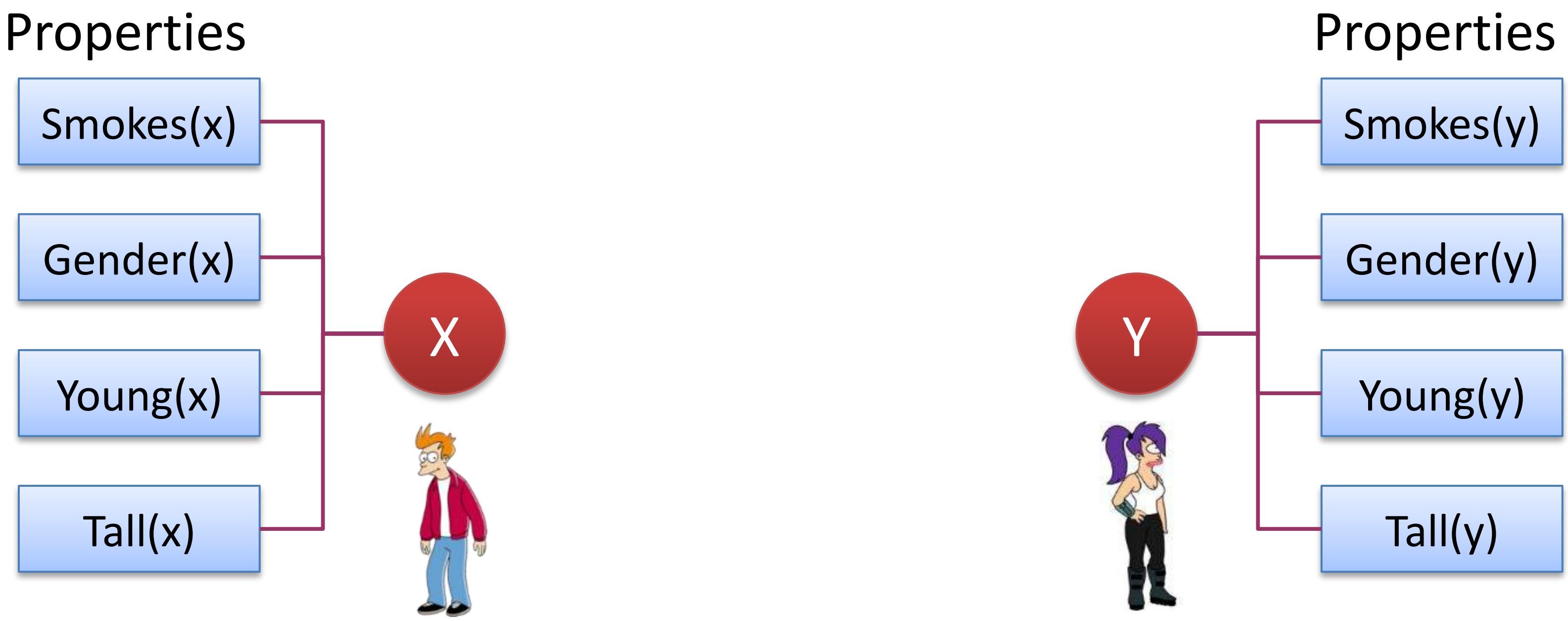
[VdB; NIPS’11], [VdB et al.; KR’14], [Gribkoff, VdB, Suciu; UAI’15], [Beame, VdB, Gribkoff, Suciu; PODS’15], etc.

