Lifted Inference in Statistical Relational Models

Guy Van den Broeck

BUDA Invited Tutorial
June 22nd 2014

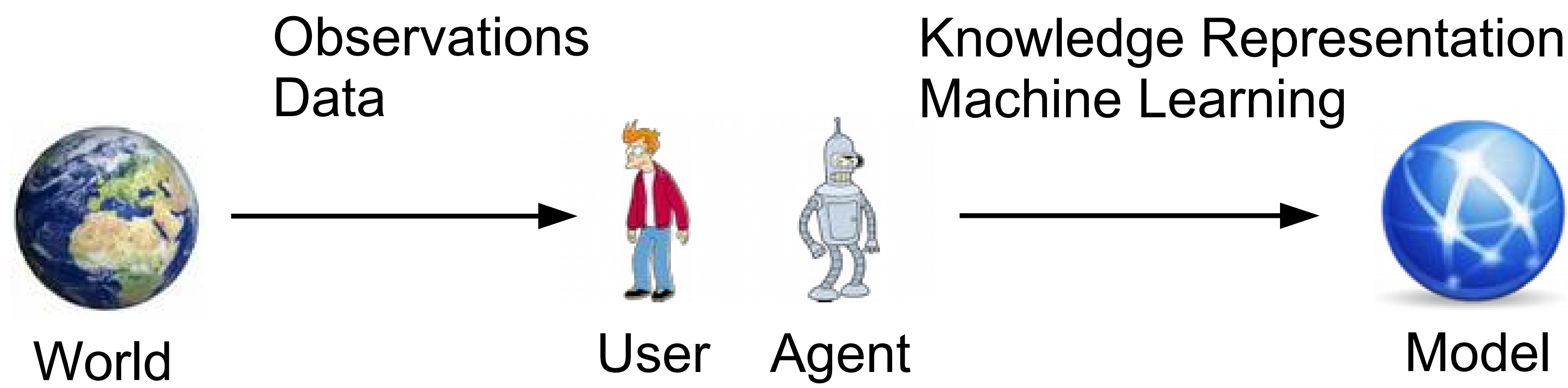
Overview

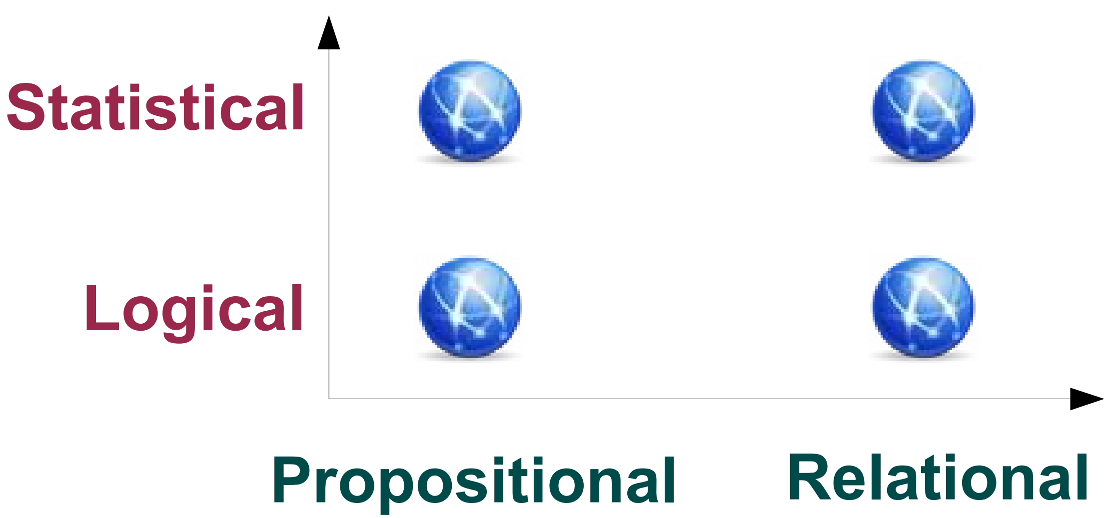
1. What are statistical relational models?
2. What is lifted inference?
3. How does lifted inference work?
4. Theoretical insights
5. Practical applications

Overview

1. **What are statistical relational models?**
2. What is lifted inference?
3. How does lifted inference work?
4. Theoretical insights
5. Practical applications

