First-Order Probabilistic Reasoning: Successes and Challenges

Guy Van den Broeck

UCLA

IJCAI Early Career Spotlight
Jul 14, 2016

Overview

1. *Why first-order probabilistic **models**?*

2. *Why first-order probabilistic **reasoning**?*

3. *How does lifted inference **work**?*

4. *What are the **successes**?*

5. *What are the **challenges**?*

Why do we need first-order probabilistic ***models****?*

Graphical Model Learning

Medical Records

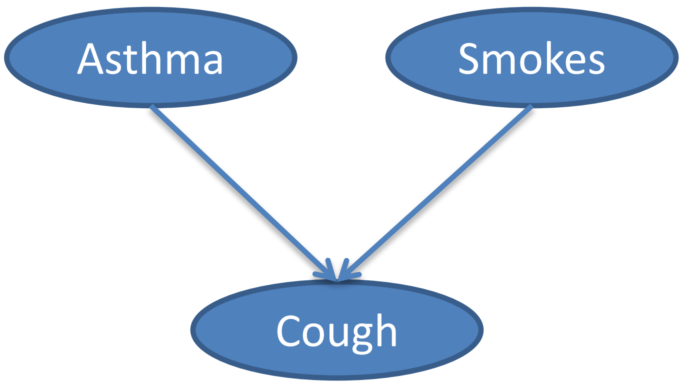
Name	Cough	Asthma	Smokes
Alice	1	1	0
Bob	0	0	0
Charlie	0	1	0
Dave	1	0	1
Eve	1	0	0

Graphical Model Learning

Medical Records

Bayesian Network

Name	Cough	Asthma	Smokes
Alice	1	1	0
Bob	0	0	0
Charlie	0	1	0
Dave	1	0	1
Eve	1	0	0



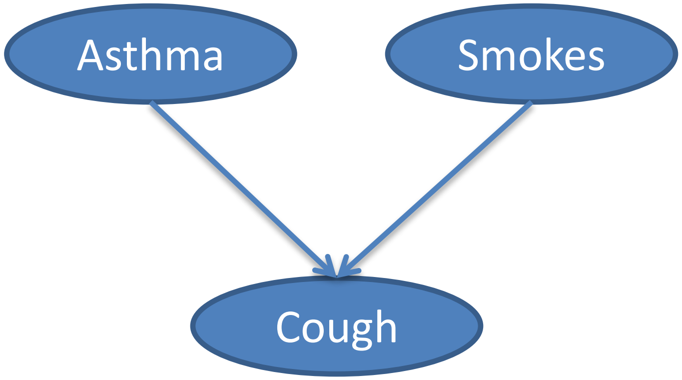
Graphical Model Learning

Medical Records

Bayesian Network

Name	Cough	Asthma	Smokes
Alice	1	1	0
Bob	0	0	0
Charlie	0	1	0
Dave	1	0	1
Eve	1	0	0

Frank	1	?	?



Graphical Model Learning

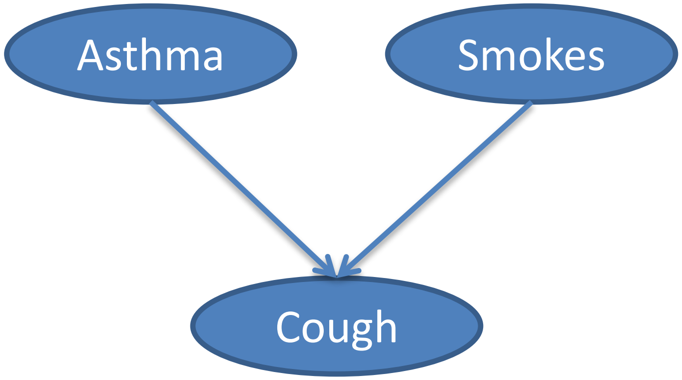
Medical Records

Bayesian Network

Name	Cough	Asthma	Smokes
Alice	1	1	0
Bob	0	0	0
Charlie	0	1	0
Dave	1	0	1
Eve	1	0	0

Frank	1	?	?

Frank	1	0.3	0.2



Statistical Relational Learning

Medical Records → Bayesian Network

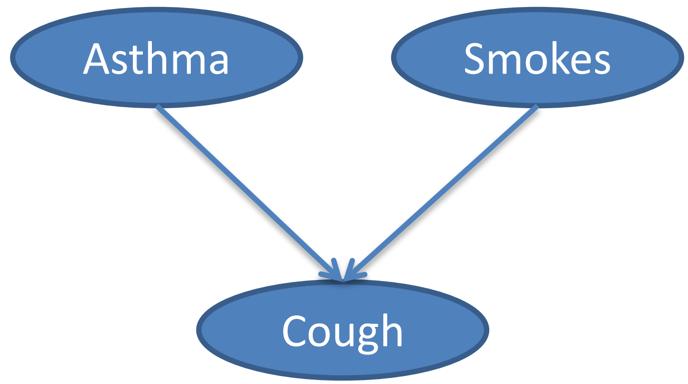
Name	Cough	Asthma	Smokes
Alice	1	1	0
Bob	0	0	0
Charlie	0	1	0
Dave	1	0	1
Eve	1	0	0

Friends

Brothers

Frank	1	?	?

Frank	1	0.3	0.2



Statistical Relational Learning

Medical Records

Bayesian Network

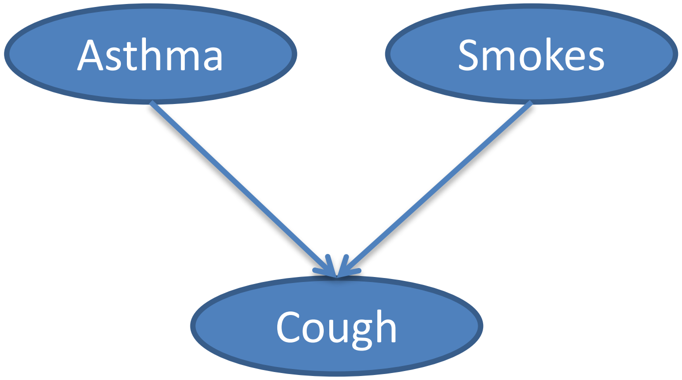
Name	Cough	Asthma	Smokes
Alice	1	1	0
Bob	0	0	0
Charlie	0	1	0
Dave	1	0	1
Eve	1	0	0

Friends

Brothers

Frank	1	?	?

Frank	1	0.3	0.2



Statistical Relational Learning

Medical Records

Bayesian Network

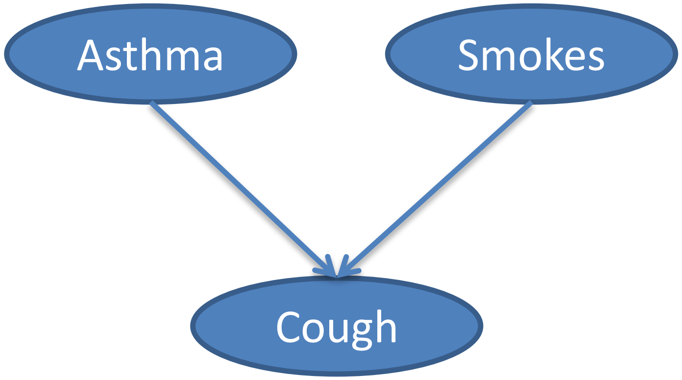
Name	Cough	Asthma	Smokes
Alice	1	1	0
Bob	0	0	0
Charlie	0	1	0
Dave	1	0	1
Eve	1	0	0

Friends

Brothers

Frank	1	?	?

Frank	1	0.3	0.2
Frank	1	0.2	0.6



Statistical Relational Learning

Medical Records → Bayesian Network

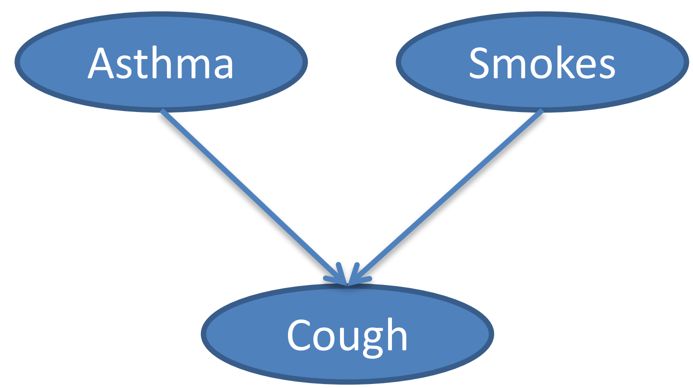
Name	Cough	Asthma	Smokes
Alice	1	1	0
Bob	0	0	0
Charlie	0	1	0
Dave	1	0	1
Eve	1	0	0

Friends

Brothers

Frank	1	?	?

Frank	1	0.3	0.2
Frank	1	0.2	0.6

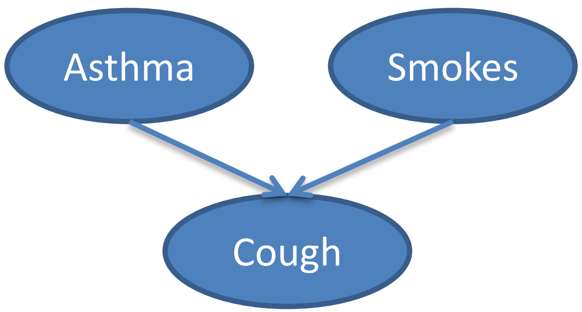


Rows are **independent** during learning and inference!

Statistical Relational Learning

Augment graphical model with relations between entities (rows).

<u>Intuition</u>



+ Friends have similar smoking habits

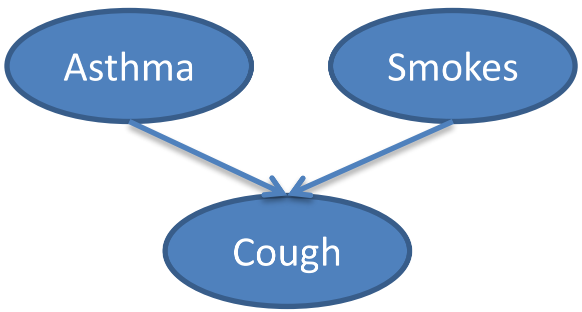
+ Asthma can be hereditary

<u>Markov Logic</u>

Statistical Relational Learning

Augment graphical model with relations between entities (rows).

Intuition



+ Friends have similar smoking habits
+ Asthma can be hereditary

Markov Logic

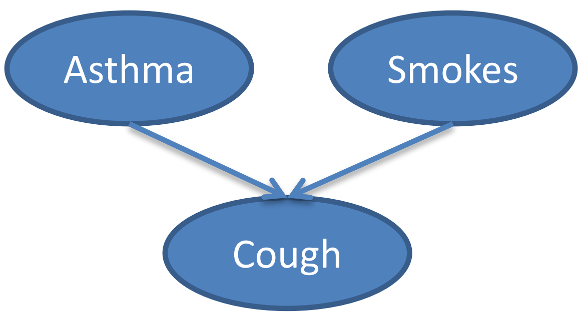
2.1 Asthma ⇒ Cough

3.5 Smokes ⇒ Cough

Statistical Relational Learning

Augment graphical model with relations between entities (rows).

Intuition



+ Friends have similar smoking habits

+ Asthma can be hereditary

Markov Logic

2.1 Asthma(x) ⇒ Cough(x)

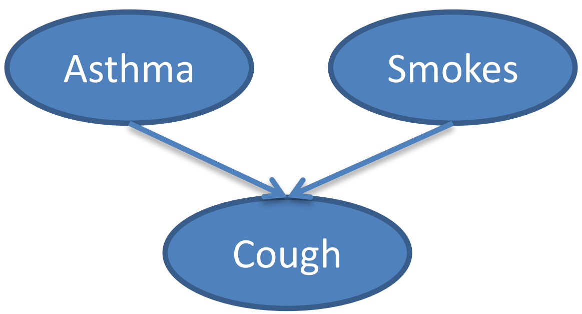
3.5 Smokes(x) ⇒ Cough(x)

↑ Logical variables refer to entities

Statistical Relational Learning

Augment graphical model with relations between entities (rows).

<u>Intuition</u>



+ Friends have similar smoking habits

+ Asthma can be hereditary

<u>Markov Logic</u>

2.1 Asthma(x) ⇒ Cough(x)

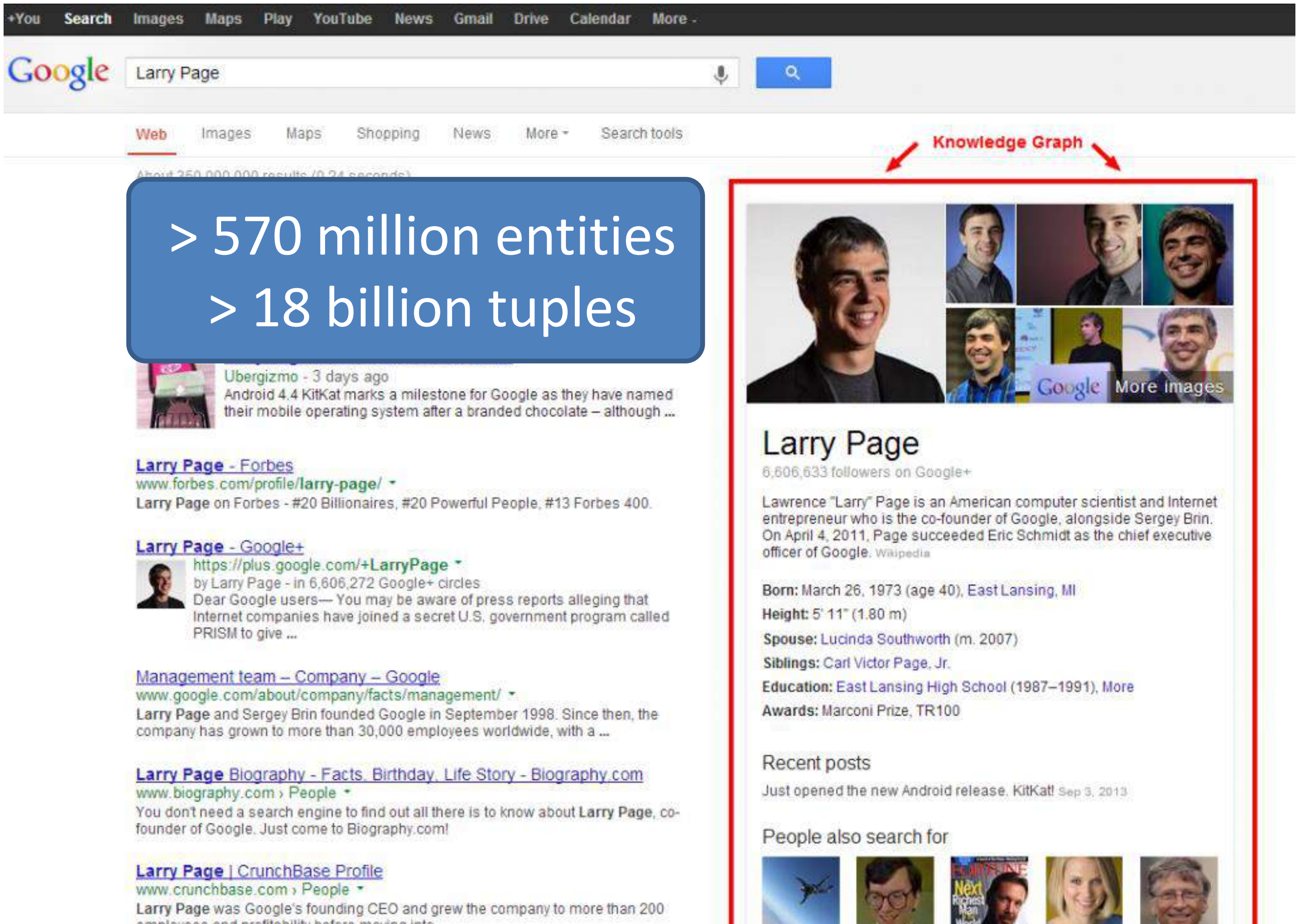
3.5 Smokes(x) ⇒ Cough(x)

1.9 Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

1.5 Asthma (x) ∧ Family(x,y) ⇒ Asthma (y)

Google Knowledge Graph

> 570 million entities
> 18 billion tuples



What we'd like to do…

Has anyone published a paper with both Erdos and Einstein

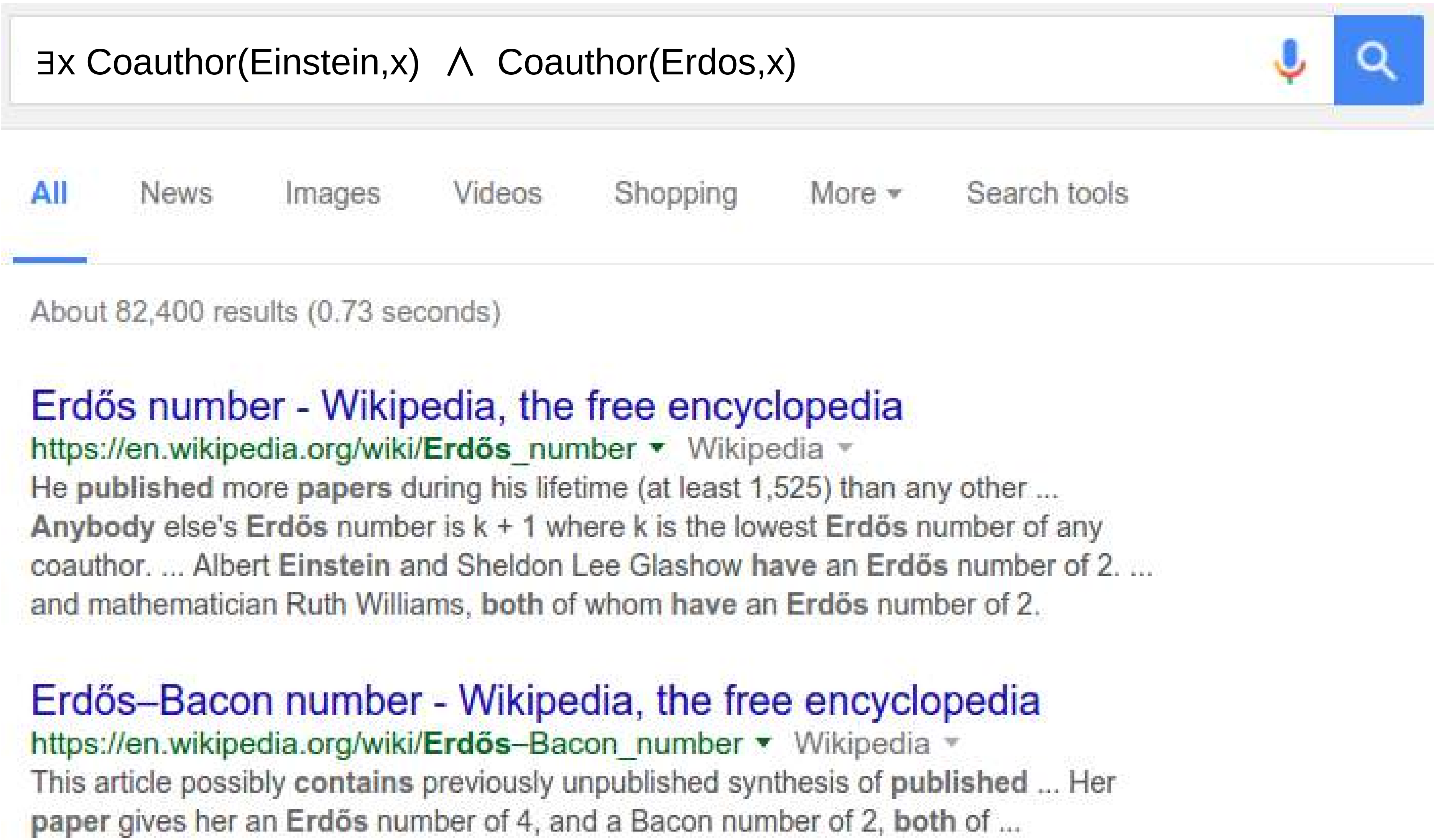
All News Images Videos Shopping More Search tools