Tractable Classes

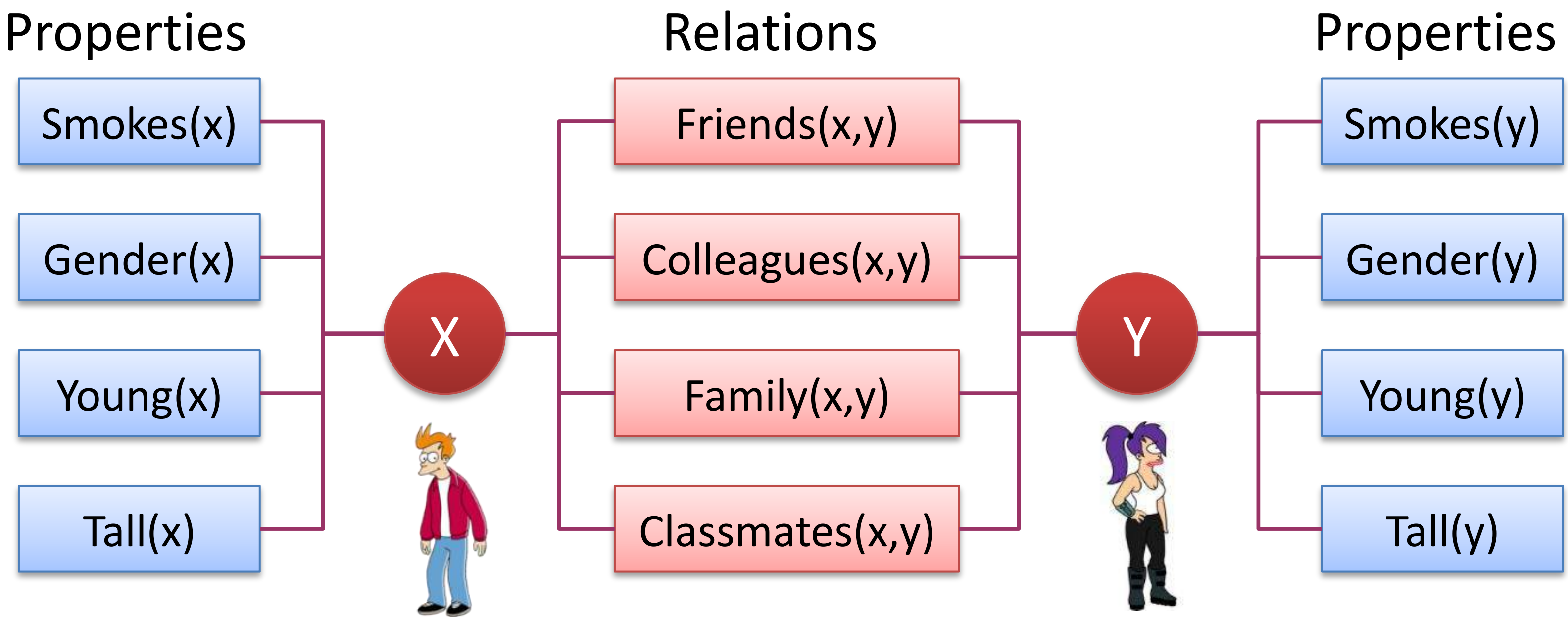


[VdB; NIPS'11], [VdB et al.; KR'14], [Gribkoff, VdB, Suciu; UAI'15], [Beame, VdB, Gribkoff, Suciu; PODS'15], etc.

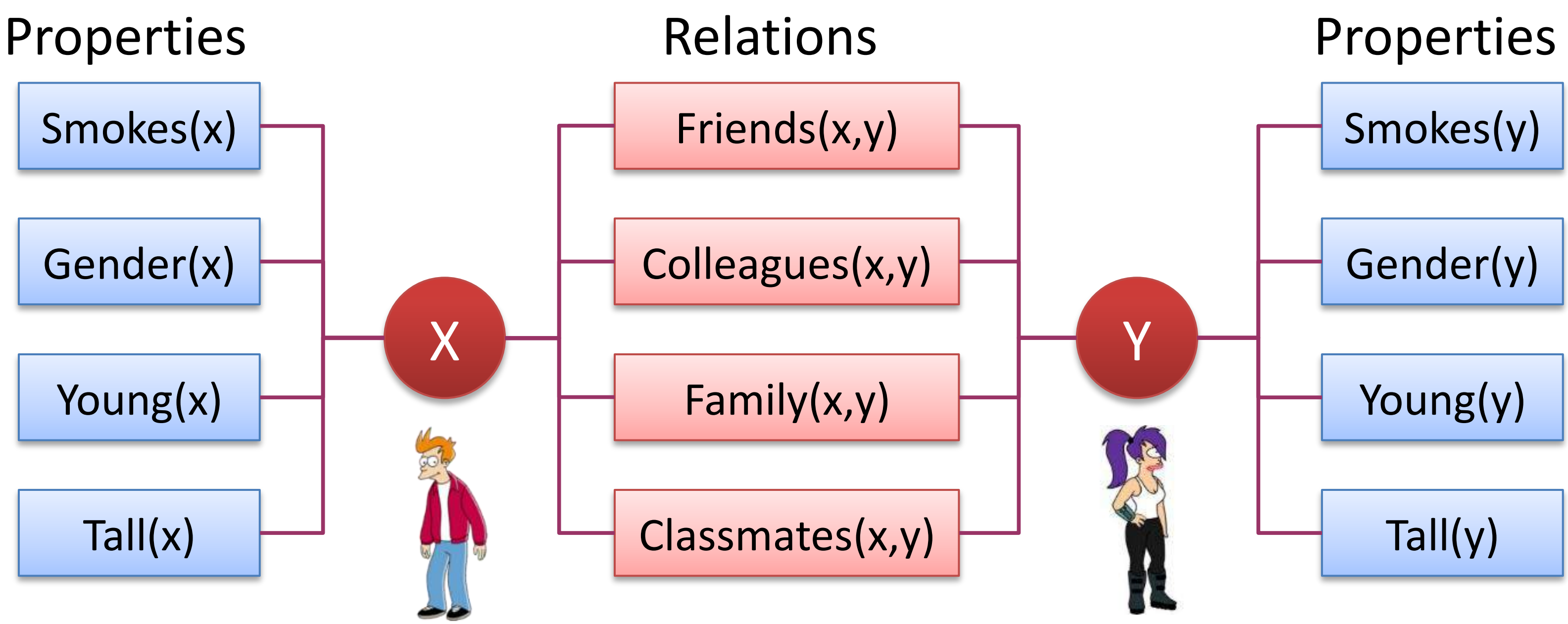
FO^2 is liftable!