Types of Models

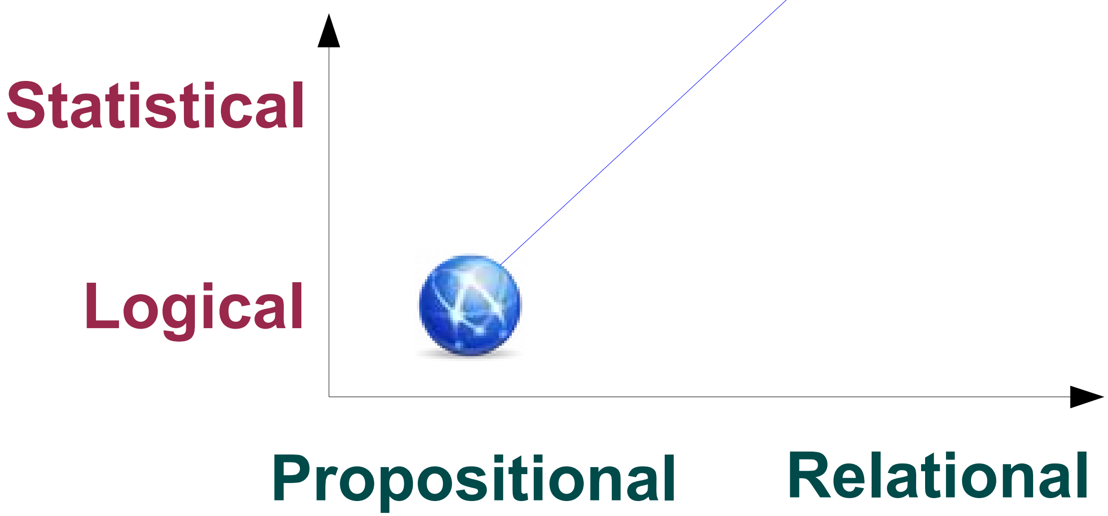




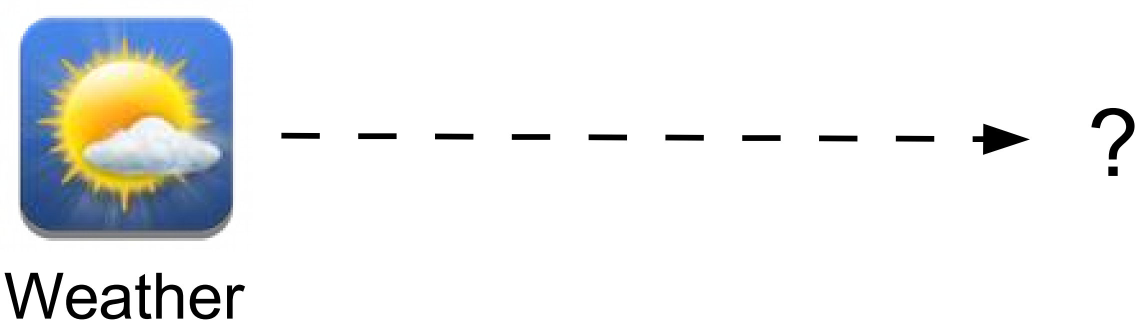
Logical Propositional Models

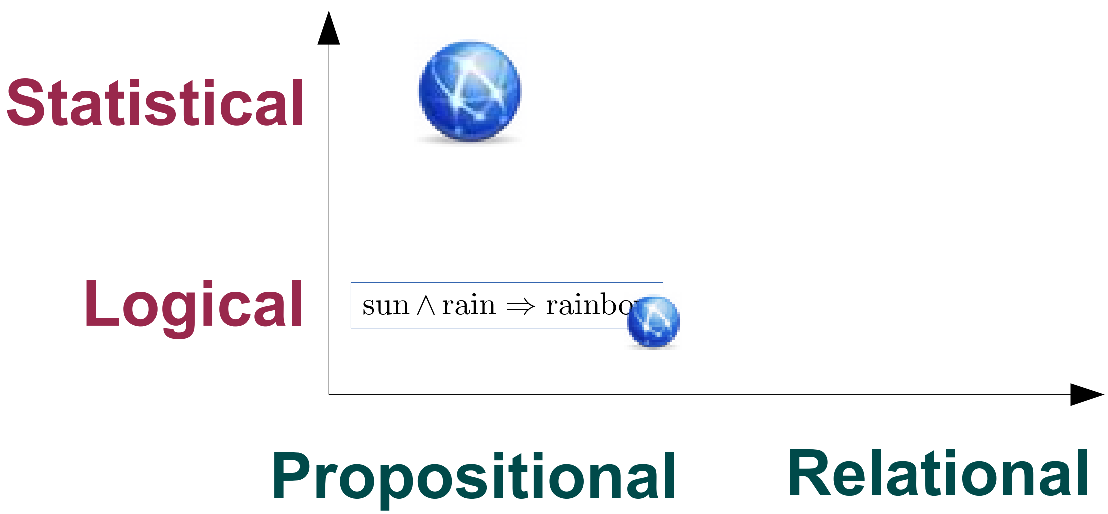
Weather

$$\boxed{\text{sun} \wedge \text{rain} \Rightarrow \text{rainbow}}$$



Statistical Propositional Models





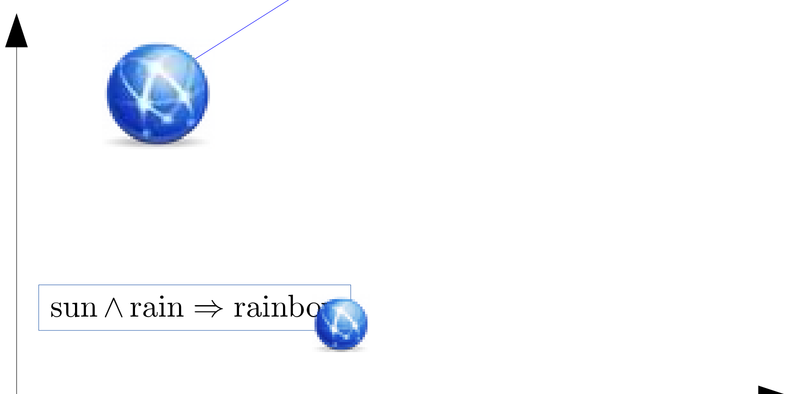
Statistical Propositional Models

Weather

rain	sun	rainbow	Prob.
T	T	T	0.018
T	T	F	0.002
T	F	T	0.009
T	F	F	0.171
F	T	T	0.024
F	T	F	0.456
F	F	T	0
F	F	F	0.32

Statistical

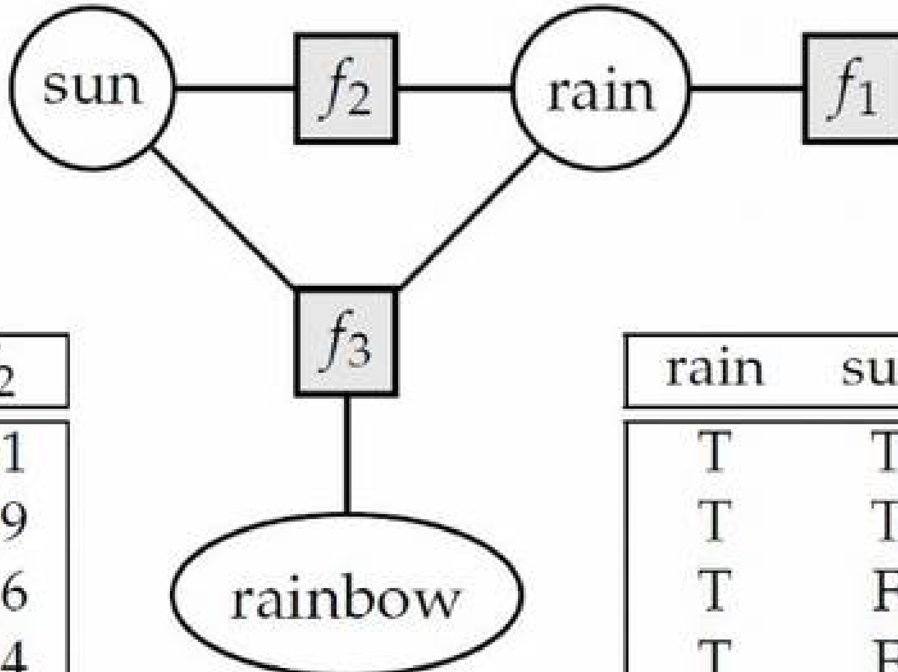
Logical



Propositional

Relational

Probabilistic Graphical Models: Factor Graphs



rain	f_1
T	0.2
F	0.8

rain	sun	f_2
T	T	0.1
T	F	0.9
F	T	0.6
F	F	0.4

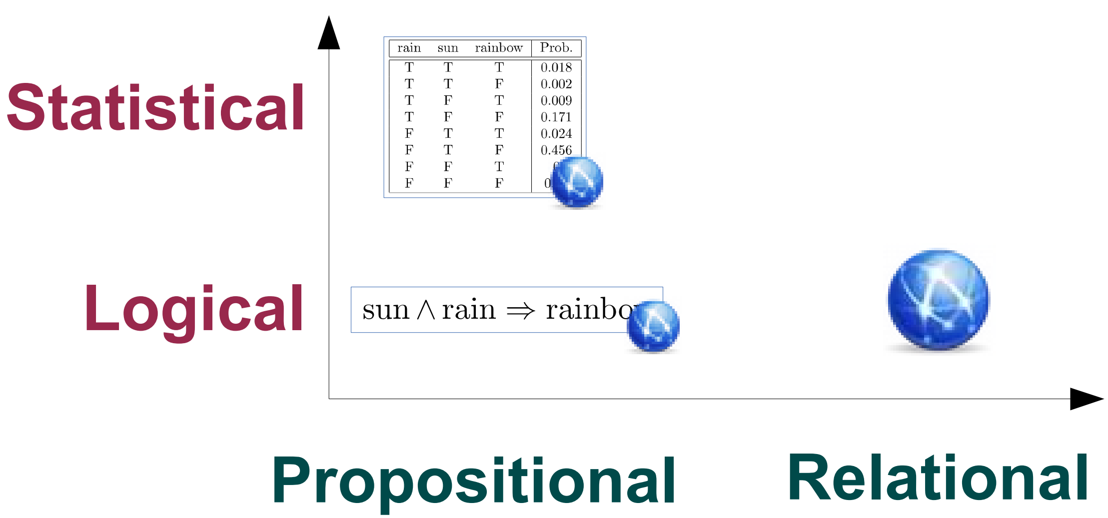
rain	sun	rainbow	f_3
T	T	T	0.9
T	T	F	0.1
T	F	T	0.05
T	F	F	0.95
F	T	T	0.05
F	T	F	0.95
F	F	T	0
F	F	F	1

$$\Pr(\omega) = \frac{1}{Z} \prod_i f_i(\omega_i)$$ where $$Z = \sum_\omega \prod_i f_i(\omega_i)$$

Logical Relational Models

Social
Network

?