About 82,400 results (0.73 seconds)

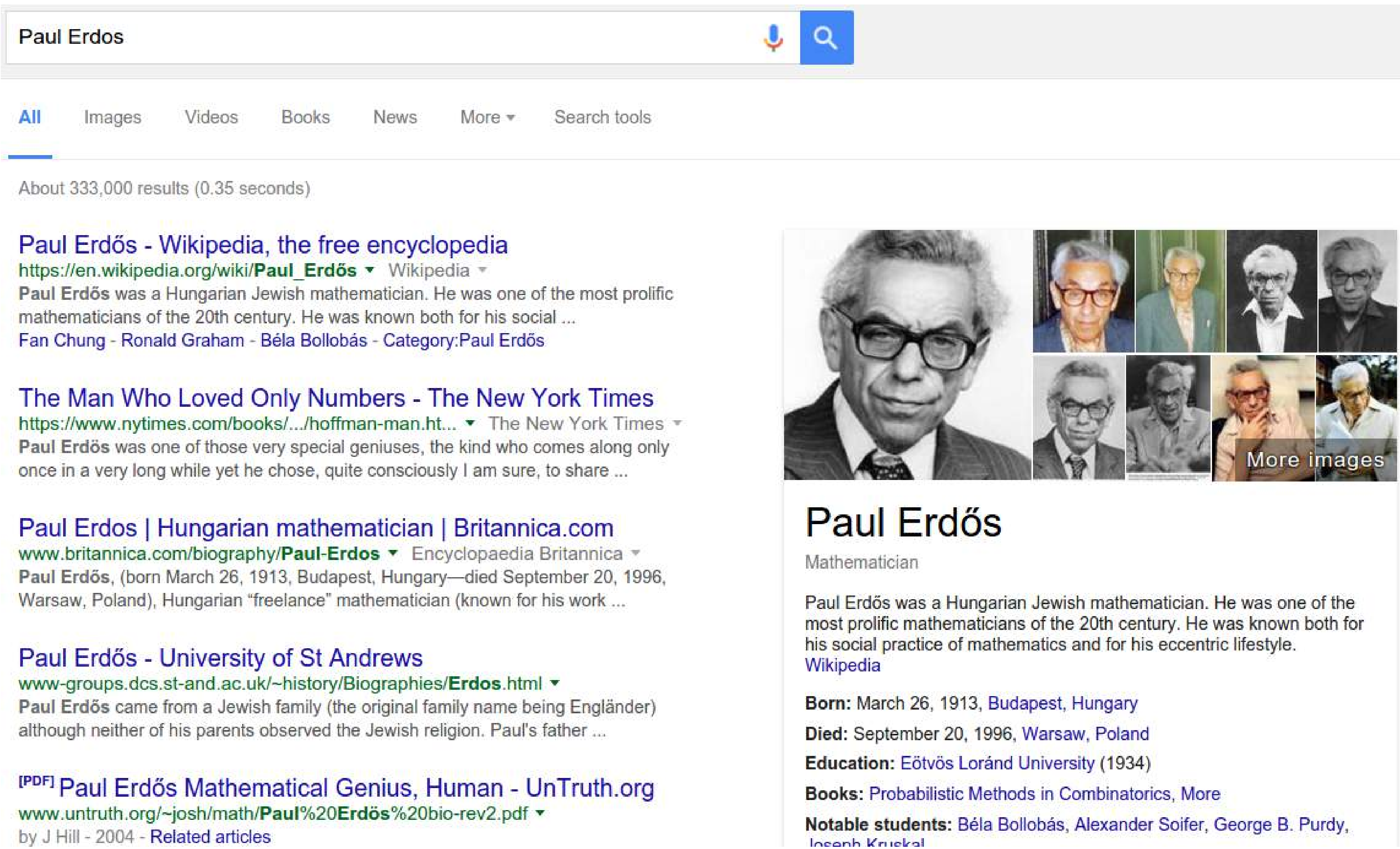
Erdős number - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/**Erdős**_number Wikipedia
He **published** more **papers** during his lifetime (at least 1,525) than any other ... **Anybody** else's **Erdős** number is k + 1 where k is the lowest **Erdős** number of any coauthor. ... Albert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

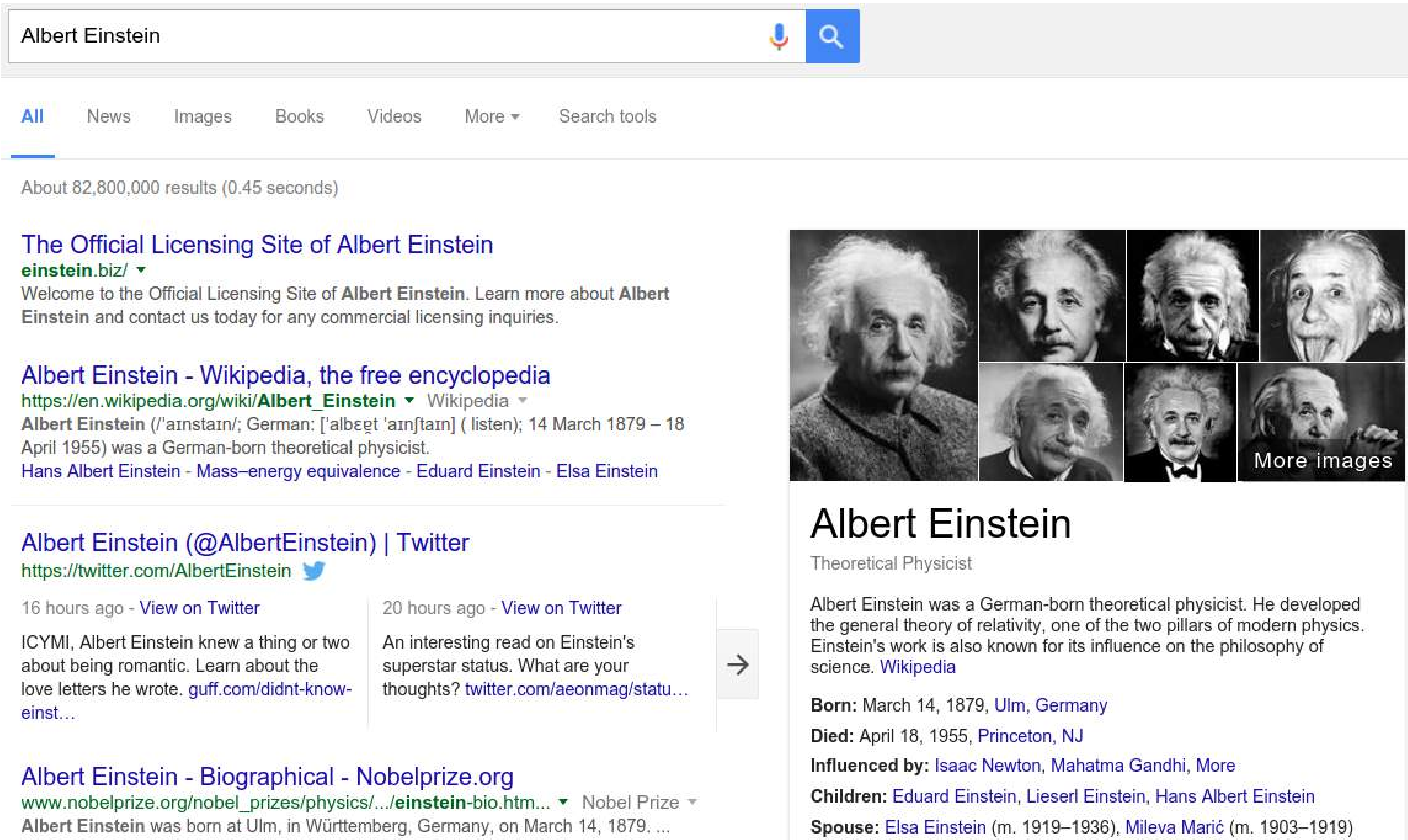
What we'd like to do…



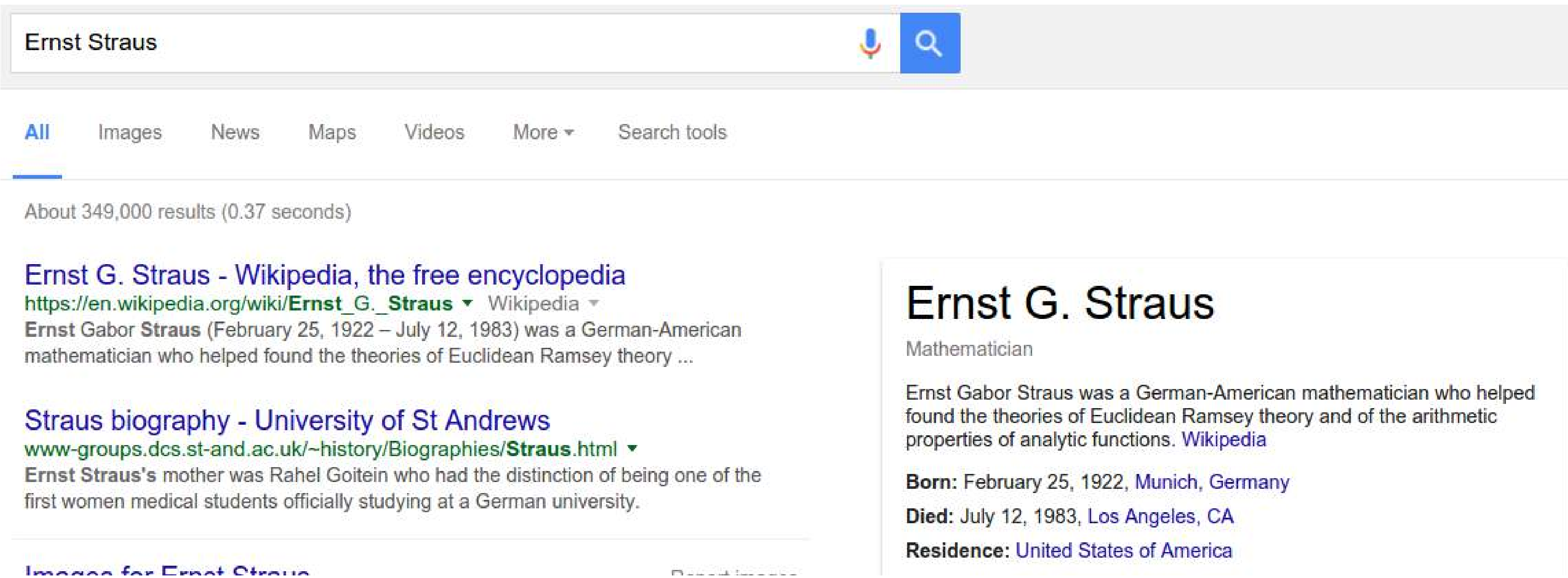
Erdős is in the Knowledge Graph



Einstein 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, …

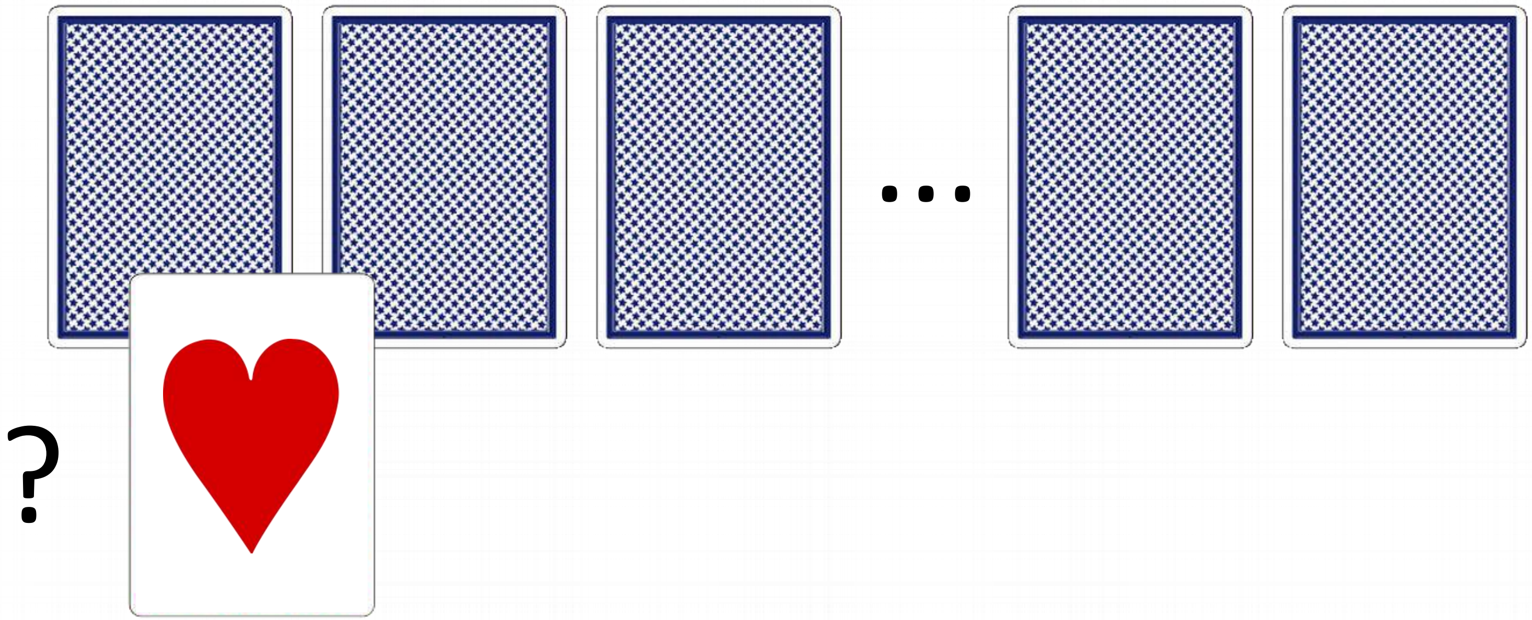
Justin Bieber, …

- Cannot come from labeled data
- Fuse uncertain information from many web pages

⇒ Embrace probability!

Why do we need
first-order probabilistic
***reasoning**?*

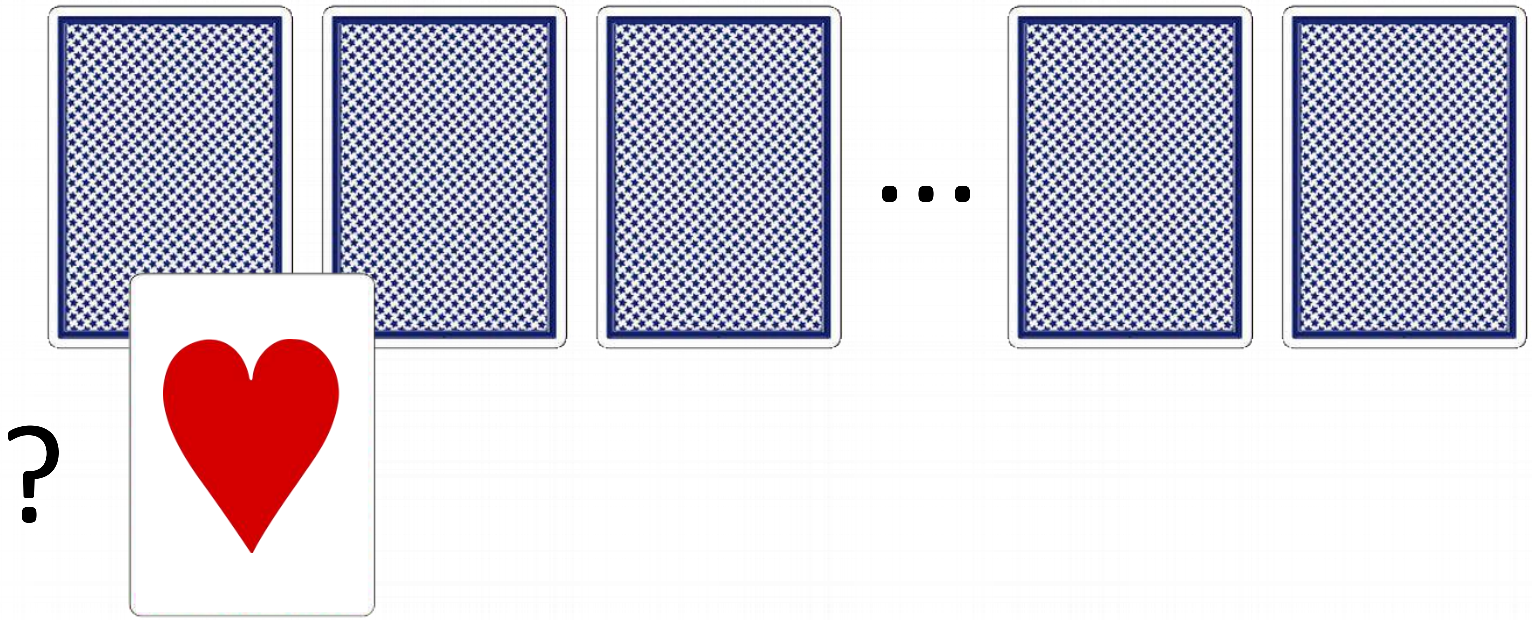
A Simple Reasoning Problem



Probability that Card1 is Hearts?