FO^2 is liftable!



FO^2 is liftable!



"Smokers are more likely to be friends with other smokers."
"Colleagues of the same age are more likely to be friends."
"People are either family or friends, but never both."
"If X is family of Y, then Y is also family of X."
"If X is a parent of Y, then Y cannot be a parent of X."

Uncertainty in AI

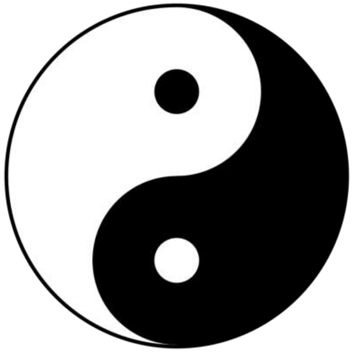
Probability Distribution

=

Qualitative

+

Quantitative



Probabilistic Graphical Models

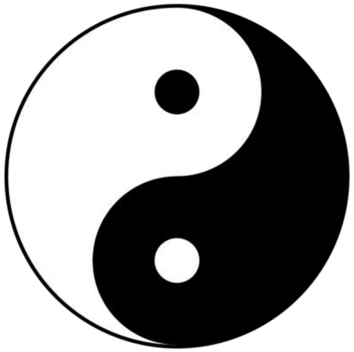
Probability Distribution

=

Graph Structure

+

Parameterization



Probabilistic Graphical Models

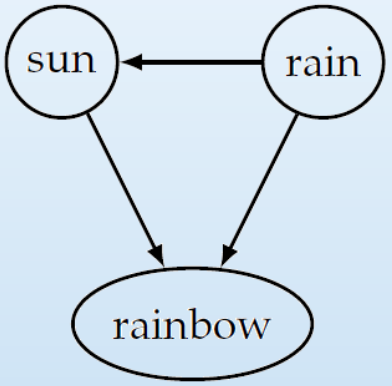
Probability Distribution

=

Graph Structure

+

Parameterization



+

rain	Pr(sun | rain)
T	0.1
F	0.6

rain	sun	Pr(rainbow | rain, sun)
T	T	0.9
T	F	0.05
F	T	0.05
F	F	0

Pr(rain)
0.2

Weighted Model Counting

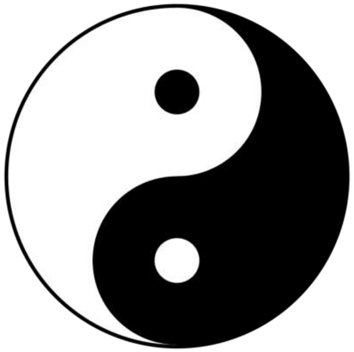
Probability Distribution

=

SAT Formula

+

Weights



[Chavira et al. 2008, Sang et al. 2005]

Weighted Model Counting

Probability Distribution

=

SAT Formula

+

Weights

$\text{Rain} \Rightarrow \text{Cloudy}$

$\text{Sun} \wedge \text{Rain} \Rightarrow \text{Rainbow}$

+

$w(\text{Rain})=1$

$w(\neg\text{Rain})=2$

$w(\text{Cloudy})=3$

$w(\neg\text{Cloudy})=5$

...