Logical Relational Models

- Example: First-Order Logic

Formula

$$\forall x,y,\ \text{Smokes}(x) \land \text{Friends}(x,y) \Rightarrow \text{Smokes}(y)$$

Atom

Logical Variables

- Logical variables have domain of constants

  e.g., x,y range over domain People = {Alice,Bob}

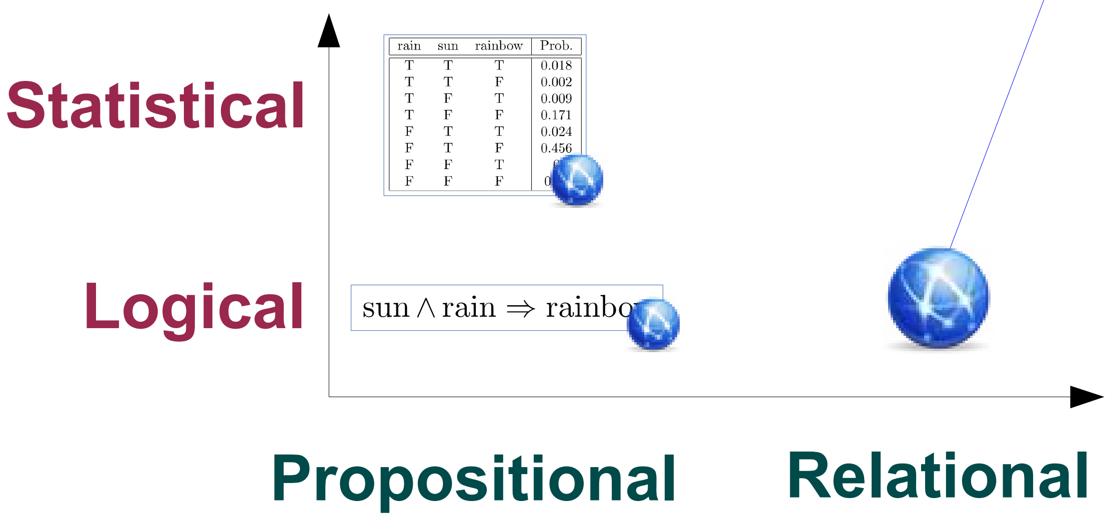
- Ground formula has no logical variables

  e.g., Smokes(Alice) ∧ Friends(Alice,Bob) ⇒ Smokes(Bob)

Logical Relational Models

Social Network

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

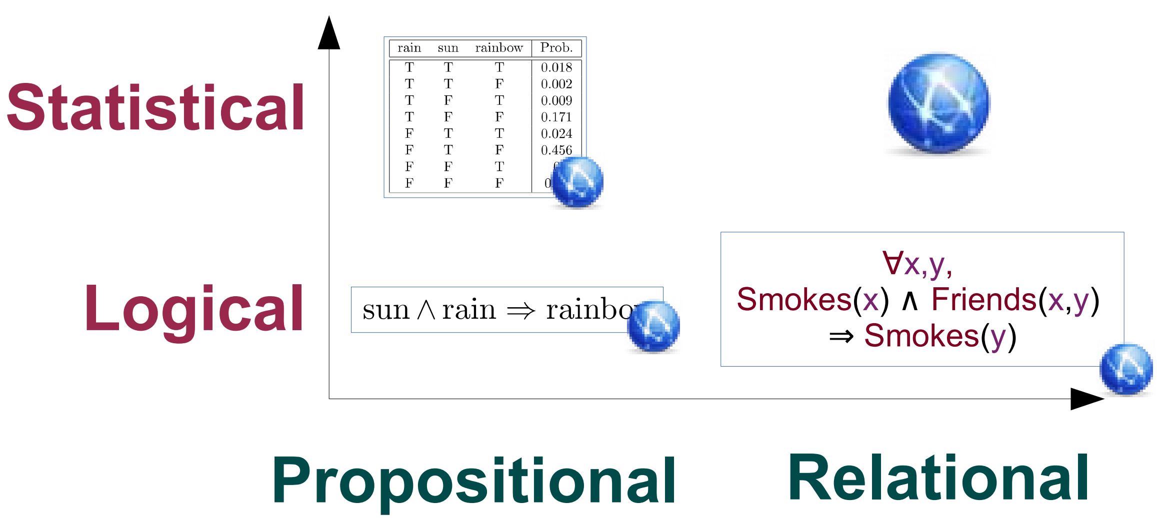


rain	sun	rainbow	Prob.
T	T	T	0.018
T	T	F	0.002
T	F	T	0.009
T	F	F	0.171
F	T	T	0.024
F	T	F	0.456
F	F	T	[illegible]
F	F	F	[illegible]

Statistical Relational Models

?

Social Network



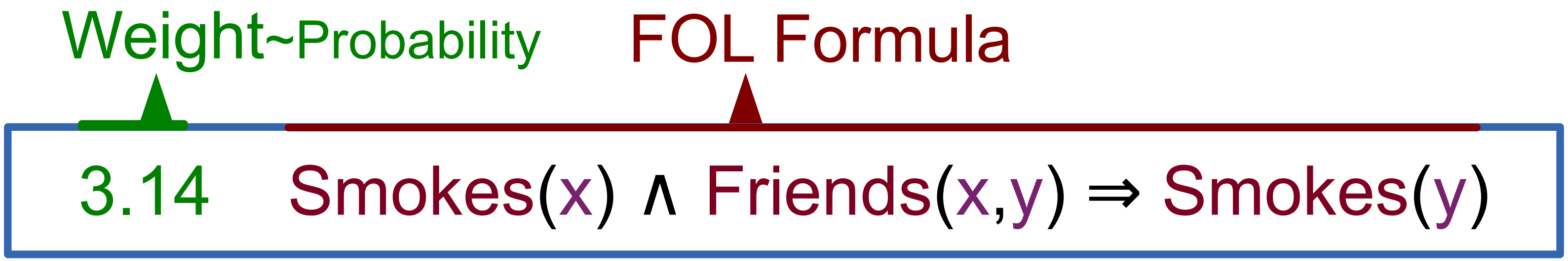
Why Statistical Relational Models?