[Van den Broeck; AAAI-KRR'15]

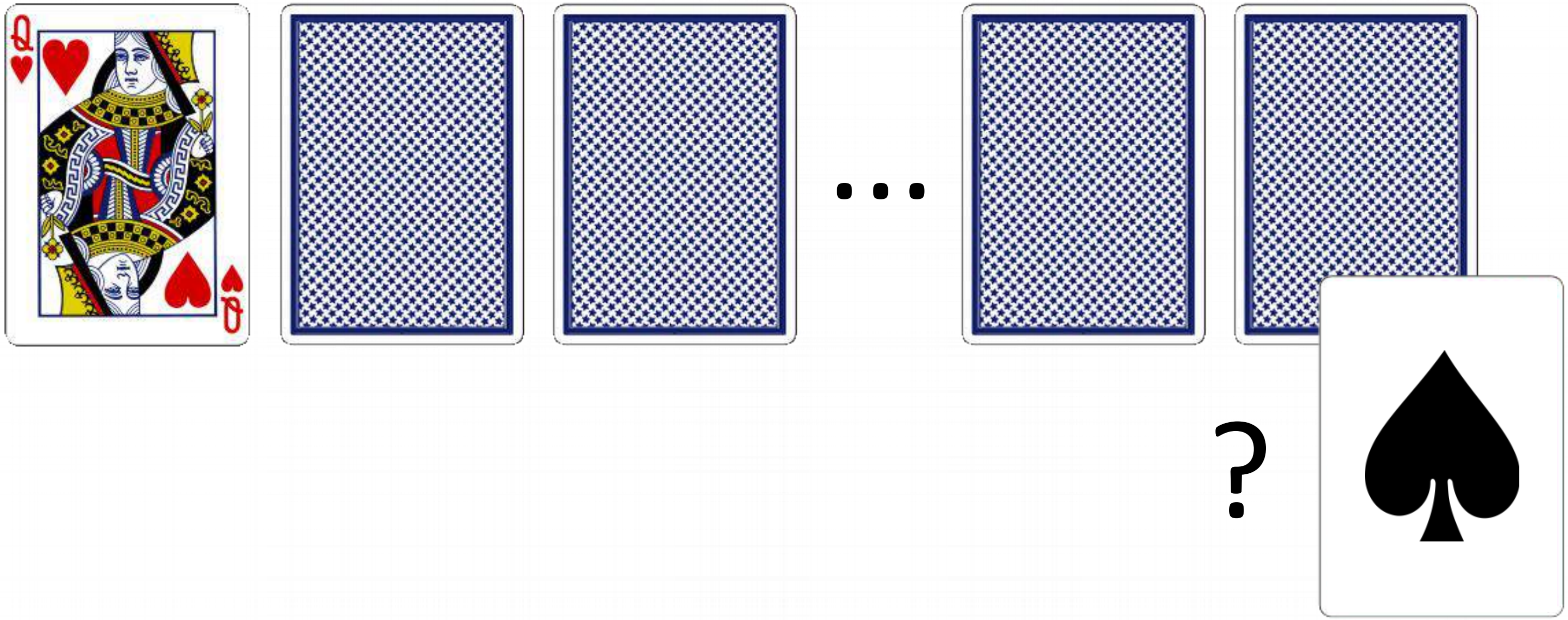
A Simple Reasoning Problem



Probability that Card1 is Hearts? 1/4

[Van den Broeck; AAAI-KRR'15]

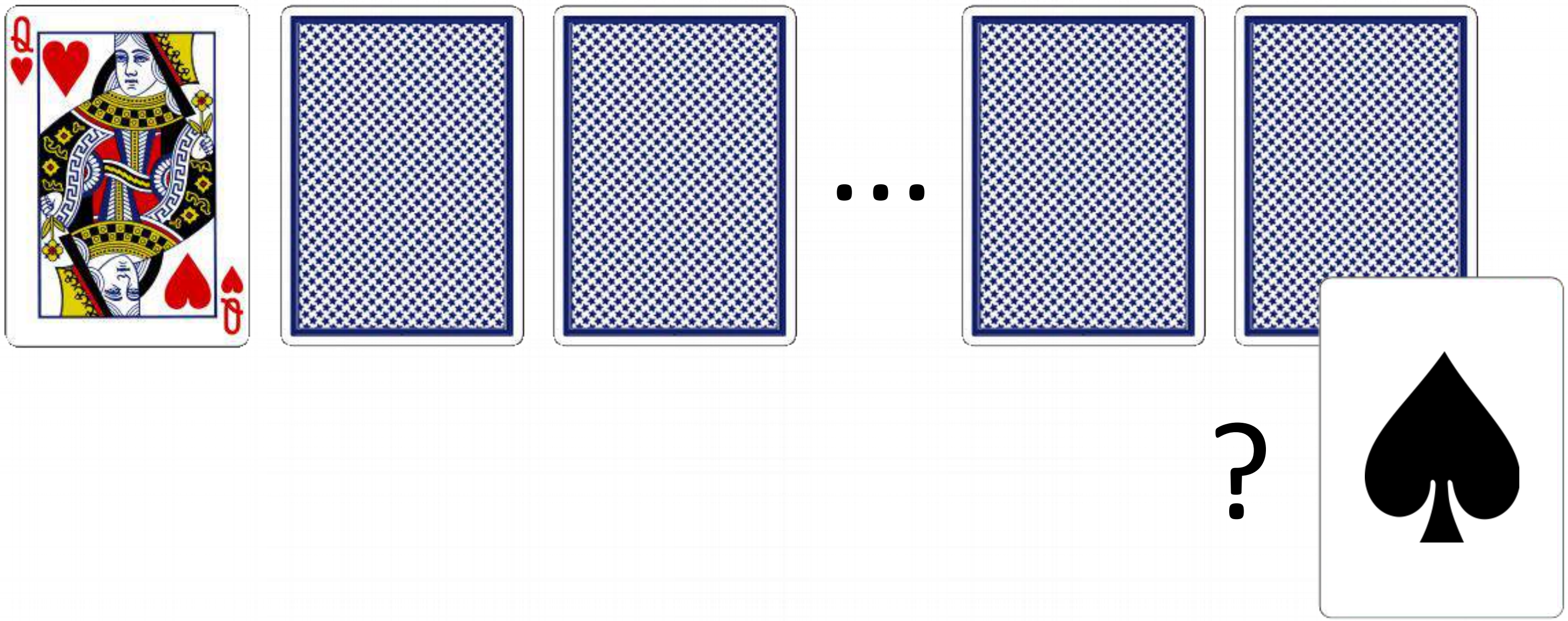
A Simple Reasoning Problem



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?

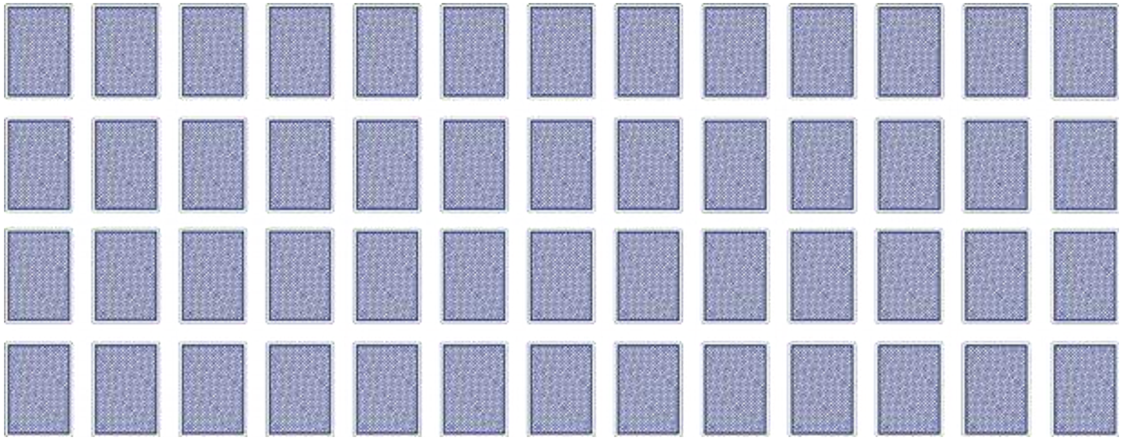
13/51

[Van den Broeck; AAAI-KRR'15]

Automated Reasoning

Let us automate this:

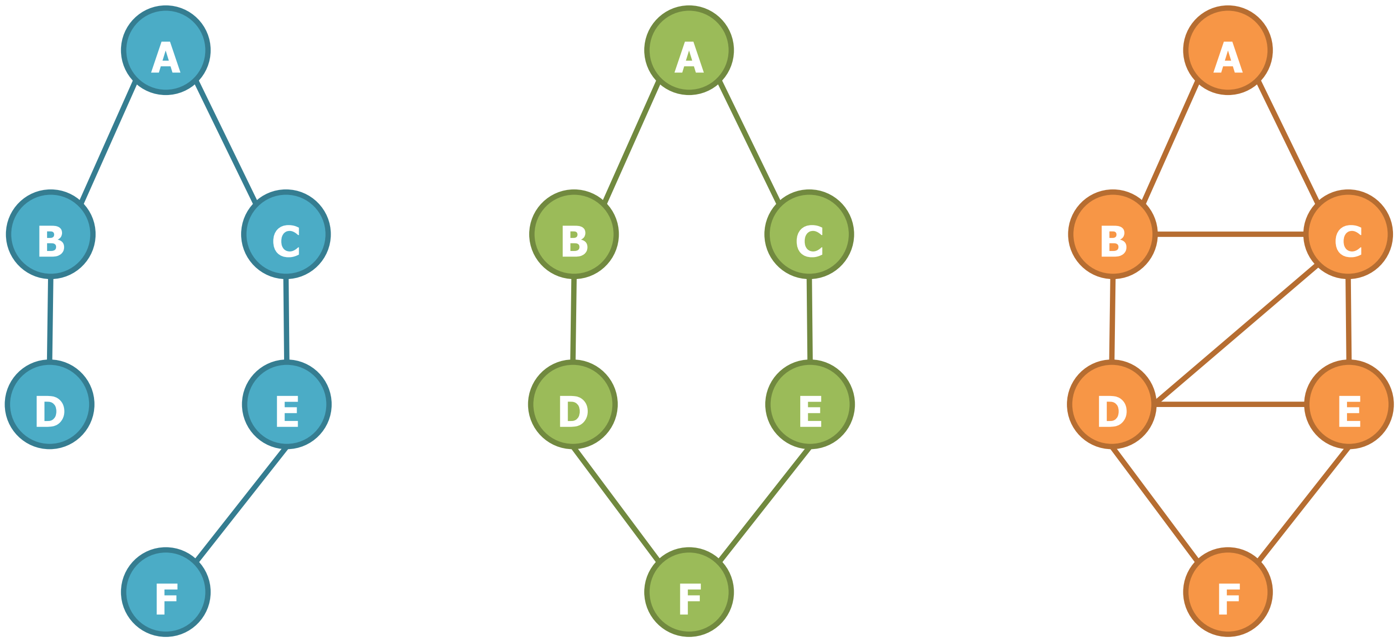
1. Probabilistic graphical model (e.g., factor graph)



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

[Van den Broeck; AAAI-KRR'15]

Classical Reasoning



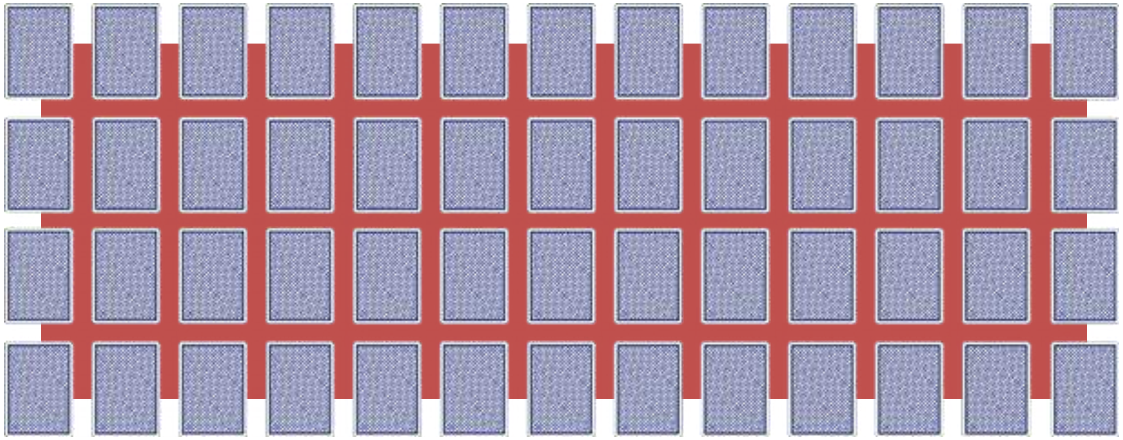
Tree *Sparse Graph* *Dense Graph*

- 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

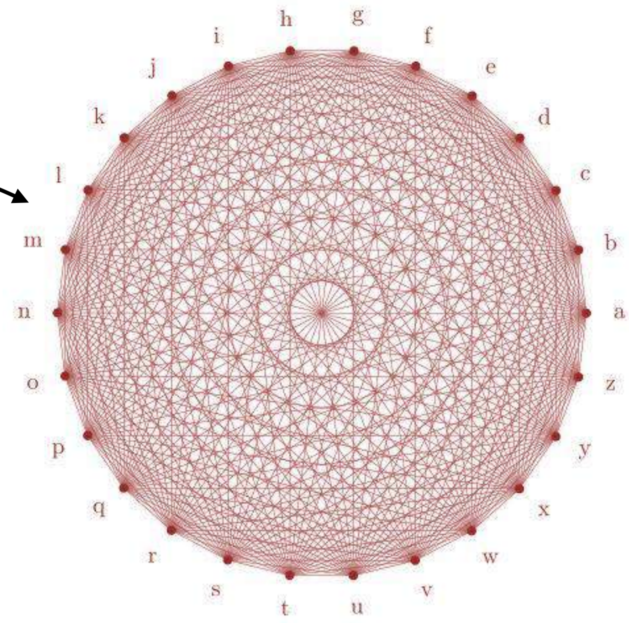
[Van den Broeck; AAAI-KRR'15]

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!

*How does lifted inference **work**?*

Uncertainty in AI

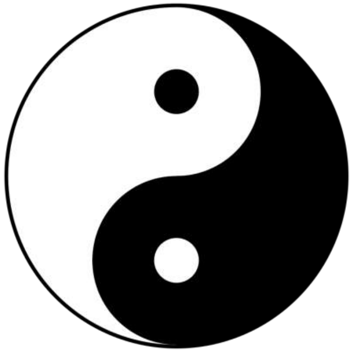
Probability Distribution

=

Qualitative

+

Quantitative



Probabilistic Graphical Models

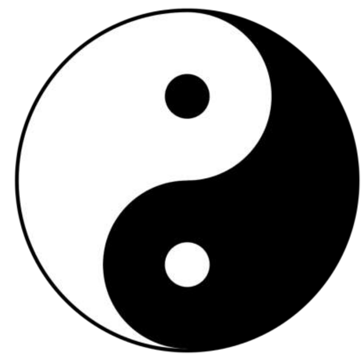
Probability Distribution

=

Graph Structure

+

Parameterization



Probabilistic Graphical Models

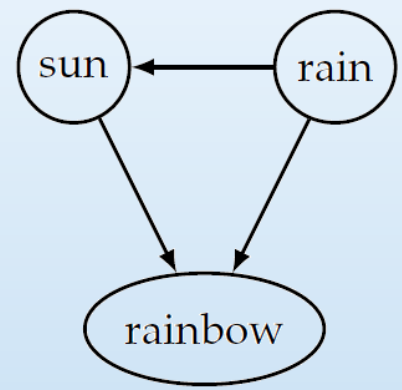
Probability Distribution

=

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

Model Counting

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

Δ = (Rain ⇒ Cloudy)

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

+ ———

#SAT = 3

Weighted Model Counting

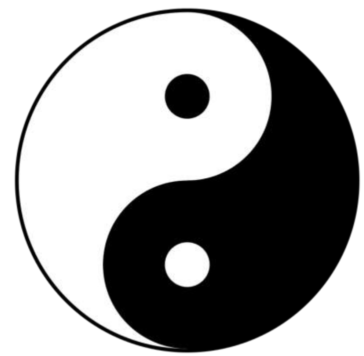
Probability Distribution

=

SAT Formula

+

Weights



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

Weighted Model Counting

Probability Distribution

=

SAT Formula

+

Weights

Rain ⇒ Cloudy

Sun ∧ Rain ⇒ Rainbow

+

w(Rain)=1

w(¬Rain)=2

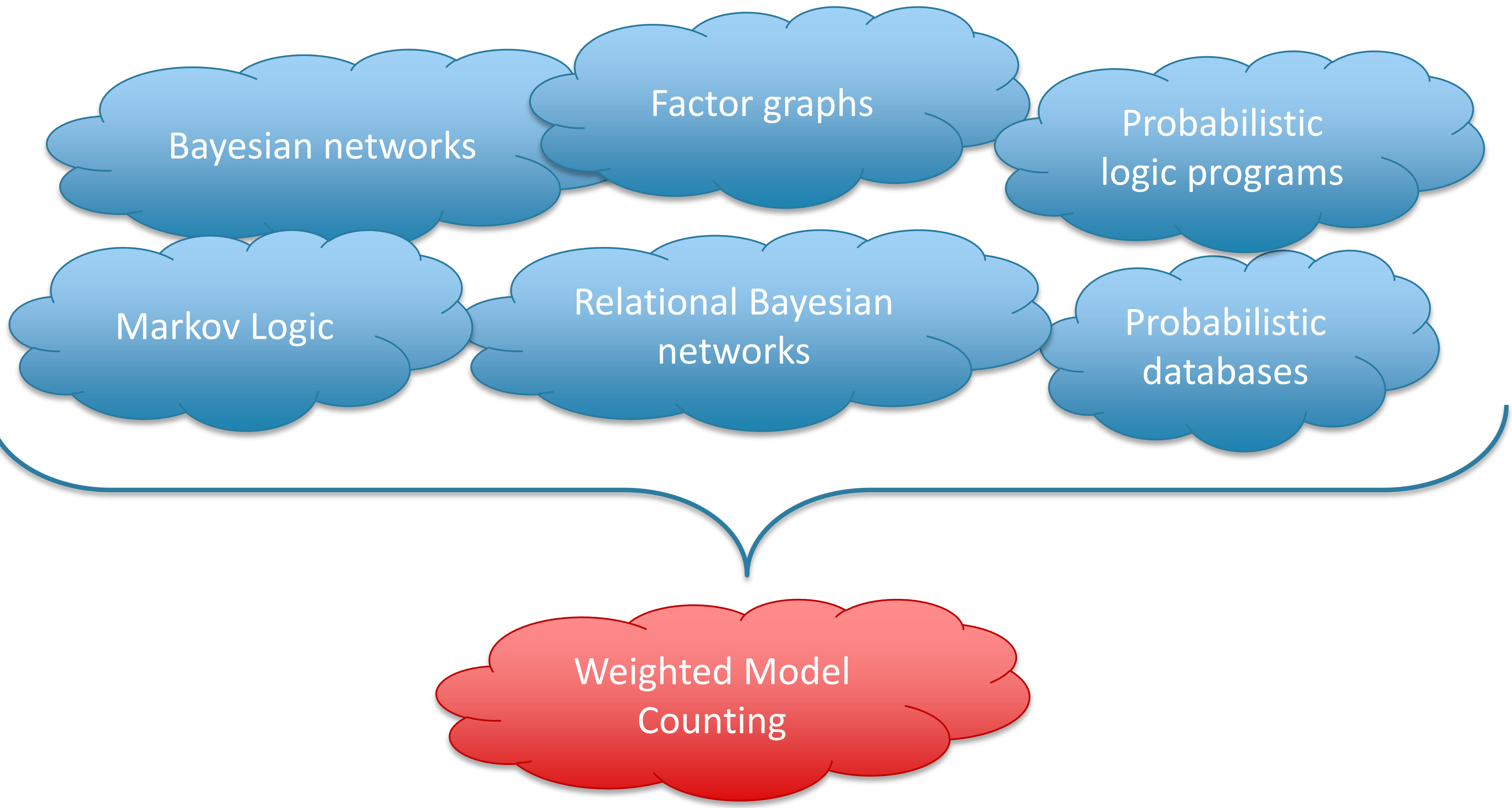
w(Cloudy)=3

w(¬Cloudy)=5

...