[Chavira et al. 2008, Sang et al. 2005]

Weighted First-Order Model Counting

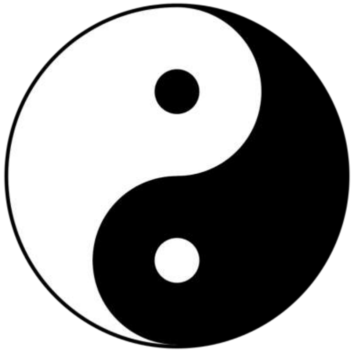
Probability Distribution

=

First-Order Logic

+

Weights



[Van den Broeck 2011, 2013, Gogate 2011]

Weighted First-Order Model Counting

Probability Distribution

=

First-Order Logic

+

Weights

Smokes(x) ∧ Friends(x,y)
⇒ Smokes(y)

+

w(Smokes(a))=1
w(¬Smokes(a))=2
w(Smokes(b))=1
w(¬Smokes(b))=2
w(Friends(a,b))=3
w(¬Friends(a,b))=5
…

[Van den Broeck 2011, 2013, Gogate 2011]

Generalized Model Counting

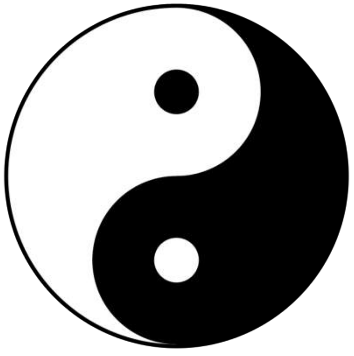
Probability Distribution

=

Logic

+

Weights



Generalized Model Counting

Probability Distribution

=

Logic

+

Weights

Logical Syntax

Model-theoretic Semantics

+

Weight function w(.)

Weighted Model Integration

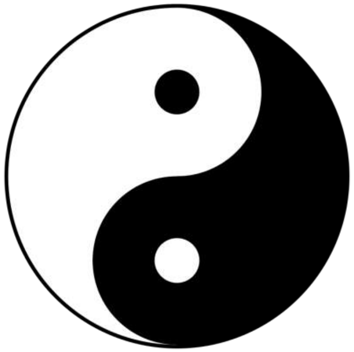
Probability Distribution

=

SMT(LRA)

+

Weights



[Belle et al. IJCAI'15, UAI'15]

Weighted Model Integration

Probability Distribution

=

SMT(LRA)

+

Weights

$0 \leq height \leq 200$
$0 \leq weight \leq 200$
$0 \leq age \leq 100$
$age < 1 \Rightarrow$
$height + weight \leq 90$

+

w(height))=height-10
w(¬height)=3*height^2
w(¬weight)=5
…

[Belle et al. IJCAI'15, UAI'15]

Probabilistic Programming

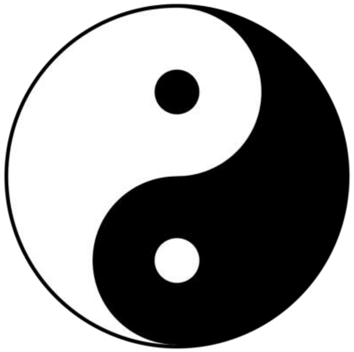
Probability Distribution

=

Logic Programs

+

Weights



[Fierens et al., TPLP'15]

Probabilistic Programming

Probability Distribution

=

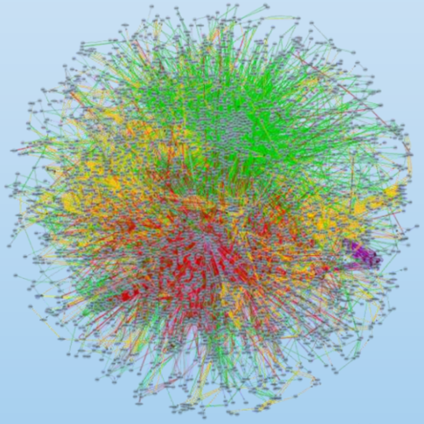
Logic Programs

+

Weights

```
path(X,Y) :-
        edge(X,Y).
path(X,Y) :-
        edge(X,Z), path(Z,Y).
```

+



[Fierens et al., TPLP'15]

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 make sense
  - FREE for UCQs
  - expensive otherwise

Long-Term Outlook

Probabilistic inference and learning exploit

~ 1988: conditional independence

~ 2000: contextual independence (local structure)

Long-Term Outlook

Probabilistic inference and learning exploit

~ 1988: conditional independence

~ 2000: contextual independence (local structure)

~ 201?: **symmetry & exchangeability & first-order**

References

- Ceylan, Ismail Ilkan, Adnan Darwiche, and Guy Van den Broeck. "Open-world probabilistic databases." Proceedings of KR (2016).