- Probabilistic graphical models
  - ✖ Not very expressive
    Rules of chess in ~100,000 pages
  - ✔ Quantify uncertainty and noise
- Relational representations
  - ✔ Very expressive
    Rules of chess in 1 page
  - ✔ Relational data is everywhere
  - ✖ Hard to express uncertainty

→ **Need probability distribution over databases**

Markov Logic Networks (MLNs)

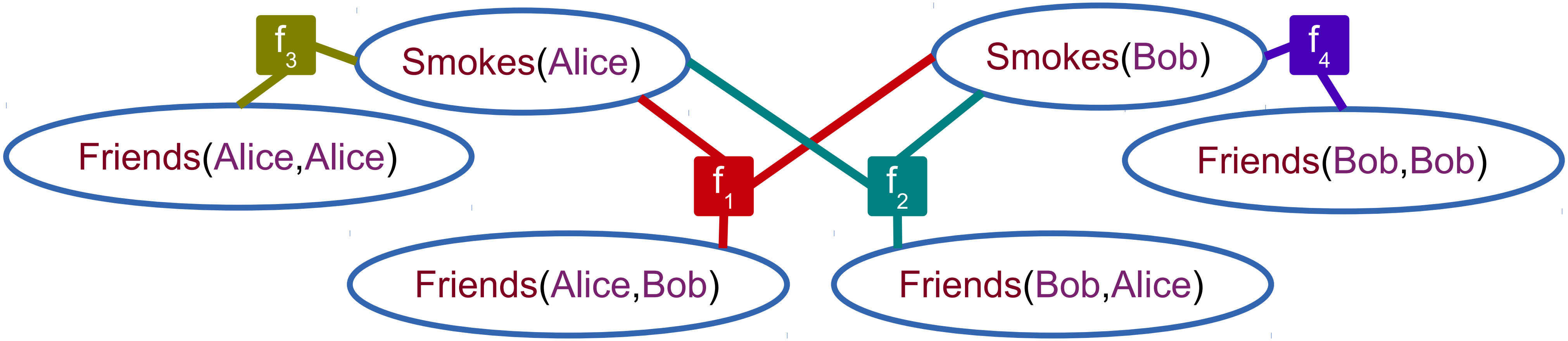
- Weighted First-Order Logic



- Ground atom/tuple = random variable in {true,false}

  e.g., Smokes(Alice), Friends(Alice,Bob), etc.

- Ground formula = factor in propositional factor graph



[Richardson-MLJ06]

Statistical Relational Models

Social Network

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

Statistical

rain	sun	rainbow	Prob.
T	T	T	0.018
T	T	F	0.002
T	F	T	0.009
T	F	F	0.171
F	T	T	0.024
F	T	F	0.456
F	F	T	[illegible]
F	F	F	[illegible]

Logical

$\text{sun} \wedge \text{rain} \Rightarrow \text{rainbo}$

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

Propositional **Relational**

Reasoning about Statistical Models: Probabilistic Inference

- Model:

0.7	Actor(a) ⇒ ¬Director(a)
1.2	Director(a) ⇒ ¬WorkedFor(a,b)
1.4	InMovie(m,a) ∧ WorkedFor(a,b) ⇒ InMovie(m,b)

- Inference query:
  - Given database tables for Actor, Director, WorkedFor

    Actor(Brando), Actor(Cruise), Director(Coppola),
    WorkedFor(Brando, Coppola), etc.

  - What is the probability of each tuple in table InMovie?

    Pr(InMovie(GodFather, Brando)) = ?

  - What is the most likely table for InMovie?

What about Probabilistic Databases?

- Tuple-independent probabilistic databases

Prob	Actor
0.9	Brando
0.8	Cruise
0.1	Coppola

Prob	WorkedFor	
0.9	Brando	Coppola
0.2	Coppola	Brando
0.1	Cruise	Coppola

...

- Also a distribution over deterministic databases
- Different purpose (query seen data vs. generalize to unseen data)
- Underlying reasoning task identical:

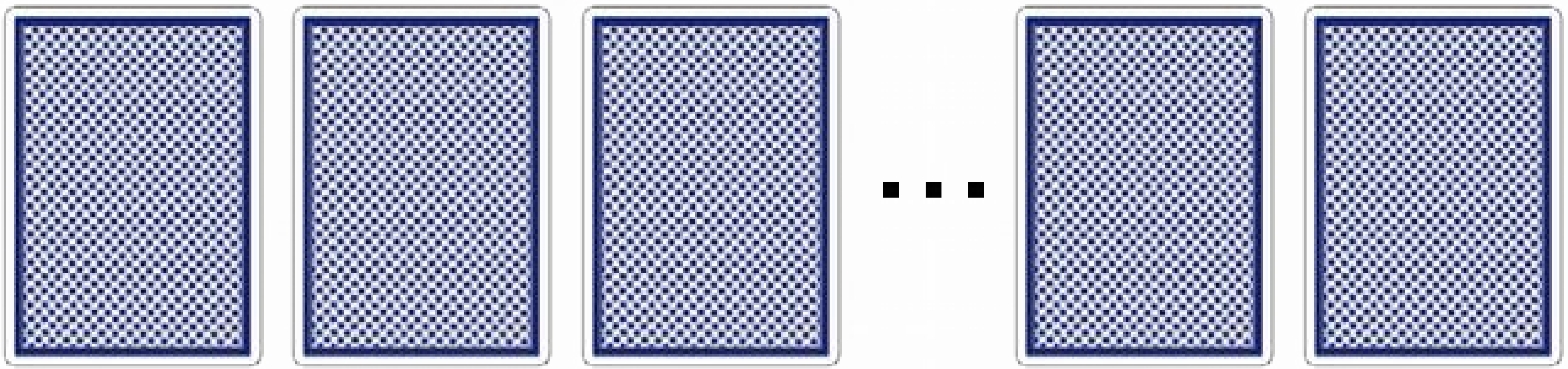
  Weighted (First-Order) Model Counting

[Suciu-Book11, Jha-TCS13, Olteanu-SUM08, VdB-IJCAI11, Gogate-UAI11, Gribkoff-UAI14]

Overview

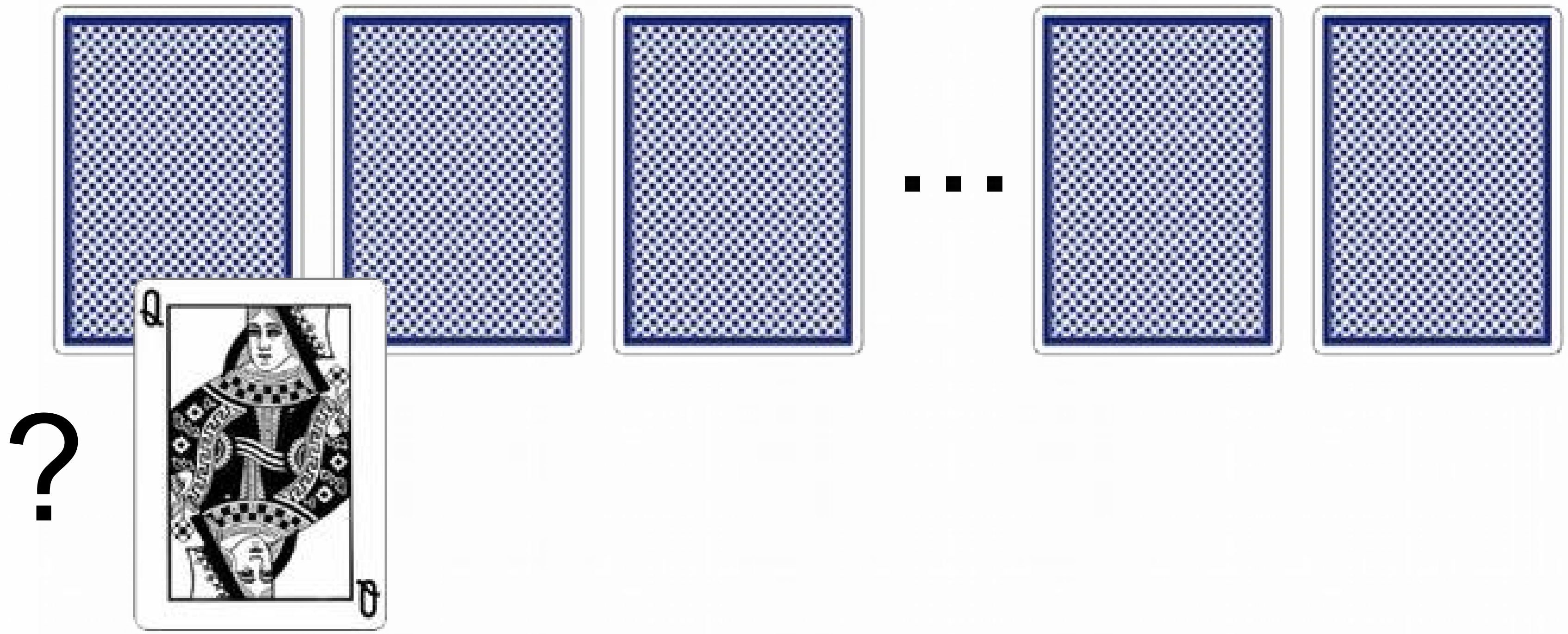
1. What are statistical relational models?
2. **What is lifted inference?**
3. How does lifted inference work?
4. Theoretical insights
5. Practical applications