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

Assembly language for probabilistic reasoning



[Chavira 2006, Chavira 2008, Sang 2005, Fierens 2015]

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

Δ = ∀d (Rain(d)
⇒ 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

Weighted First-Order Model Counting

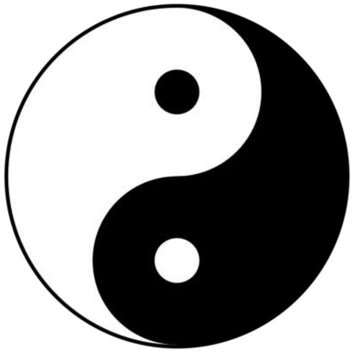
Probability Distribution

=

First-Order Logic

+

Weights



[Van den Broeck 2011, 2013, Gogate 2011]

Weighted First-Order Model Counting

Probability Distribution

=

First-Order Logic

+

Weights

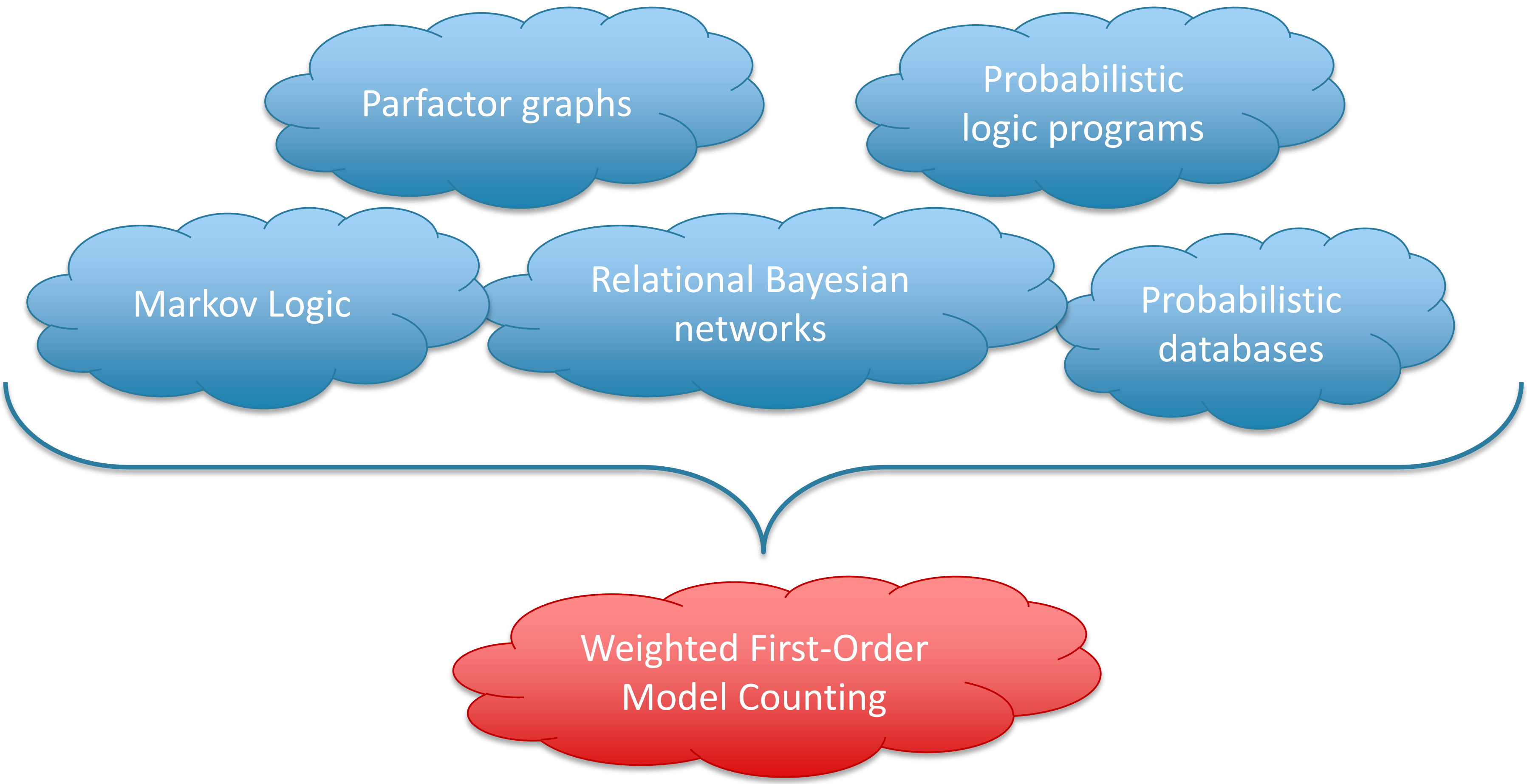
Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

+

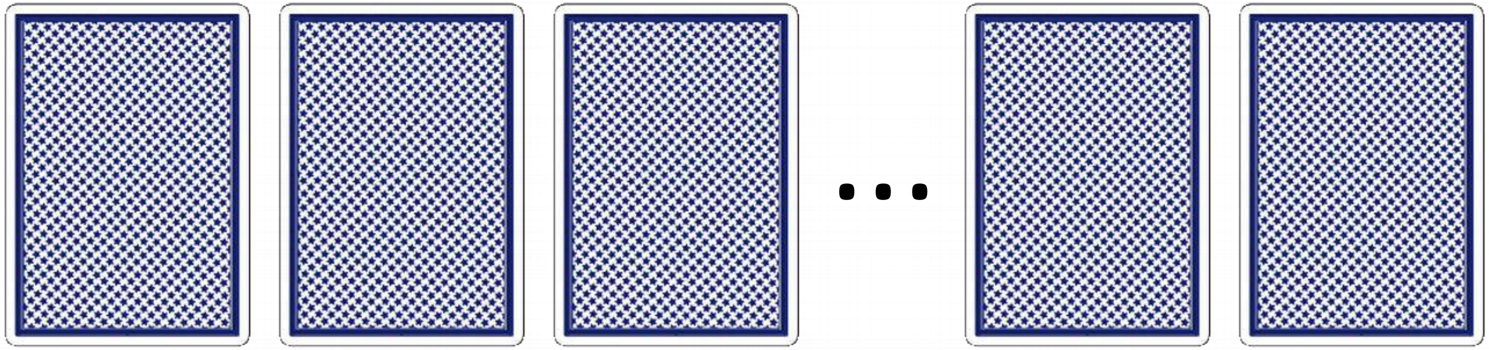
w(Smokes(a))=1
w(¬Smokes(a))=2
w(Smokes(b))=1
w(¬Smokes(b))=2
w(Friends(a,b))=3
w(¬Friends(a,b))=5
…

[Van den Broeck 2011, 2013, Gogate 2011]

Assembly language for first-order probabilistic reasoning



[Van den Broeck 2011, 2013, Gogate 2011, Gribkoff 2014]



Let us automate this:

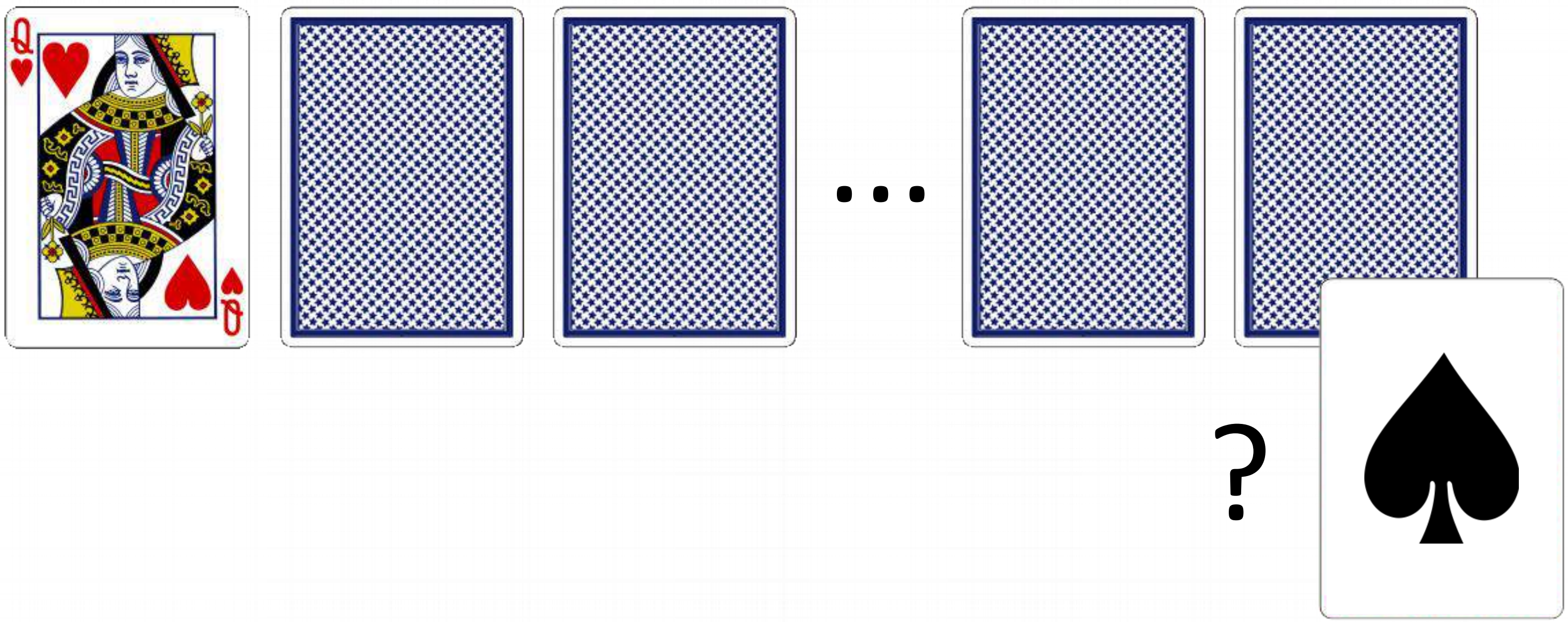
- **Relational** model

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

- **Lifted** probabilistic inference algorithm

[Van den Broeck; AAAI-KRR'15]

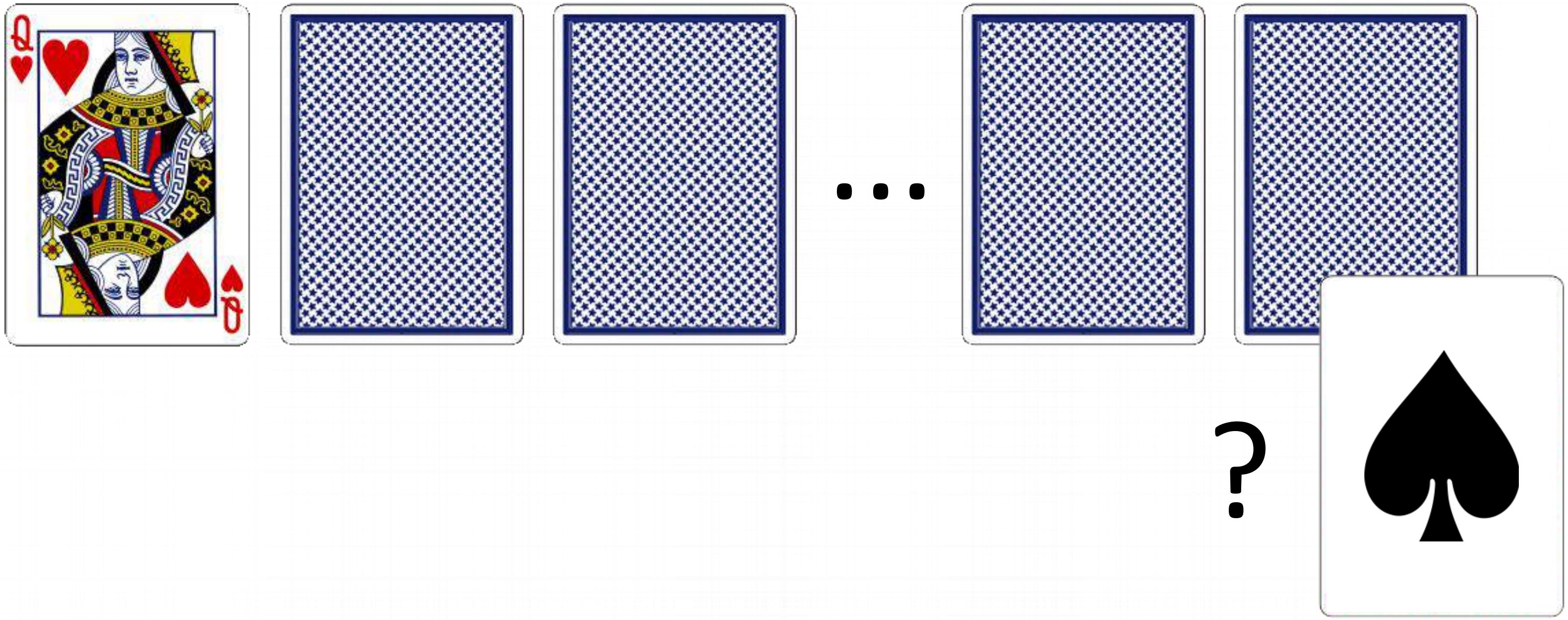
What's Going On Here?



Probability that Card52 is Spades given that Card1 is QH?

[Van den Broeck; AAAI-KRR’15]

What's Going On Here?

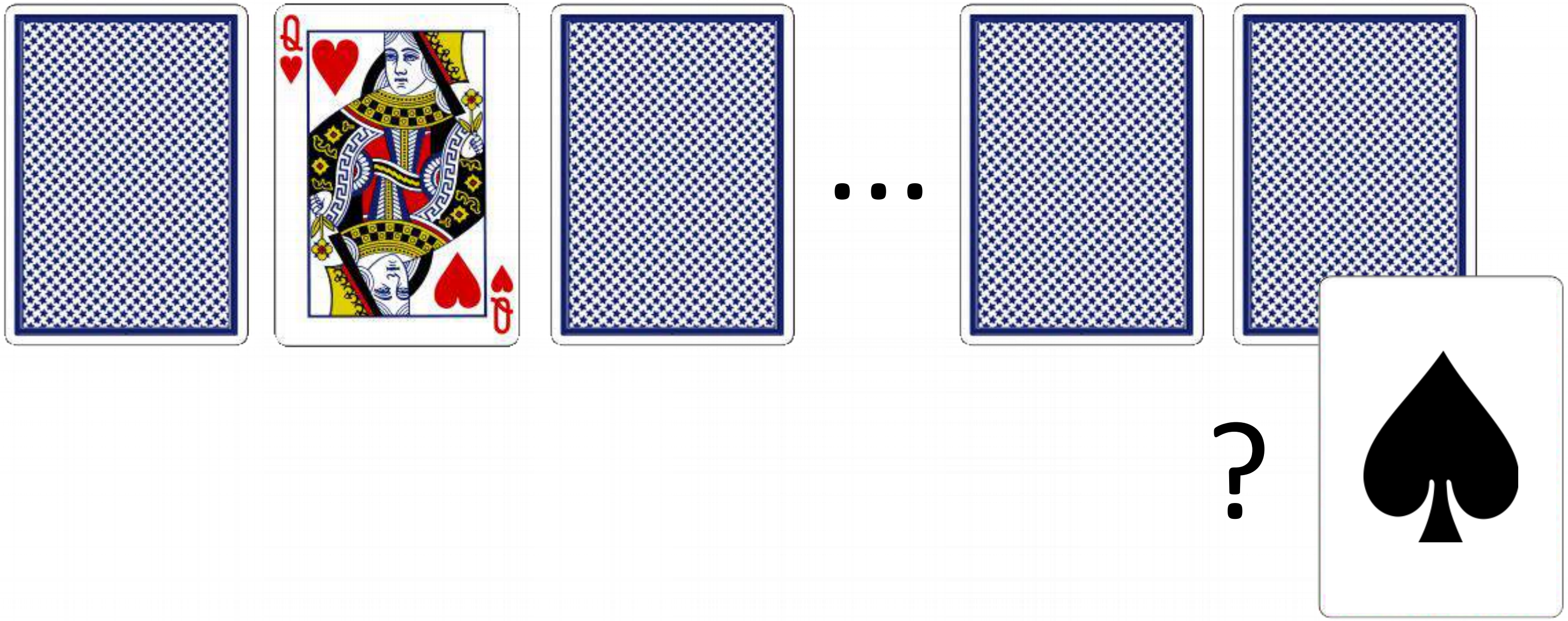


Probability that Card52 is Spades given that Card1 is QH?