- Suciu, Dan, Dan Olteanu, Christopher Ré, and Christoph Koch. "Probabilistic databases." Synthesis Lectures on Data Management 3, no. 2 (2011): 1-180.

- Dong, Xin, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. "Knowledge vault: A web-scale approach to probabilistic knowledge fusion." In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 601-610. ACM, 2014.

- Carlson, Andrew, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr, and Tom M. Mitchell. "Toward an Architecture for Never-Ending Language Learning." In AAAI, vol. 5, p. 3. 2010.

- Niu, Feng, Ce Zhang, Christopher Ré, and Jude W. Shavlik. "DeepDive: Web-scale Knowledge-base Construction using Statistical Learning and Inference." VLDS 12 (2012): 25-28.

References

- Chen, Brian X. "Siri, Alexa and Other Virtual Assistants Put to the Test" The New York Times (2016).
- Dalvi, Nilesh, and Dan Suciu. "The dichotomy of probabilistic inference for unions of conjunctive queries." Journal of the ACM (JACM) 59, no. 6 (2012): 30.
- De Raedt, Luc, Anton Dries, Ingo Thon, Guy Van den Broeck, and Mathias Verbeke. "Inducing probabilistic relational rules from probabilistic examples." In Proceedings of the 24th International Conference on Artificial Intelligence, pp. 1835-1843. AAAI Press, 2015.
- Van den Broeck, Guy. "Towards high-level probabilistic reasoning with lifted inference." AAAI Spring Symposium on KRR (2015).
- Niepert, Mathias, and Guy Van den Broeck. "Tractability through exchangeability: A new perspective on efficient probabilistic inference." AAAI (2014).
- Van den Broeck, Guy. "On the completeness of first-order knowledge compilation for lifted probabilistic inference." In Advances in Neural Information Processing Systems, pp. 1386-1394. 2011.

References

- Van den Broeck, Guy, Wannes Meert, and Adnan Darwiche. "Skolemization for weighted first-order model counting." In Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR). 2014.

- Gribkoff, Eric, Guy Van den Broeck, and Dan Suciu. "Understanding the complexity of lifted inference and asymmetric weighted model counting." UAI, 2014.

- Beame, Paul, Guy Van den Broeck, Eric Gribkoff, and Dan Suciu. "Symmetric weighted first-order model counting." In Proceedings of the 34th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, pp. 313-328. ACM, 2015.

- Chavira, Mark, and Adnan Darwiche. "On probabilistic inference by weighted model counting." Artificial Intelligence 172.6 (2008): 772-799.

- Sang, Tian, Paul Beame, and Henry A. Kautz. "Performing Bayesian inference by weighted model counting." AAAI. Vol. 5. 2005.

References

- Van den Broeck, Guy, Nima Taghipour, Wannes Meert, Jesse Davis, and Luc De Raedt. "Lifted probabilistic inference by first-order knowledge compilation." In Proceedings of the Twenty-Second international joint conference on Artificial Intelligence, pp. 2178-2185. AAAI Press/International Joint Conferences on Artificial Intelligence, 2011.

- Van den Broeck, Guy. Lifted inference and learning in statistical relational models. Diss. Ph. D. Dissertation, KU Leuven, 2013.

- Gogate, Vibhav, and Pedro Domingos. "Probabilistic theorem proving." UAI (2011).

References

- Belle, Vaishak, Andrea Passerini, and Guy Van den Broeck. "Probabilistic inference in hybrid domains by weighted model integration." Proceedings of 24th International Joint Conference on Artificial Intelligence (IJCAI). 2015.

- Belle, Vaishak, Guy Van den Broeck, and Andrea Passerini. "Hashing-based approximate probabilistic inference in hybrid domains." In Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence (UAI). 2015.

- Fierens, Daan, Guy Van den Broeck, Joris Renkens, Dimitar Shterionov, Bernd Gutmann, Ingo Thon, Gerda Janssens, and Luc De Raedt. "Inference and learning in probabilistic logic programs using weighted boolean formulas." Theory and Practice of Logic Programming 15, no. 03 (2015): 358-401.