A Simple Reasoning Problem

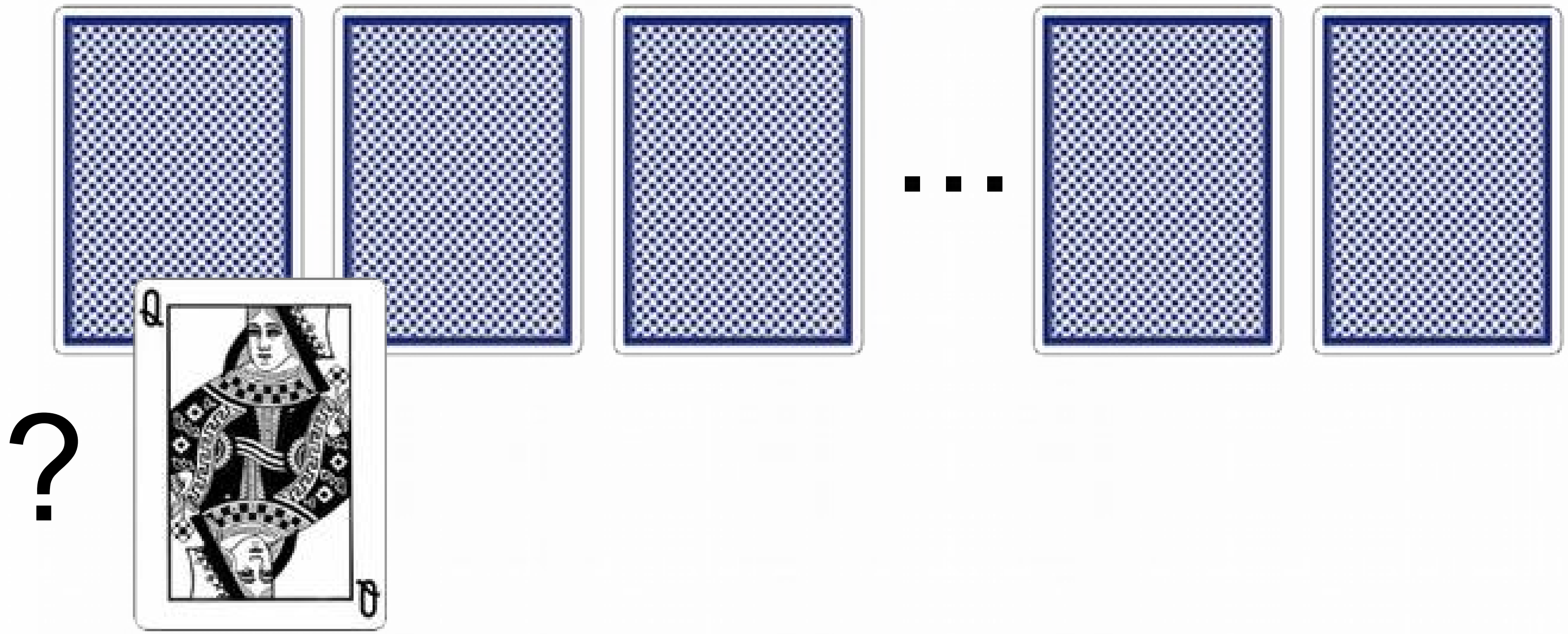


- 52 playing cards
- Let us ask some simple questions

A Simple Reasoning Problem

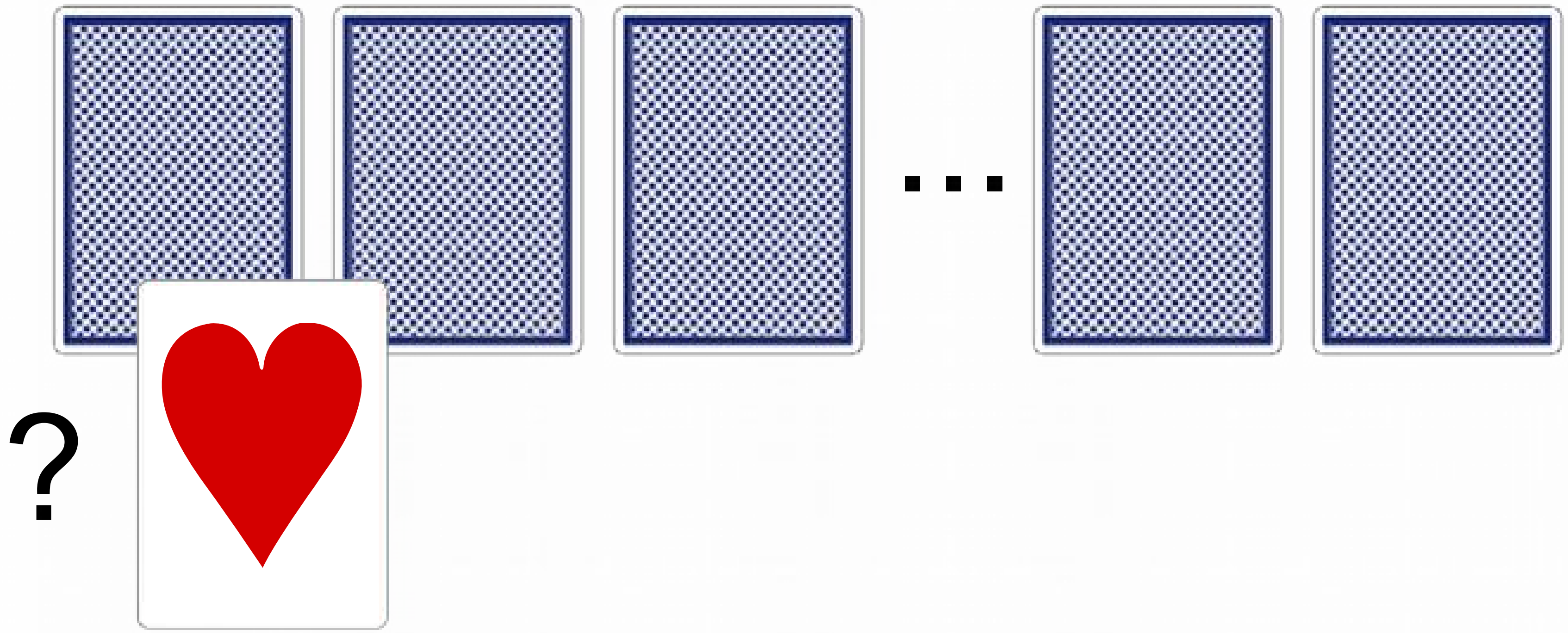


A Simple Reasoning Problem