13/51

[Van den Broeck; AAAI-KRR'15]

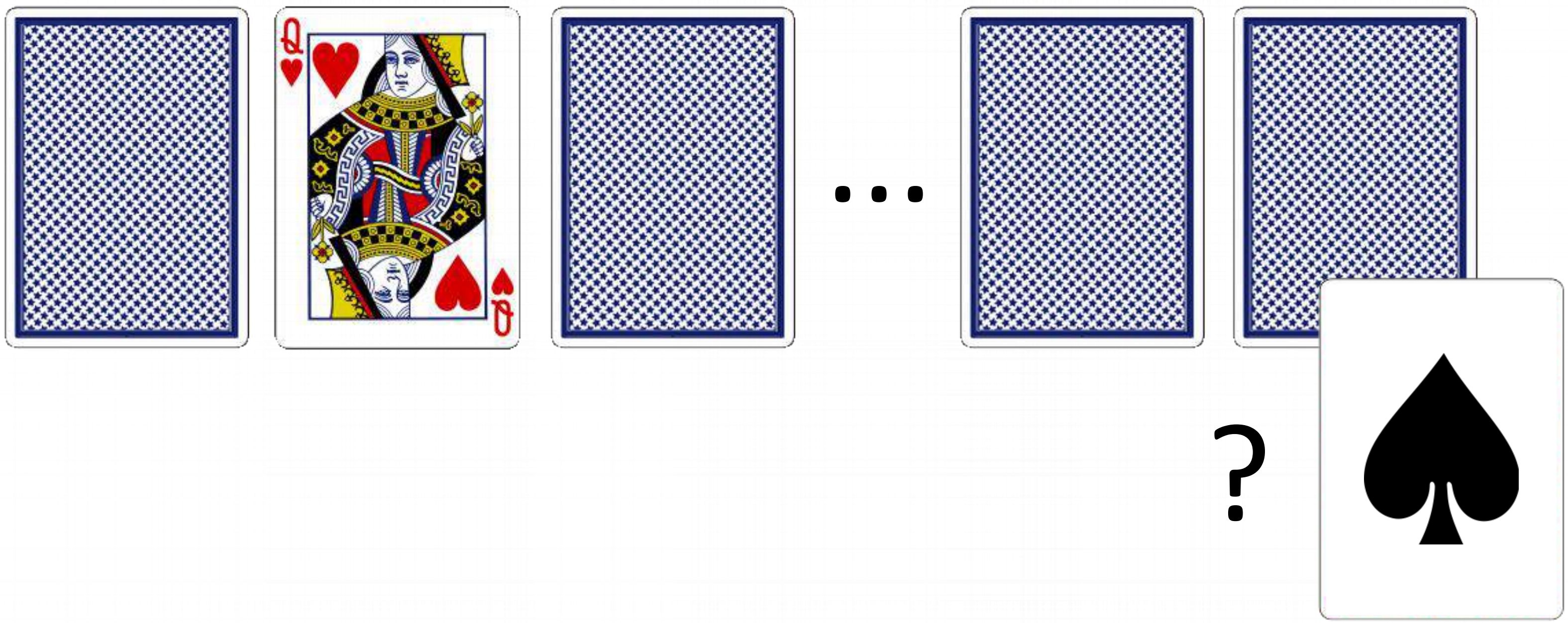
What's Going On Here?



Probability that Card52 is Spades
given that Card2 is QH?

[Van den Broeck; AAAI-KRR'15]

What's Going On Here?

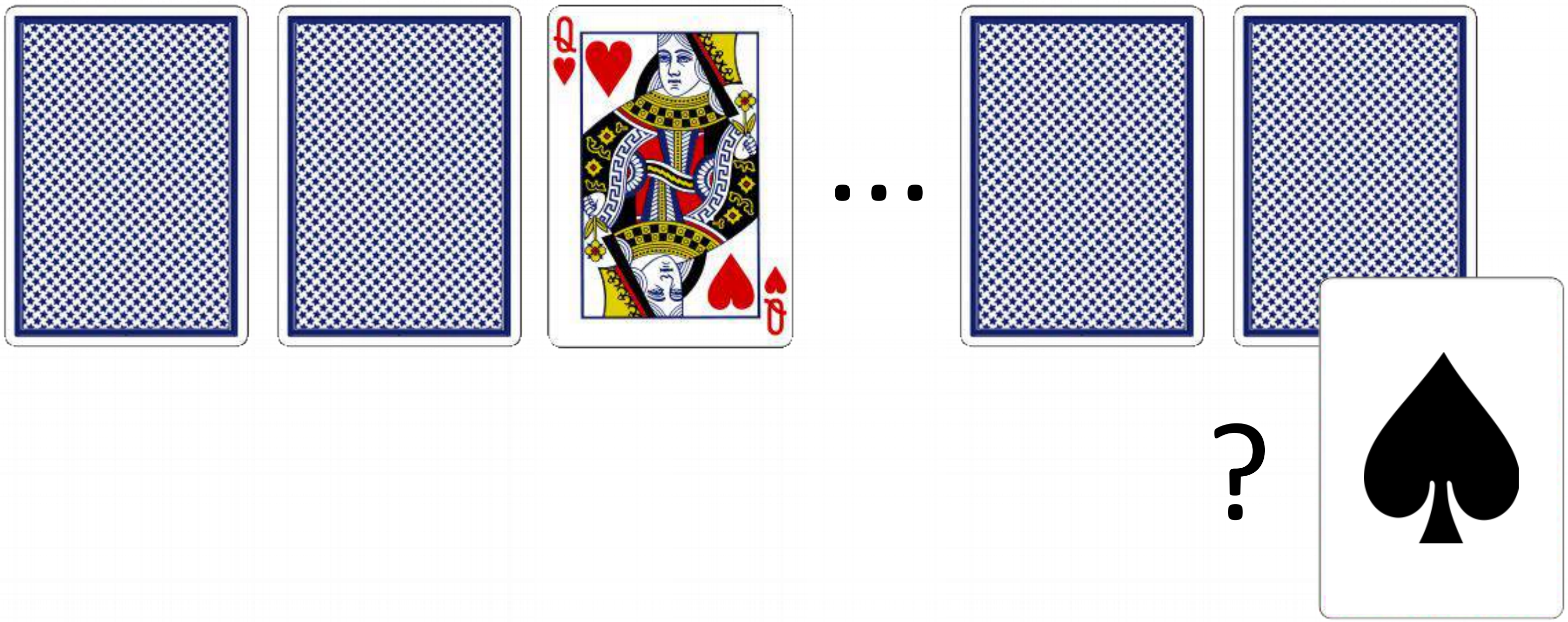


Probability that Card52 is Spades given that Card2 is QH?

13/51

[Van den Broeck; AAAI-KRR’15]

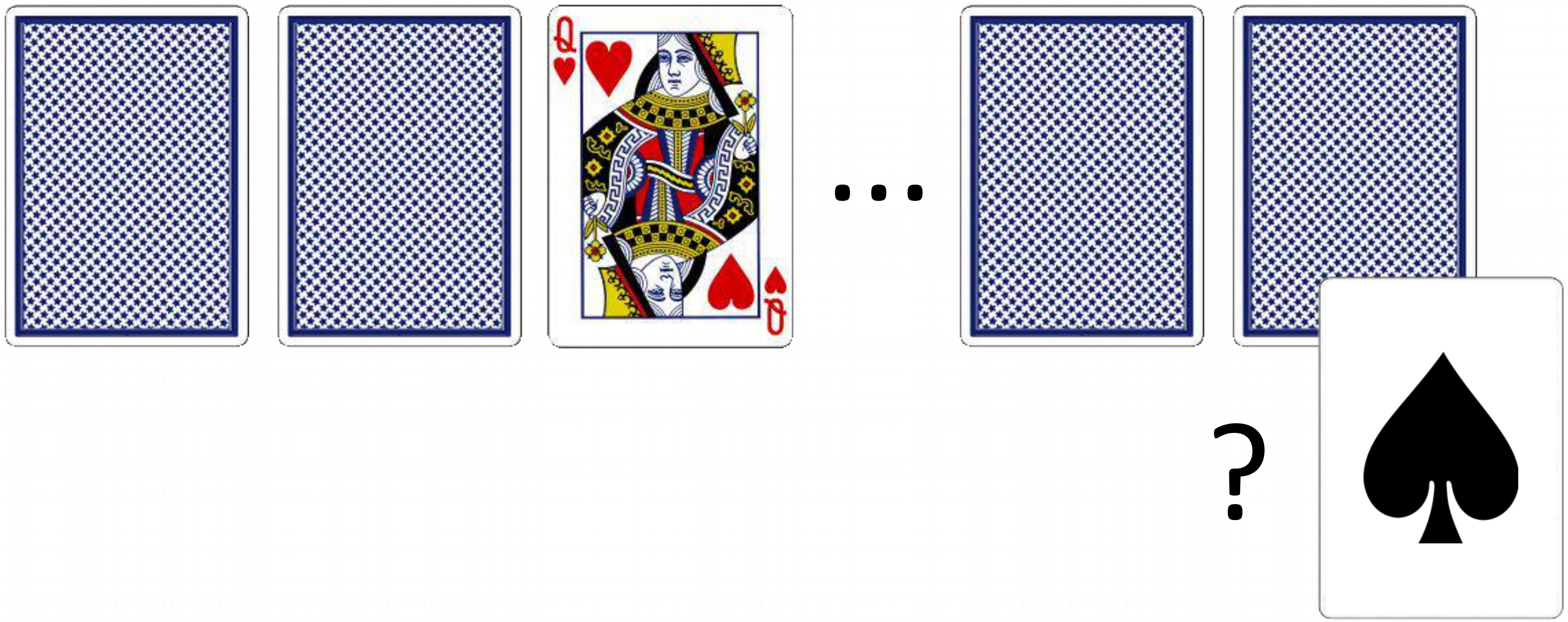
What's Going On Here?



Probability that Card52 is Spades given that Card3 is QH?

[Van den Broeck; AAAI-KRR'15]

What's Going On Here?

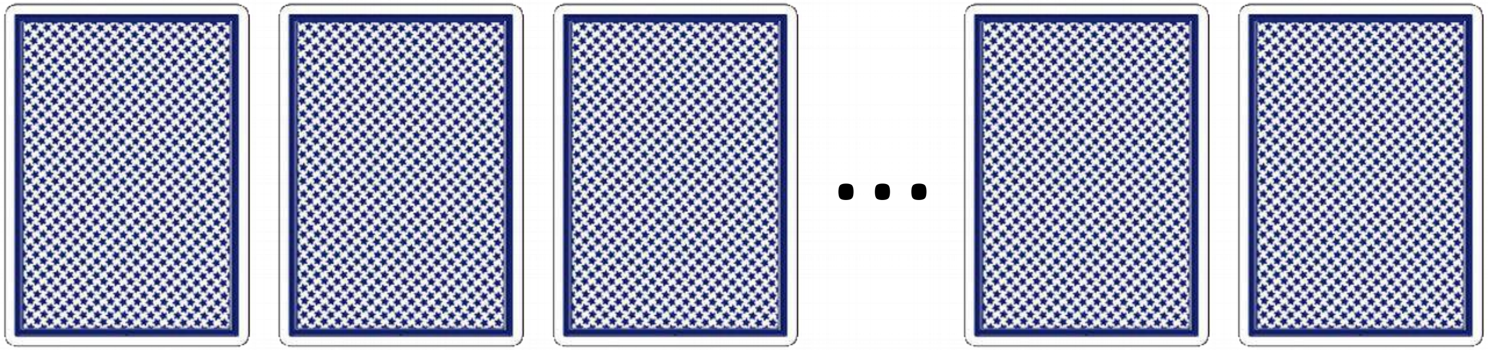


Probability that Card52 is Spades given that Card3 is QH?

13/51

[Van den Broeck; AAAI-KRR'15]

Tractable Reasoning

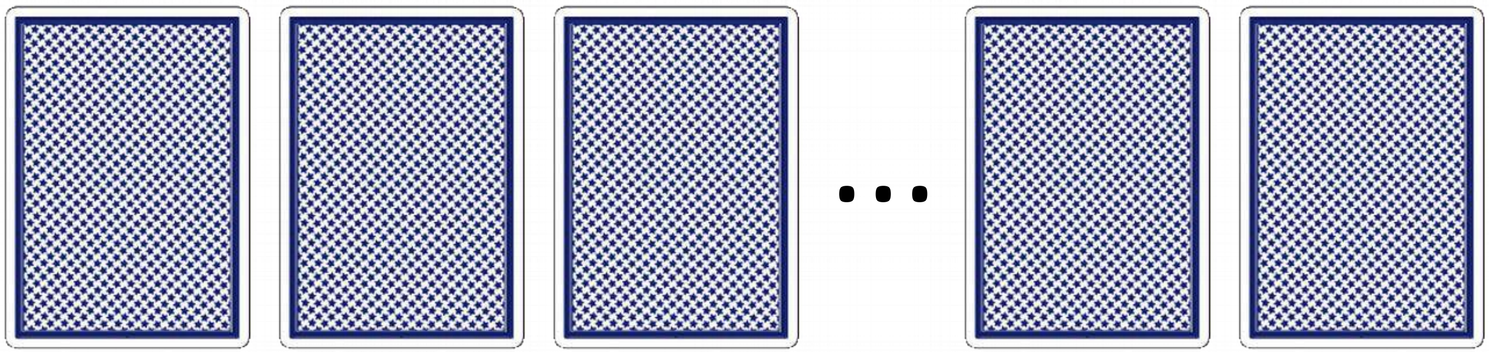


What's going on here?

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]

FOMC Inference

3. $\Delta = \forall x, (Stress(x) \Rightarrow Smokes(x))$

Domain = {n people}

FOMC Inference

3. Δ = ∀x, (Stress(x) ⇒ Smokes(x))

Domain = {n people}

→ 3^n models

2. Δ = ∀y, (ParentOf(y) ∧ Female ⇒ MotherOf(y))

D = {n people}

FOMC Inference

3. $\Delta = \forall x, (Stress(x) \Rightarrow Smokes(x))$

Domain = {n people}

→ 3^n models

2. $\Delta = \forall y, (ParentOf(y) \land Female \Rightarrow MotherOf(y))$

D = {n people}

If Female = true?

$\Delta = \forall y, (ParentOf(y) \Rightarrow MotherOf(y))$

→ 3^n models

FOMC Inference

3. $\Delta = \forall x, (\text{Stress}(x) \Rightarrow \text{Smokes}(x))$

Domain = {n people}

→ 3^n models

2. $\Delta = \forall y, (\text{ParentOf}(y) \wedge \text{Female} \Rightarrow \text{MotherOf}(y))$

D = {n people}

If Female = true? $\Delta = \forall y, (\text{ParentOf}(y) \Rightarrow \text{MotherOf}(y))$ → 3^n models

If Female = false? $\Delta = \text{true}$ → 4^n models

FOMC Inference

3. $\Delta = \forall x, (Stress(x) \Rightarrow Smokes(x))$

Domain = {n people}

→ 3^n models

2. $\Delta = \forall y, (ParentOf(y) \wedge Female \Rightarrow MotherOf(y))$

D = {n people}

If Female = true? $\Delta = \forall y, (ParentOf(y) \Rightarrow MotherOf(y))$ → 3^n models

If Female = false? Δ = true → 4^n models

→ $3^n + 4^n$ models

1. $\Delta = \forall x,y, (ParentOf(x,y) \wedge Female(x) \Rightarrow MotherOf(x,y))$

D = {n people}

FOMC Inference

3. $\Delta = \forall x, (\text{Stress}(x) \Rightarrow \text{Smokes}(x))$

Domain = {n people}

→ 3^n models

2. $\Delta = \forall y, (\text{ParentOf}(y) \wedge \text{Female} \Rightarrow \text{MotherOf}(y))$

D = {n people}

If Female = true? $\Delta = \forall y, (\text{ParentOf}(y) \Rightarrow \text{MotherOf}(y))$ → 3^n models

If Female = false? $\Delta = \text{true}$ → 4^n models

→ $3^n + 4^n$ models

1. $\Delta = \forall x,y, (\text{ParentOf}(x,y) \wedge \text{Female}(x) \Rightarrow \text{MotherOf}(x,y))$

D = {n people}

→ $(3^n + 4^n)^n$ models

FOMC Inference

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

Domain = {n people}

FOMC Inference

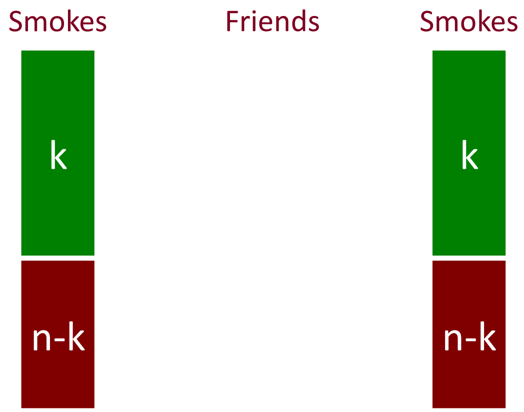
Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

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

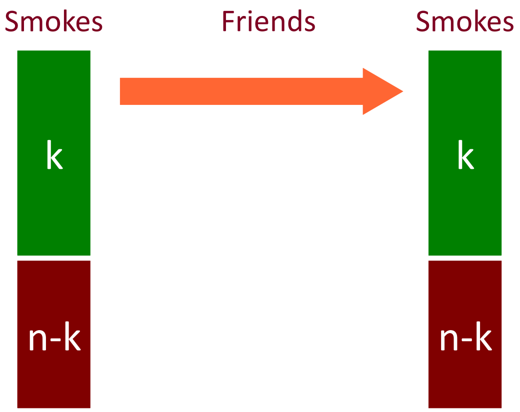
$\Delta = \forall x,y, (Smokes(x) \wedge Friends(x,y) \Rightarrow Smokes(y))$

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

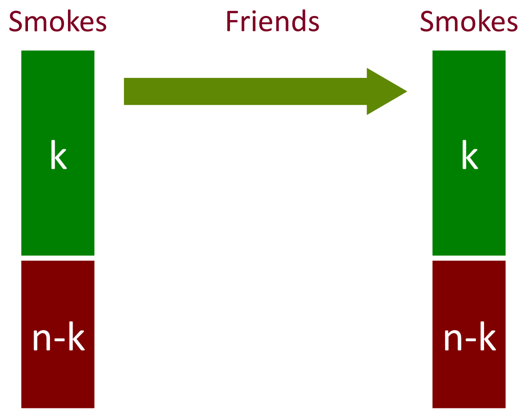
$\Delta = \forall x,y, (Smokes(x) \land Friends(x,y) \Rightarrow Smokes(y))$