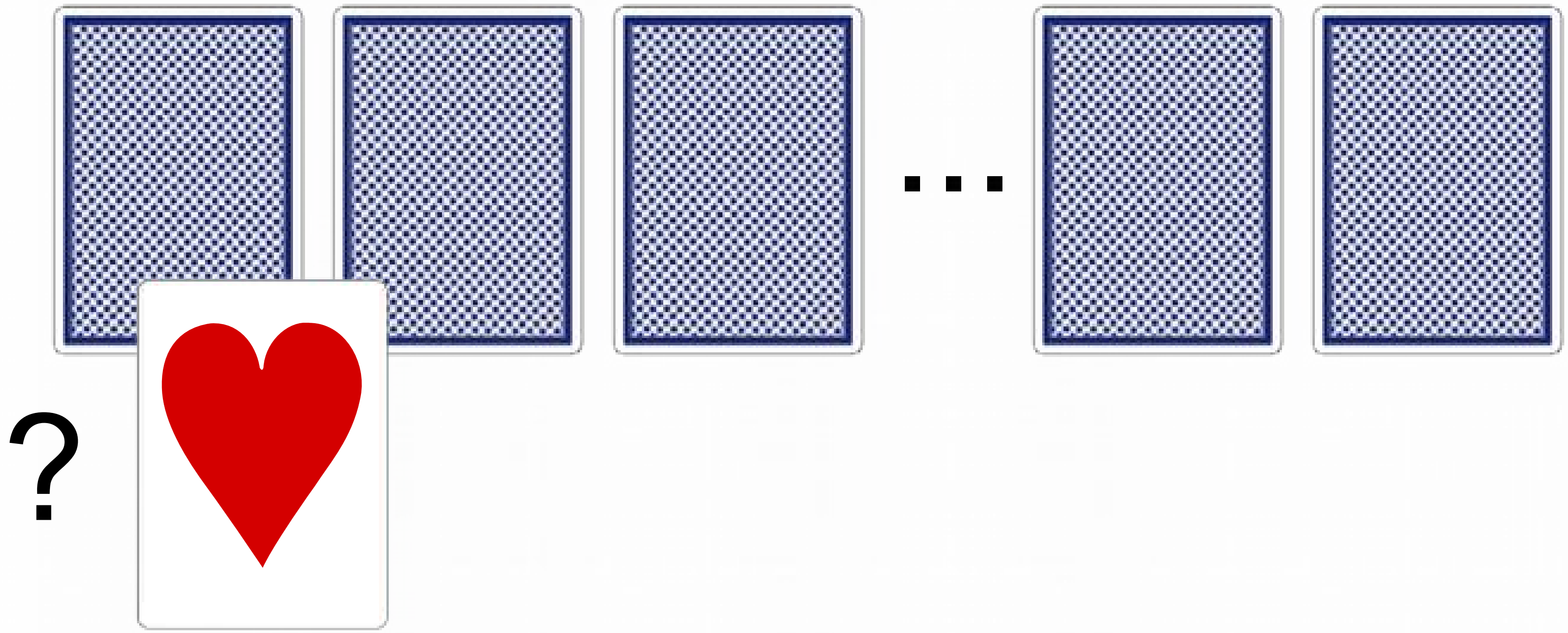


Probability 1/13

A Simple Reasoning Problem

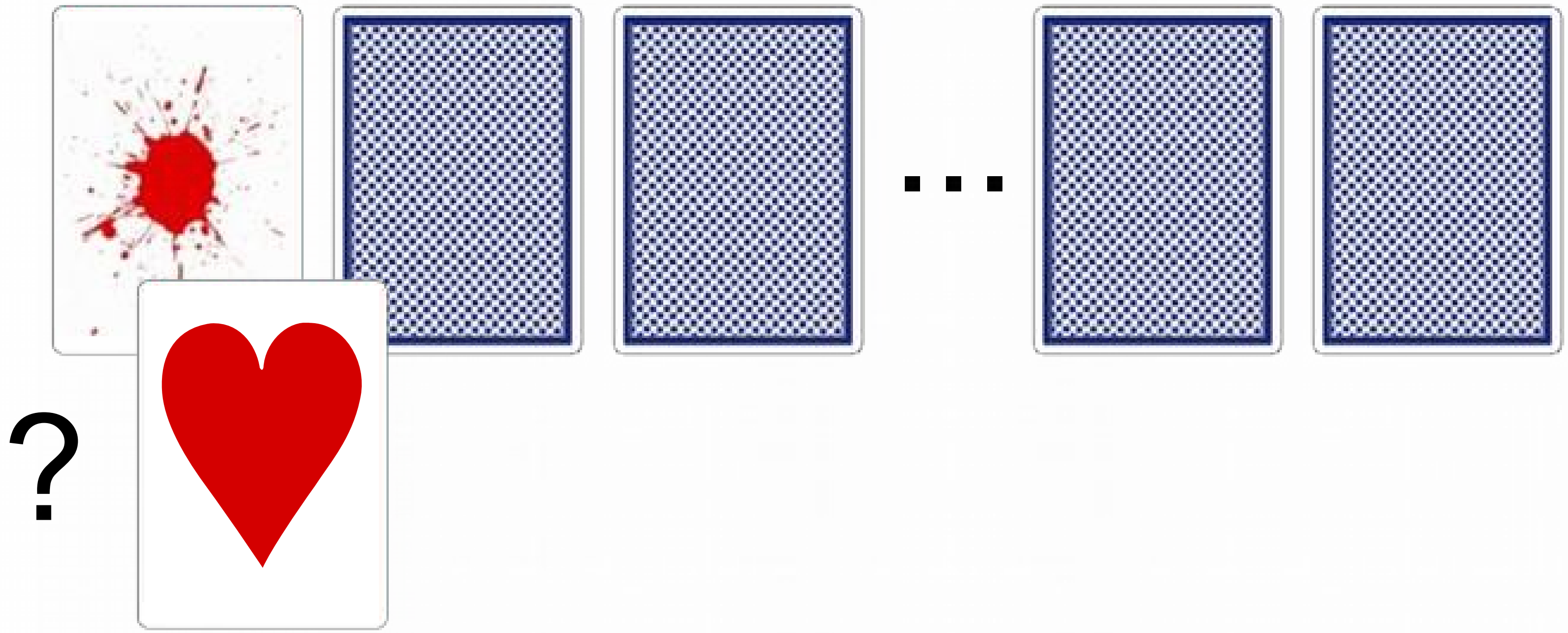


A Simple Reasoning Problem

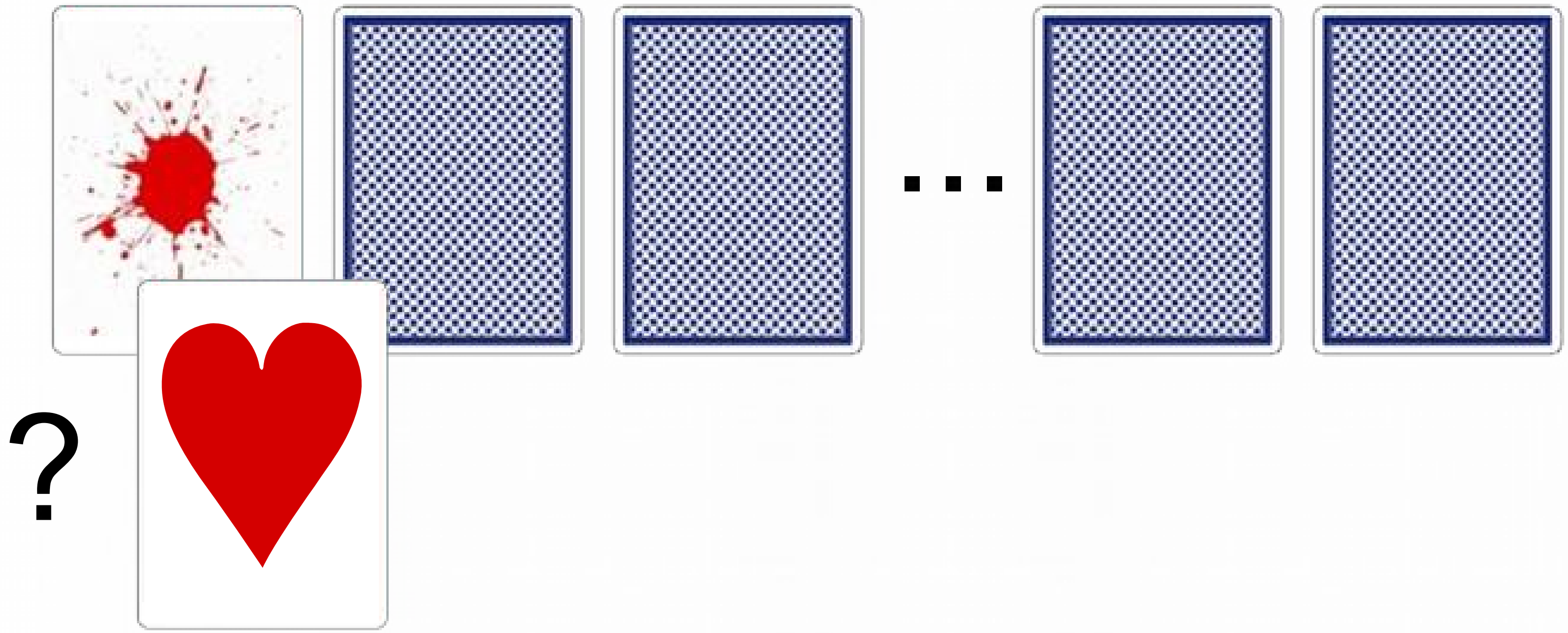


Probability 1/4

A Simple Reasoning Problem

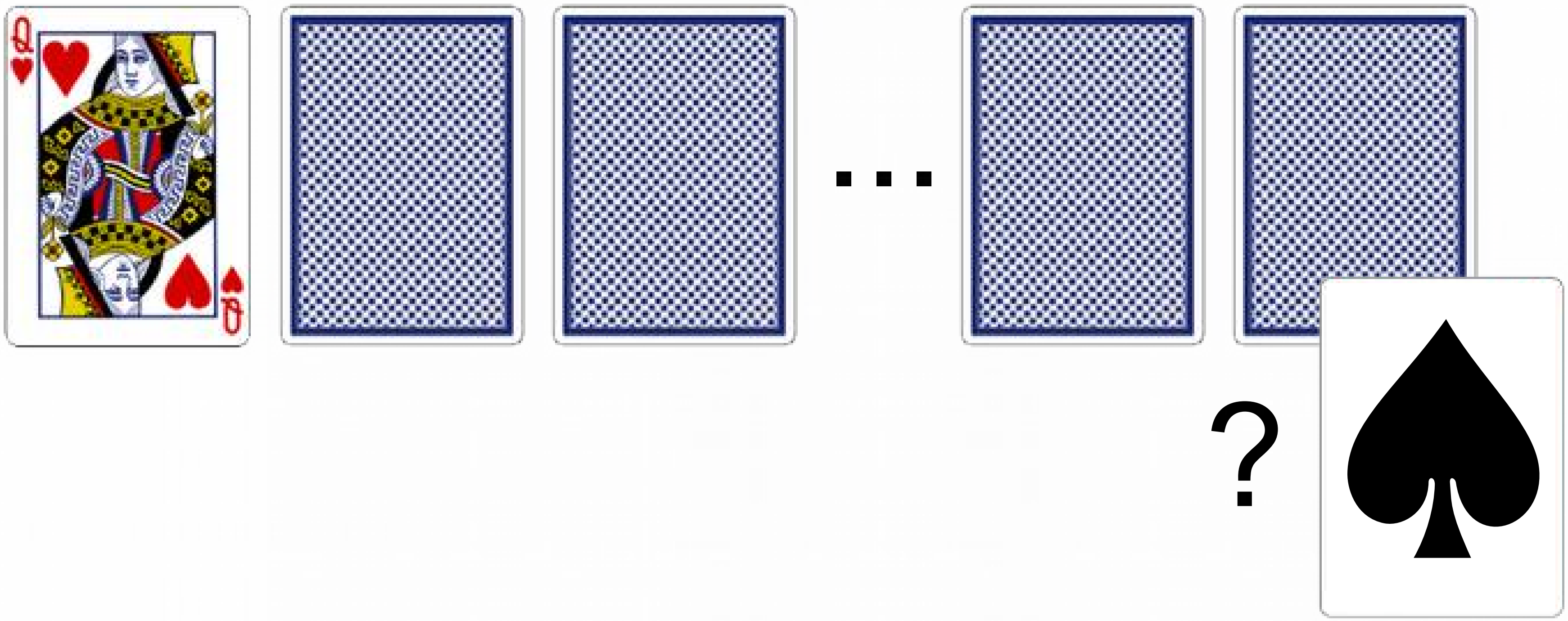


A Simple Reasoning Problem

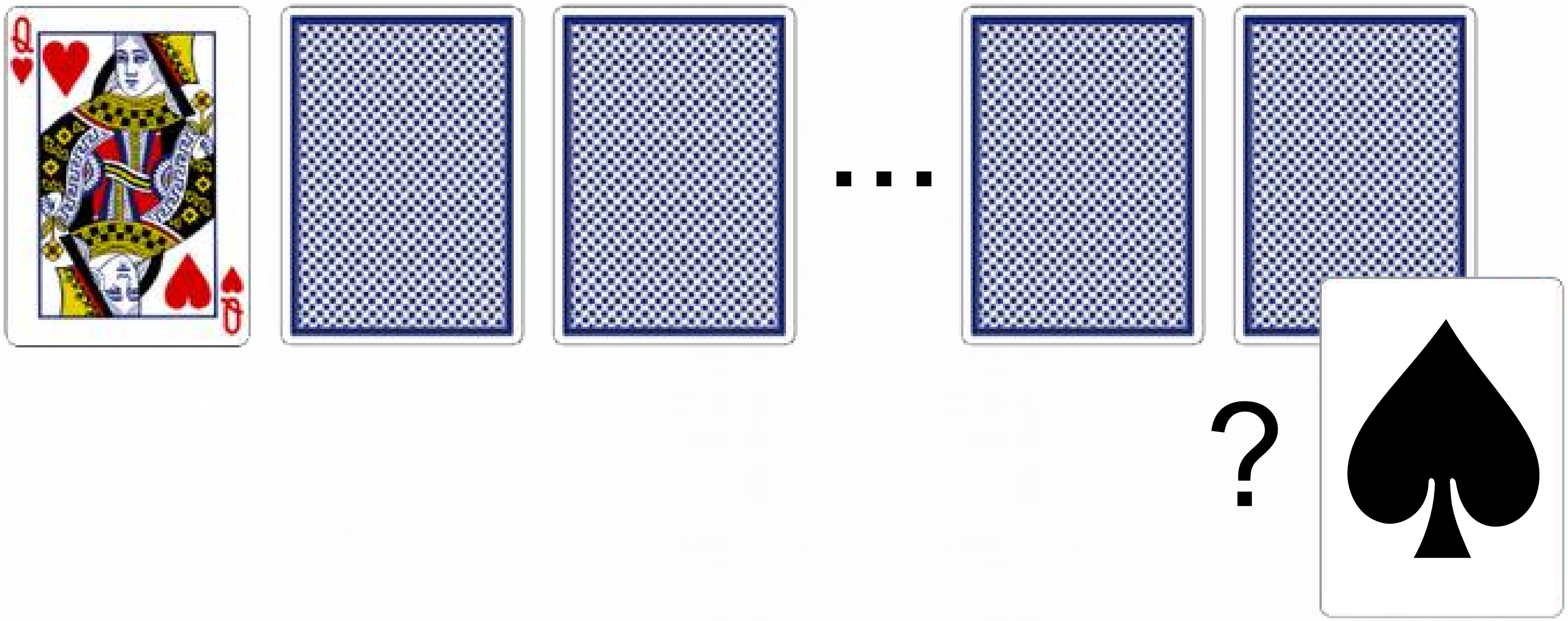


Probability 1/2

A Simple Reasoning Problem