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

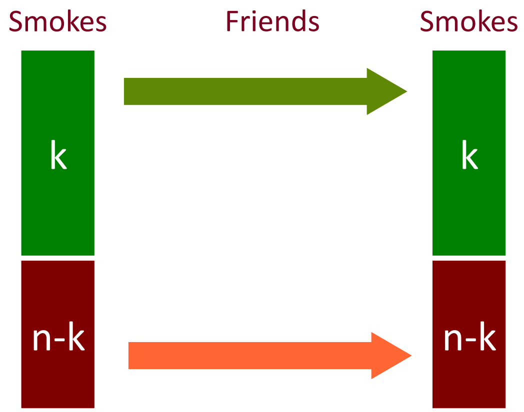
$\Delta = \forall x,y, (\text{Smokes}(x) \wedge \text{Friends}(x,y) \Rightarrow \text{Smokes}(y))$

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

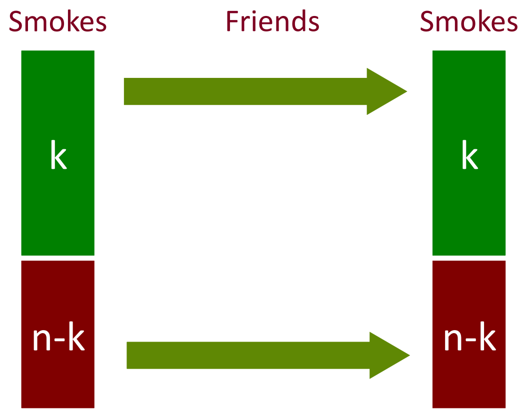
$\Delta = \forall x,y, (\text{Smokes}(x) \wedge \text{Friends}(x,y) \Rightarrow \text{Smokes}(y))$

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

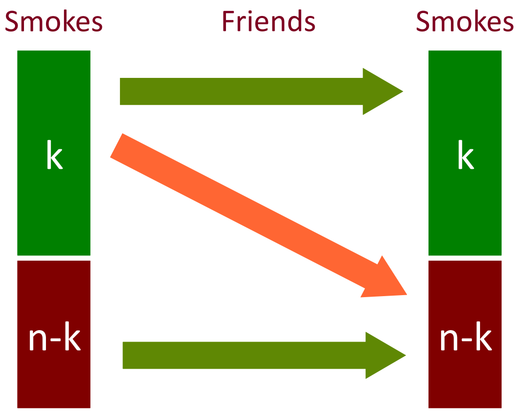
Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

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

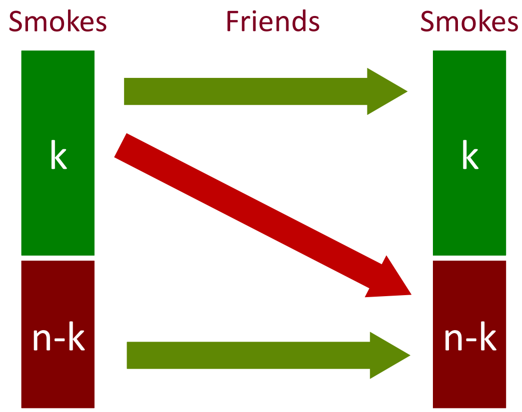
Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

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

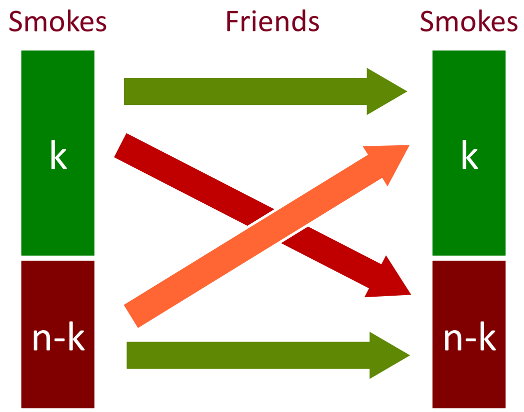
Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

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

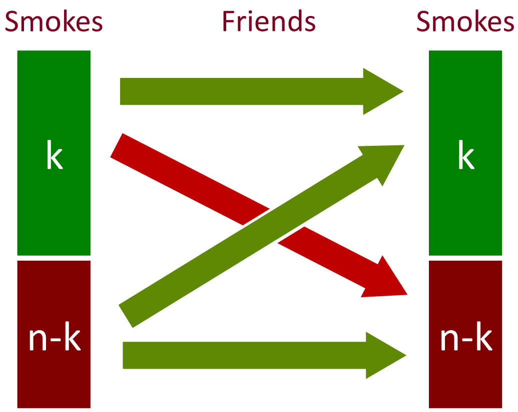
Δ = ∀x,y, (Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y))

Domain = {n people}

- If we know precisely who smokes, and there 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))

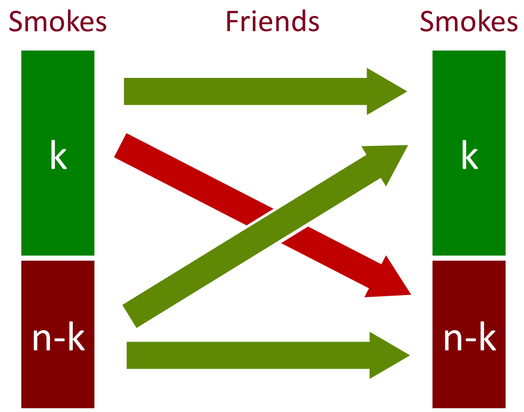
Domain = {n people}

- 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
...

→ $2^{n^2 - k(n-k)}$ models



FOMC Inference

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

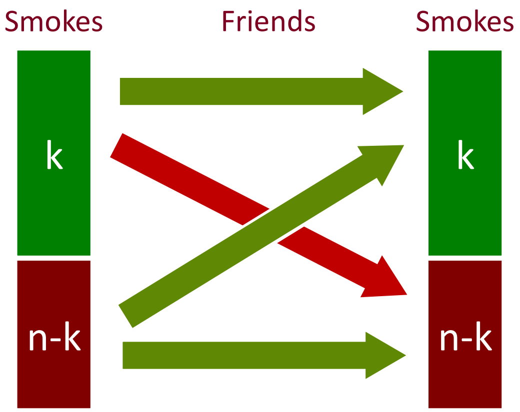
Domain = {n people}

- 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?

FOMC Inference

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

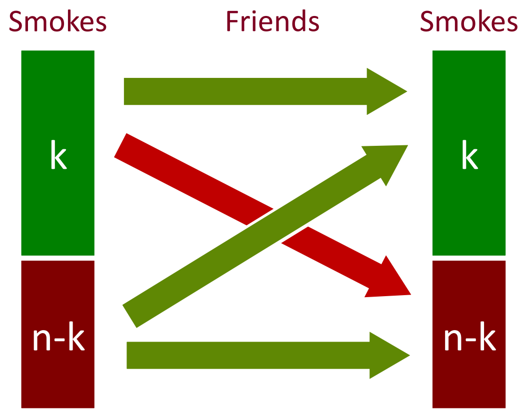
Domain = {n people}

- 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))

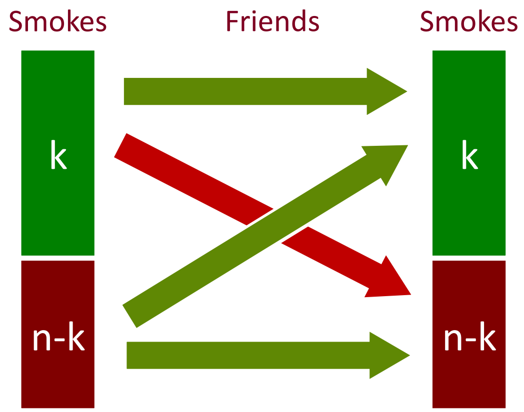
Domain = {n people}

- 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))

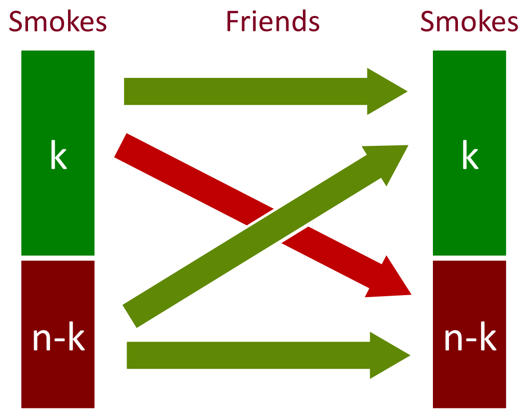
Domain = {n people}

- 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

What are the ***successes****?*

First-Order Knowledge Compilation

Markov Logic

3.14	Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

[Van den Broeck,PhD'13]

First-Order Knowledge Compilation

Markov Logic

3.14 Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

Weight Function

w(Smokes)=1
w(¬Smokes)=1
w(Friends)=1
w(¬Friends)=1
w(F)=3.14
w(¬F)=1

FOL Sentence

∀x,y, F(x,y) ⇔ [Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)]

[Van den Broeck,PhD'13]

First-Order Knowledge Compilation

Markov Logic

3.14 Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

Weight Function

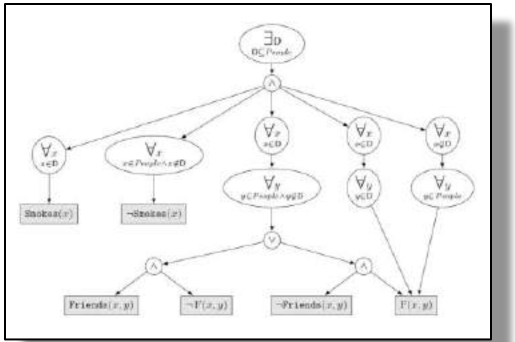
w(Smokes)=1
w(¬Smokes)=1
w(Friends)=1
w(¬Friends)=1
w(F)=3.14
w(¬F)=1

FOL Sentence

∀x,y, F(x,y) ⇔ [Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)]

Compile?

First-Order d-DNNF Circuit



[Van den Broeck,PhD'13]

First-Order Knowledge Compilation

Markov Logic

3.14 Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

Weight Function

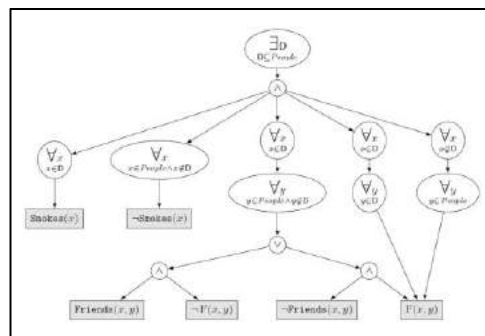
w(Smokes)=1
w(¬Smokes)=1
w(Friends)=1
w(¬Friends)=1
w(F)=3.14
w(¬F)=1

FOL Sentence

∀x,y, F(x,y) ⇔ [Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)]

Compile?

First-Order d-DNNF Circuit



Domain

Alice
Bob
Charlie

[Van den Broeck,PhD'13]

First-Order Knowledge Compilation

Markov Logic

3.14 Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

Weight Function

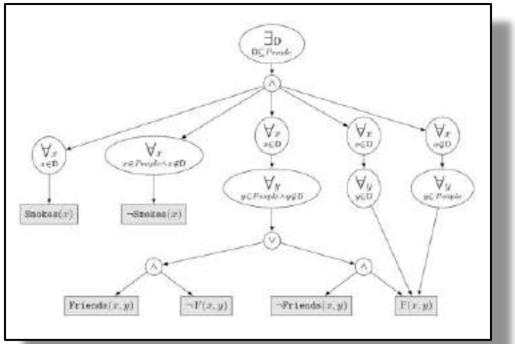
w(Smokes)=1
w(¬Smokes)=1
w(Friends)=1
w(¬Friends)=1
w(F)=3.14
w(¬F)=1

FOL Sentence

∀x,y, F(x,y) ⇔ [Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)]

Compile?

First-Order d-DNNF Circuit



Domain

Alice
Bob
Charlie

Z = WFOMC = 1479.85

[Van den Broeck,PhD’13]

First-Order Knowledge Compilation

Markov Logic: 3.14 Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

Weight Function

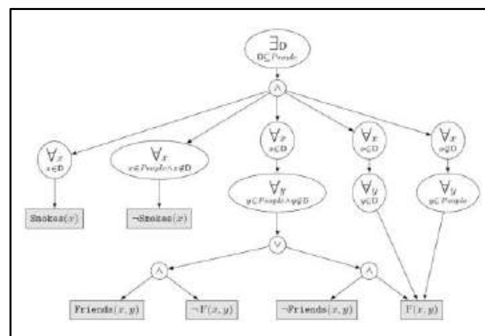
w(Smokes)=1
w(¬Smokes)=1
w(Friends)=1
w(¬Friends)=1
w(F)=3.14
w(¬F)=1

FOL Sentence

∀x,y, F(x,y) ⇔ [Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)]

Compile?

First-Order d-DNNF Circuit



Domain

Alice
Bob
Charlie

Z = WFOMC = 1479.85

Evaluation in time polynomial in domain size!

[Van den Broeck,PhD'13]

First-Order Knowledge Compilation

Markov Logic: 3.14 Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

Weight Function

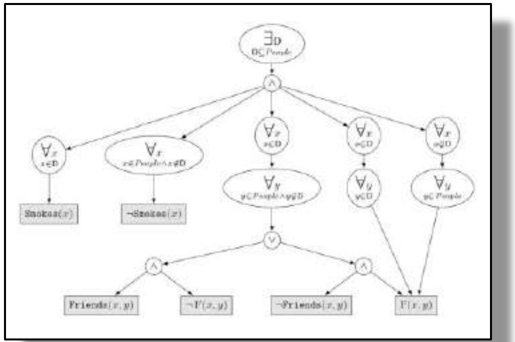
w(Smokes)=1
w(¬Smokes)=1
w(Friends)=1
w(¬Friends)=1
w(F)=3.14
w(¬F)=1

FOL Sentence

∀x,y, F(x,y) ⇔ [Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)]

Compile?

First-Order d-DNNF Circuit



Domain

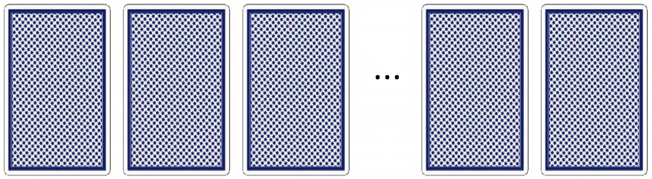
Alice
Bob
Charlie

Z = WFOMC = 1479.85

Evaluation in time polynomial in domain size! = Lifted!

[Van den Broeck,PhD'13]

Playing Cards Revisited



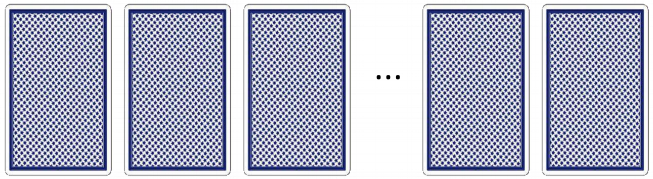
$\forall p, \exists c, Card(p,c)$

$\forall c, \exists p, Card(p,c)$

$\forall p, \forall c, \forall c', Card(p,c) \land Card(p,c') \Rightarrow c = c'$

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

Playing Cards Revisited



∀p, ∃c, Card(p,c)

∀c, ∃p, Card(p,c)

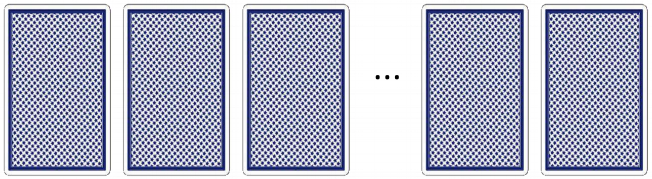
∀p, ∀c, ∀c', Card(p,c) ∧ Card(p,c') ⇒ c = c'

↓

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

[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'

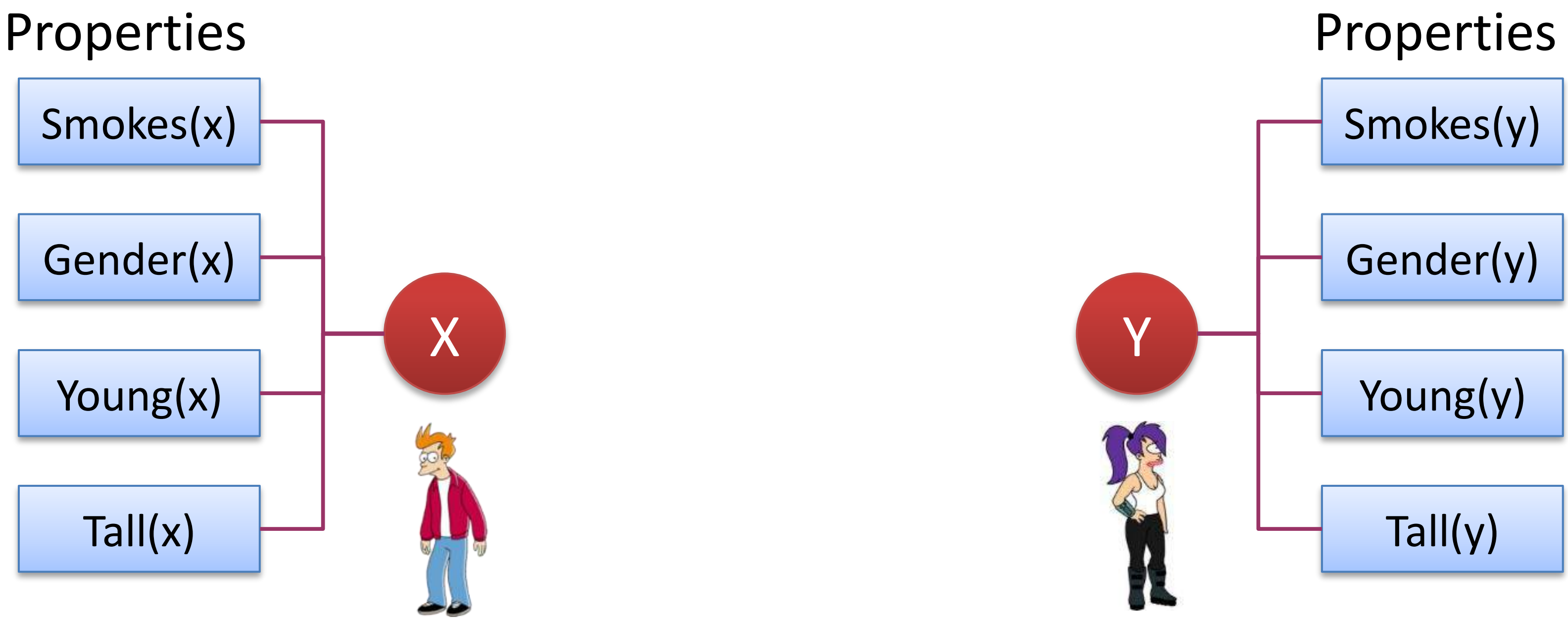
↓

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

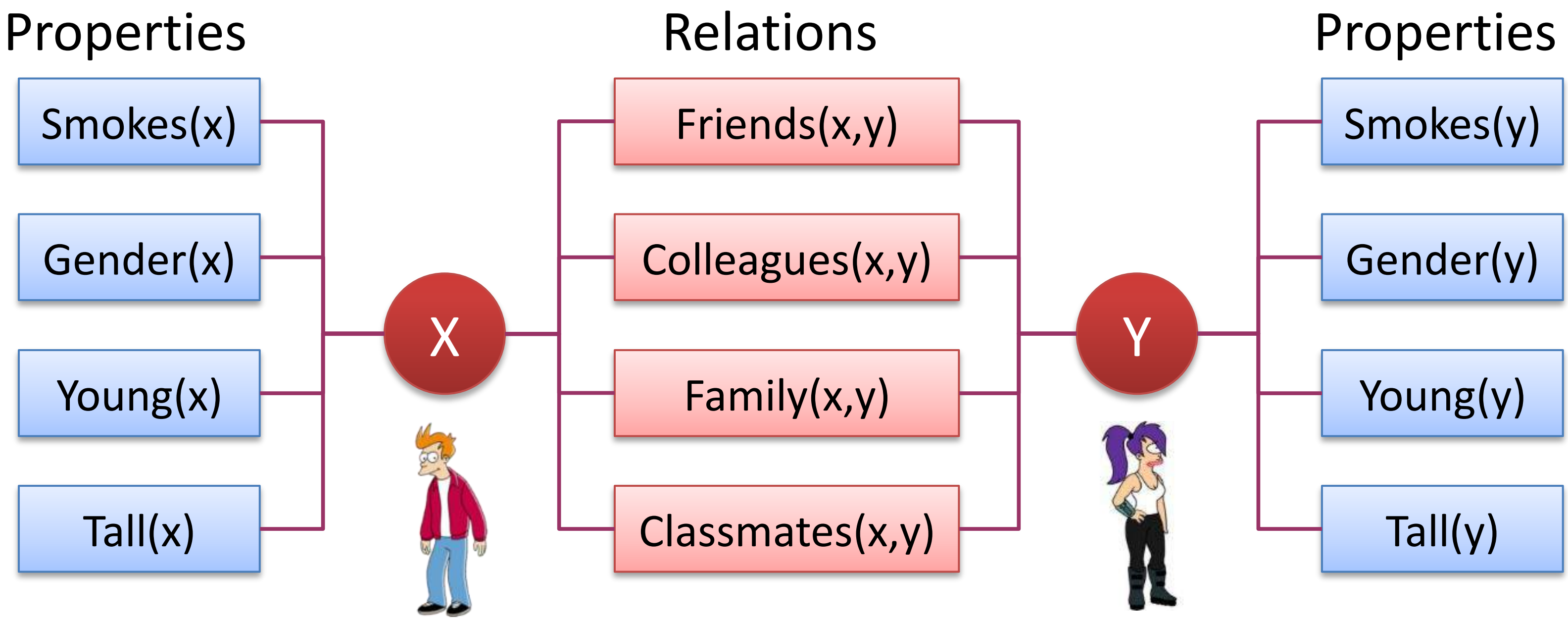
Computed in time polynomial in n

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

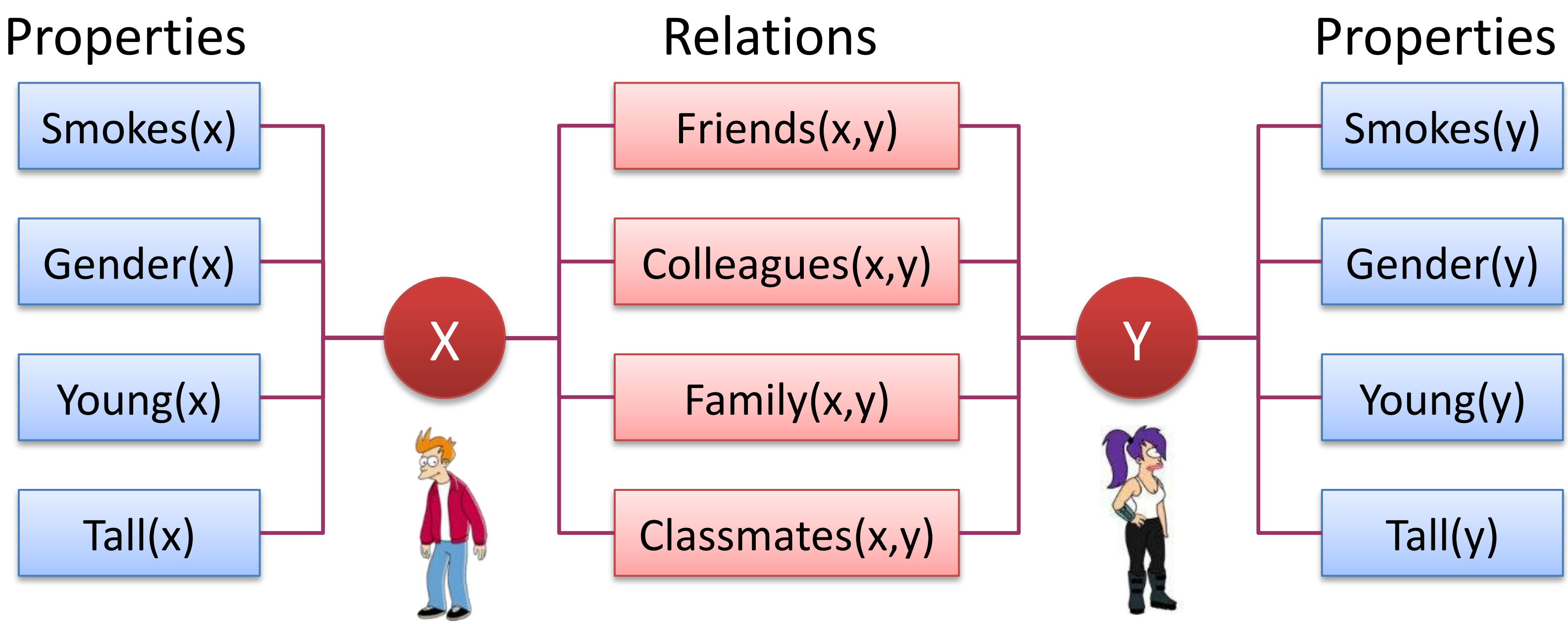
FO^2 is liftable!



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."

FO^2 is liftable!

Medical Records

Statistical Relational Model in FO^2

Name	Cough	Asthma	Smokes
Alice	1	1	0
Bob	0	0	0
Charlie	0	1	0
Dave	1	0	1
Eve	1	0	0

Frank	1	?	?

Friends

Brothers

2.1 Asthma(x) ⇒ Cough(x)
3.5 Smokes(x) ⇒ Cough(x)
1.9 Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)
1.5 Asthma (x) ∧ Family(x,y) ⇒ Asthma (y)

Frank	1	0.2	0.6

[Van den Broeck.; NIPS’11], [Van den Broeck et al.; KR’14]

FO^2 is liftable!

Medical Records → Statistical Relational Model in FO^2

Name	Cough	Asthma	Smokes
Alice	1	1	0
Bob	0	0	0
Charlie	0	1	0
Dave	1	0	1
Eve	1	0	0

Friends

Brothers

Frank	1	?	?

Big data

2.1 Asthma(x) ⇒ Cough(x)

3.5 Smokes(x) ⇒ Cough(x)

1.9 Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

1.5 Asthma (x) ∧ Family(x,y) ⇒ Asthma (y)

Frank	1	0.2	0.6

[Van den Broeck.; NIPS'11], [Van den Broeck et al.; KR'14]

Probabilistic Databases

- Tuple-independent probabilistic database

Scientist

Name	Prob
Erdos	0.9
Einstein	0.8
Straus	0.6

Coauthor

Actor	Director	Prob
Erdos	Straus	0.6
Einstein	Straus	0.7
Obama	Erdos	0.1

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

Probabilistic Databases

- Query: SQL or First-order logic

```
SELECT Actor.name
FROM Actor, WorkedFor
WHERE Actor.name = WorkedFor.actor
```

$Q(x) = \exists y\ Actor(x) \wedge WorkedFor(x,y)$

- Each UCQ query is either #P-hard, or PTIME in the size of the database.

[Dalvi and Suciu;JACM'11], [Ceylan, Darwiche, Van den Broeck; KR'16]

Probabilistic Databases

- Query: SQL or First-order logic

```
SELECT Actor.name
FROM Actor, WorkedFor
WHERE Actor.name = WorkedFor.actor
```

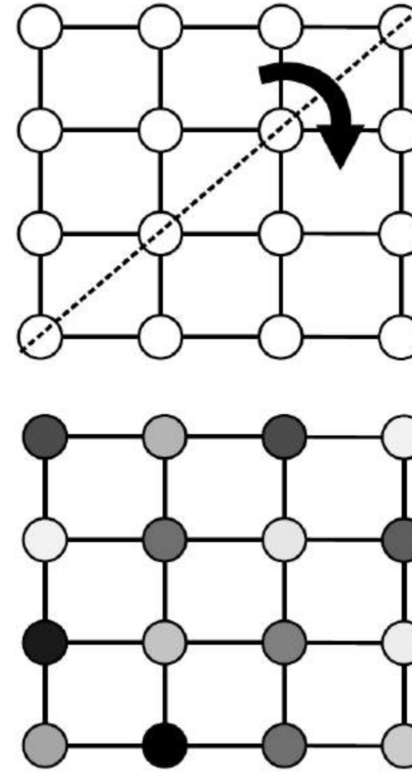
$Q(x) = \exists y \; Actor(x) \wedge WorkedFor(x,y)$

- Each UCQ query is either #P-hard, or PTIME in the size of the database.

Probabilistic query evaluation algorithm runs in **linear time** for all PTIME UCQ queries

[Dalvi and Suciu;JACM'11], [Ceylan, Darwiche, Van den Broeck; KR'16]

Approximate Symmetries



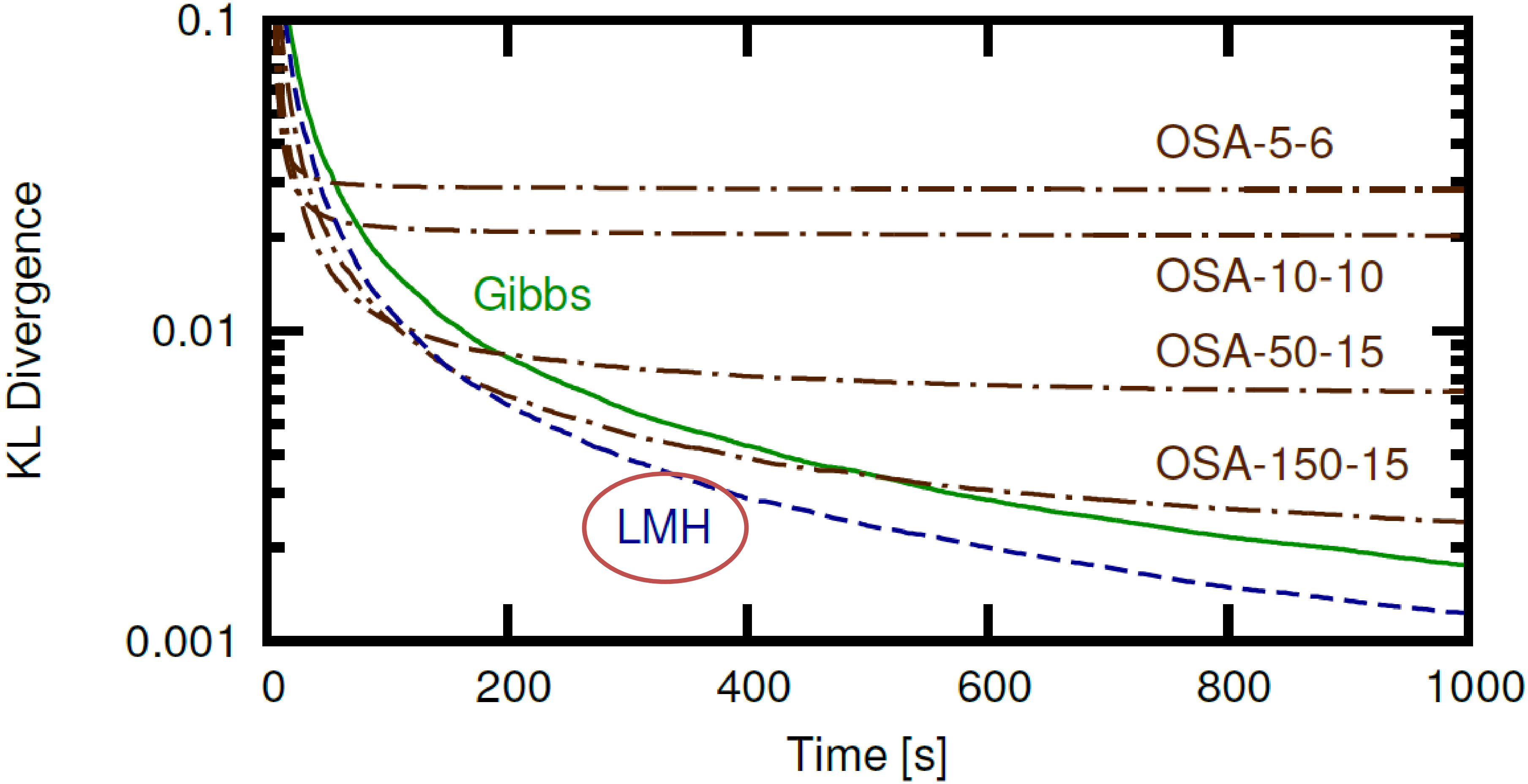
- Exploit approximate symmetries:
  - Exact symmetry g
    $\Pr(\mathbf{x}) = \Pr(\mathbf{x}^g)$
  - Approximate symmetry g
    $\Pr(\mathbf{x}) \approx \Pr(\mathbf{x}^g)$

Pr [] ≈ Pr []

- Approximate lifted inference (MCMC)

[Van den Broeck, Darwiche; NIPS'13], [Van den Broeck, Niepert; AAAI'15]

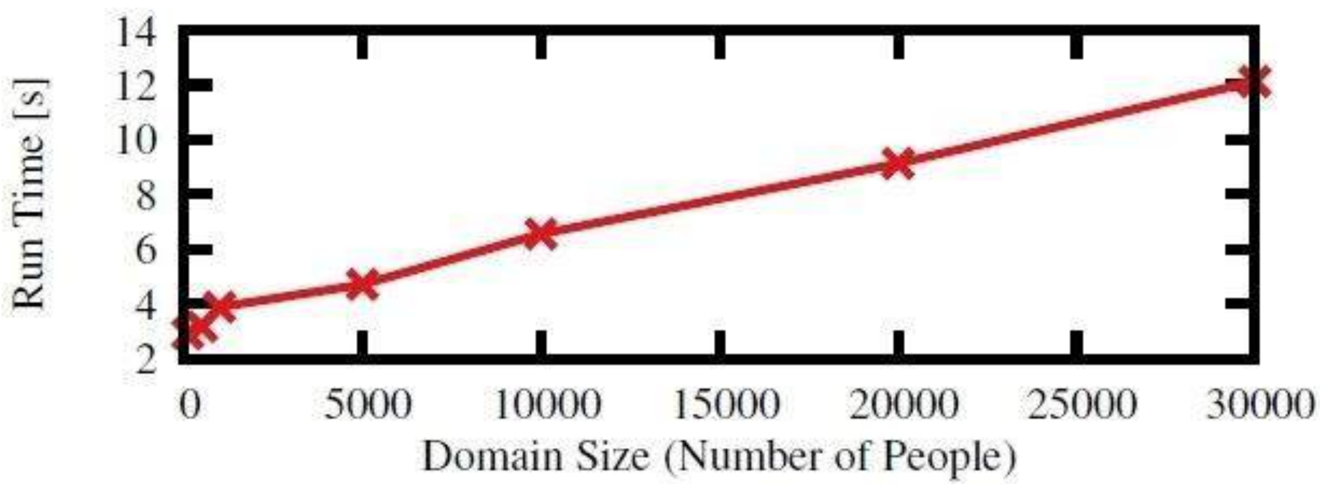
Experiments: WebKB



[Van den Broeck, Niepert; AAAI'15]

Lifted Parameter Learning

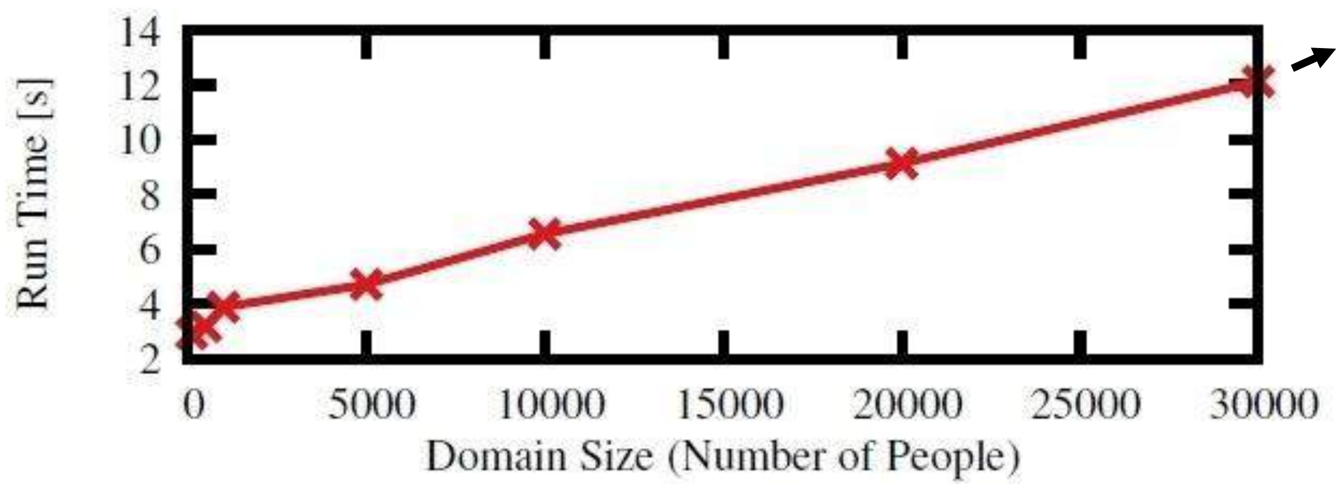
- **Given:** A set of first-order logic **formulas**
  A set of training **databases**
- **Learn:** Maximum-likelihood **weights**
- **Idea:** Lift the gradient computation



[Van Haaren et al.; MLJ'15]

Lifted Parameter Learning

- **Given:** A set of first-order logic **formulas**
  A set of training **databases**
- **Learn:** Maximum-likelihood **weights**
- **Idea:** Lift the gradient computation



900,030,000 random variables

[Van Haaren et al.; MLJ'15]

Lifted Structure Learning

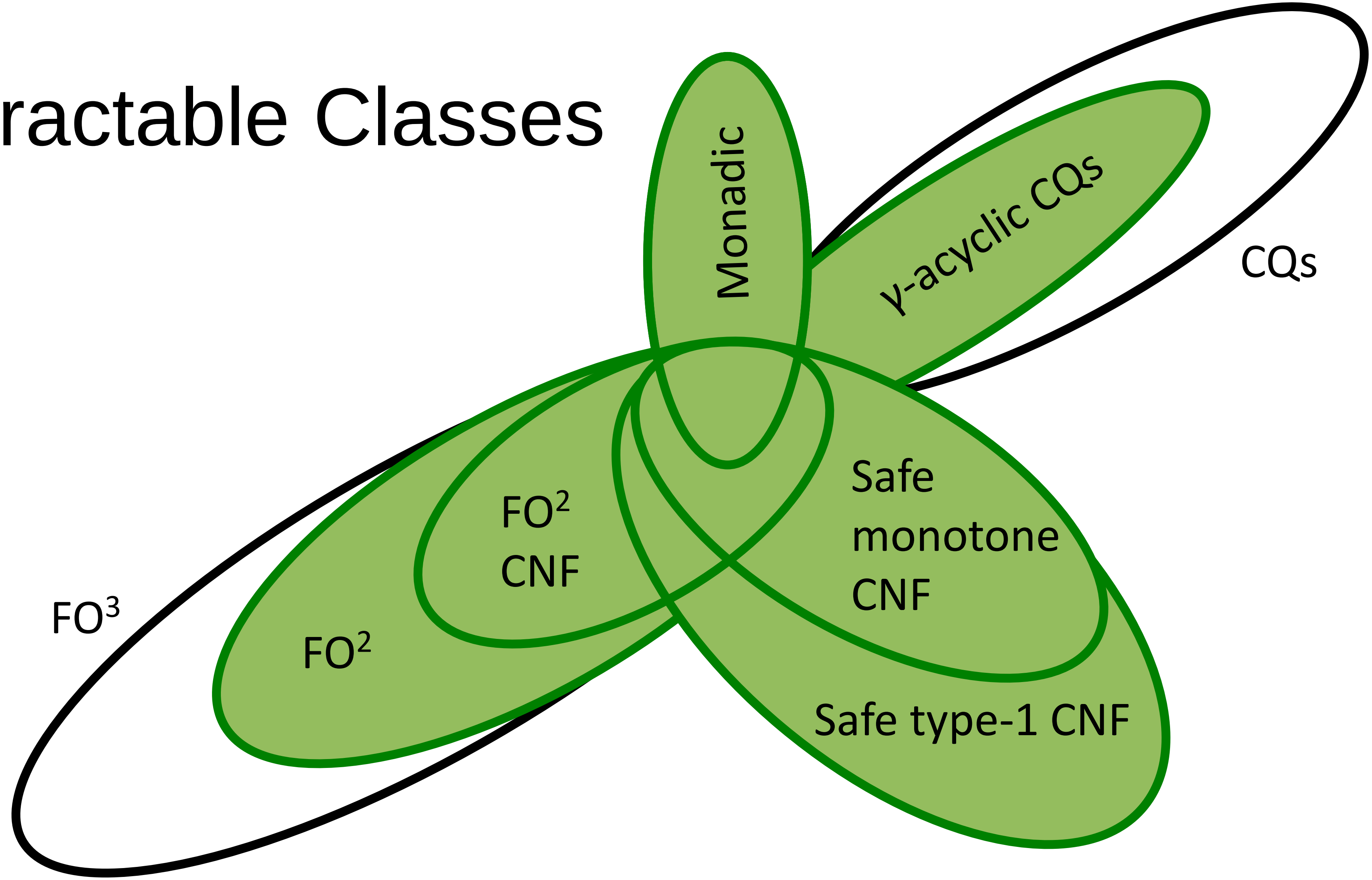
- **Given:** A set of training **databases**
- **Learn:** A set of first-order logic **formulas**
  The associated maximum-likelihood **weights**
- **Idea:** Learn liftable models (regularize with symmetry)

	IMDb			*UWCSE*		
	Baseline	Lifted Weight Learning	Lifted Structure Learning	Baseline	Lifted Weight Learning	Lifted Structure Learning
Fold 1	-548	-378	**-306**	-1,860	-1,524	**-1,477**
Fold 2	-689	-390	**-309**	-594	-535	**-511**
Fold 3	-1,157	-851	**-733**	-1,462	-1,245	**-1,167**
Fold 4	-415	-285	**-224**	-2,820	-2,510	**-2,442**
Fold 5	-413	-267	**-216**	-2,763	-2,357	**-2,227**

[Van Haaren et al.; MLJ'15]

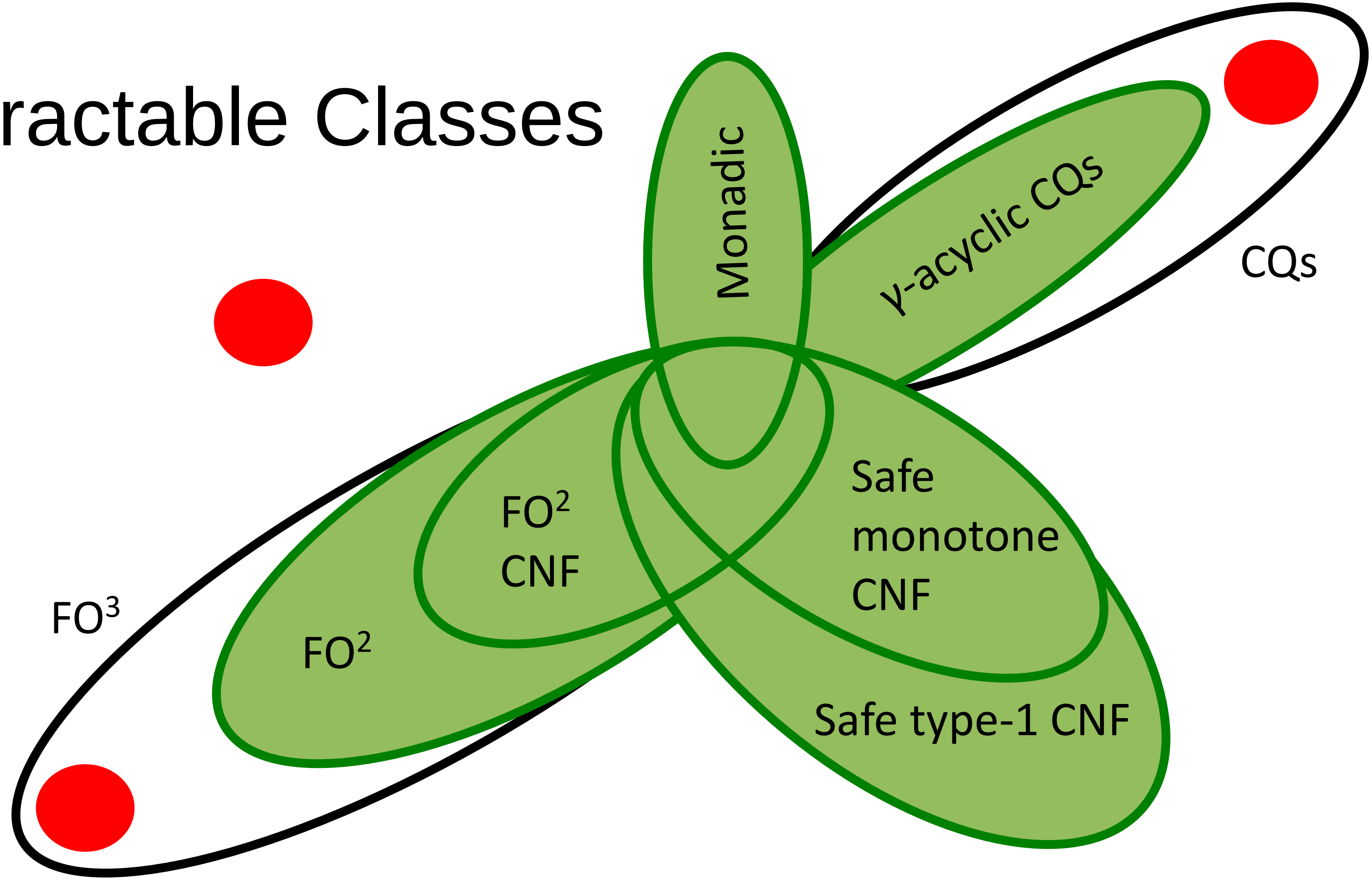
*What are the **challenges**?*

Tractable Classes



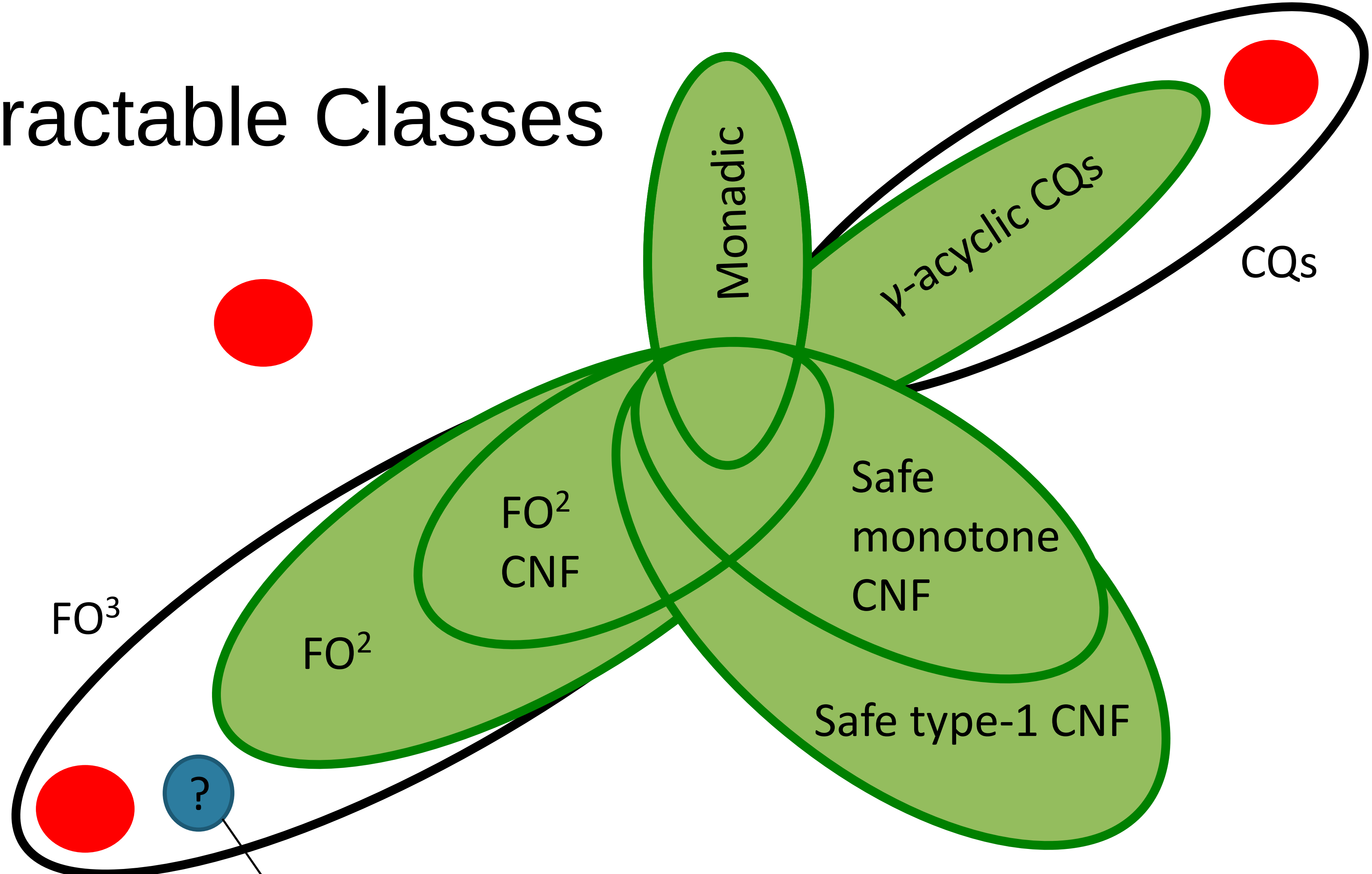
[VdB; NIPS'11], [VdB et al.; KR'14], [Gribkoff, VdB, Suciu; UAI'15], [Beame, VdB, Gribkoff, Suciu; PODS'15], etc.

Tractable Classes



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

Tractable Classes



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

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

Generalized Model Counting

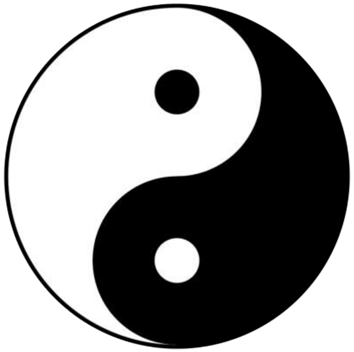
Probability Distribution

=

Logic

+

Weights



Generalized Model Counting

Probability Distribution	Logical Syntax
=	
Logic	Model-theoretic Semantics
+	+
Weights	Weight function w(.)

Weighted Model Integration

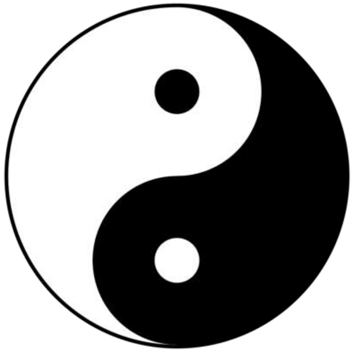
Probability Distribution

=

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

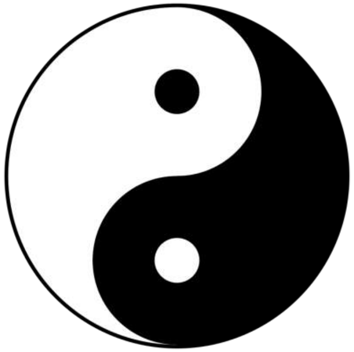
Probability Distribution

=

Logic Programs

+

Weights



[Fierens et al., TPLP'15]

Probabilistic Programming

Probability Distribution

=

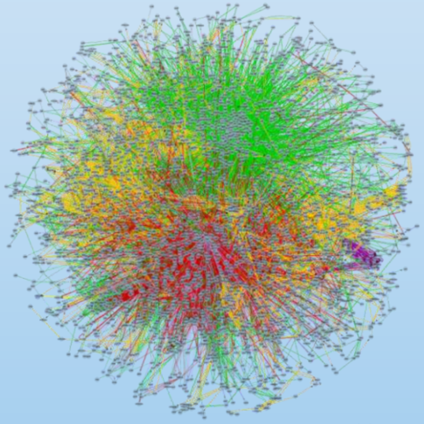
Logic Programs

+

Weights

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

+



[Fierens et al., TPLP'15]

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

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

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**)

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

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**)

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

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**)

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

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**)

[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**)

[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
…	…	…

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

$$Q3 = \text{Coauthor}(\text{Einstein},\textbf{Straus}) \wedge \text{Coauthor}(\text{Erdos},\textbf{Straus})$$

$$Q4 = \text{Coauthor}(\text{Einstein},\textbf{Bieber}) \wedge \text{Coauthor}(\text{Erdos},\textbf{Bieber})$$

$$Q5 = \text{Coauthor}(\text{Einstein},\textbf{Bieber}) \wedge \neg\text{Coauthor}(\textbf{Einstein},\textbf{Bieber})$$

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

and $P(Q3) \geq P(Q5)$, $P(Q4) \geq P(Q5)$

[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) 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]

Conclusions

- Integration of logic and probability is long-standing goal of AI
- First-order probabilistic reasoning is **frontier** and **integration** of AI, KR, ML, DBs, theory, PL, etc.
- We need
  - relational models and logic
  - probabilistic models and statistical learning
  - algorithms that scale

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**

QUESTIONS?

References

- Van den Broeck, Guy. "Towards high-level probabilistic reasoning with lifted inference." AAAI Spring Symposium on KRR (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.
- Chavira, Mark, Adnan Darwiche, and Manfred Jaeger. "Compiling relational Bayesian networks for exact inference." International Journal of Approximate Reasoning 42.1 (2006): 4-20.
- 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.

References

- Van den Broeck, Guy, Nima Taghipour, Wannes Meert, Jesse Davis, and Luc De Raedt. "Lifted probabilistic inference by first-order knowledge compilation." AAAI, 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).
- Gribkoff, Eric, Guy Van den Broeck, and Dan Suciu. "Understanding the complexity of lifted inference and asymmetric weighted model counting." UAI (2014).
- Niepert, Mathias, and Guy Van den Broeck. "Tractability through exchangeability: A new perspective on efficient probabilistic inference." AAAI (2014).

References

- Van den Broeck, Guy. "On the completeness of first-order knowledge compilation for lifted probabilistic inference." Advances in Neural Information Processing Systems. 2011.
- Van den Broeck, Guy, Wannes Meert, and Adnan Darwiche. "Skolemization for weighted first-order model counting." Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR). 2014.
- Ceylan, Ismail Ilkan, Adnan Darwiche, and Guy Van den Broeck. "Open-world probabilistic databases." Proceedings of KR (2016).
- Van den Broeck, Guy, and Adnan Darwiche. "On the complexity and approximation of binary evidence in lifted inference." Advances in Neural Information Processing Systems. 2013.
- Van den Broeck, Guy, and Mathias Niepert. "Lifted probabilistic inference for asymmetric graphical models." Proceedings of AAAI (2015).

References

- Van Haaren, Jan, Guy Van den Broeck, Wannes Meert, and Jesse Davis. "Lifted generative learning of Markov logic networks." Machine Learning 103, no. 1 (2016): 27-55.
- 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.
- 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.