A Simple Reasoning Problem

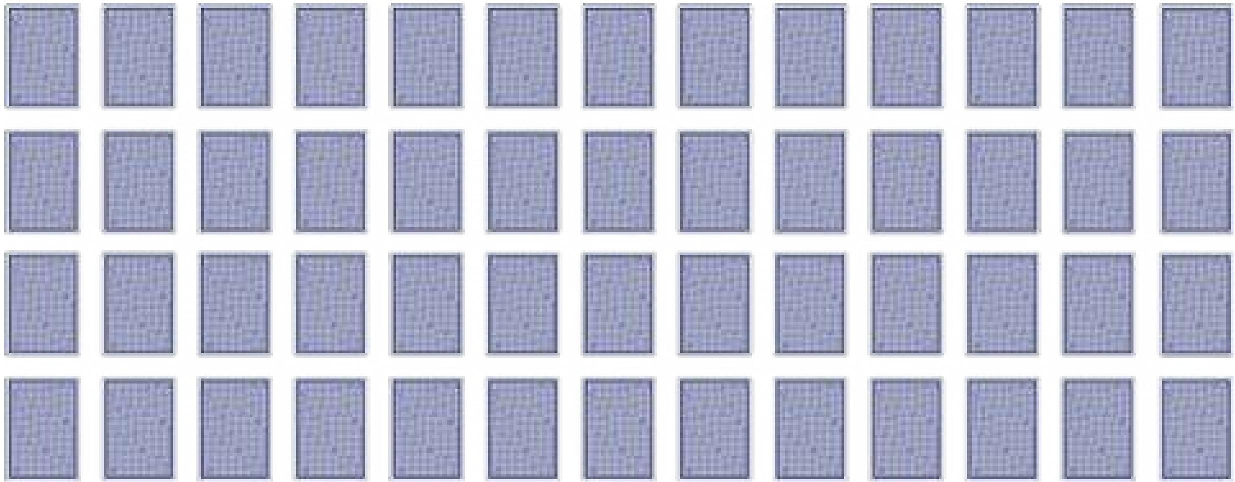


Probability 13/51

Automated Reasoning

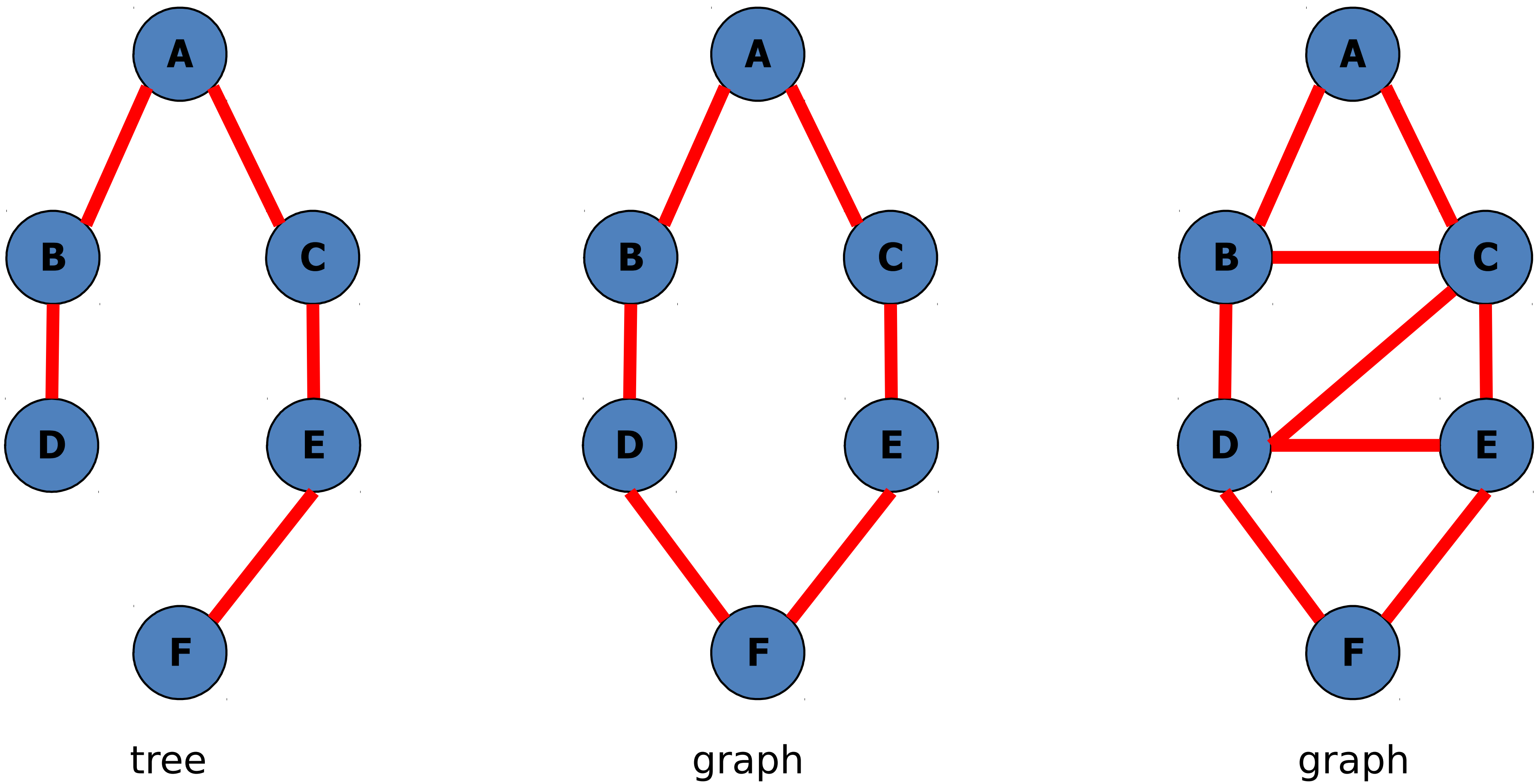
Let us automate this:

1. Probabilistic propositional model (factor graph)



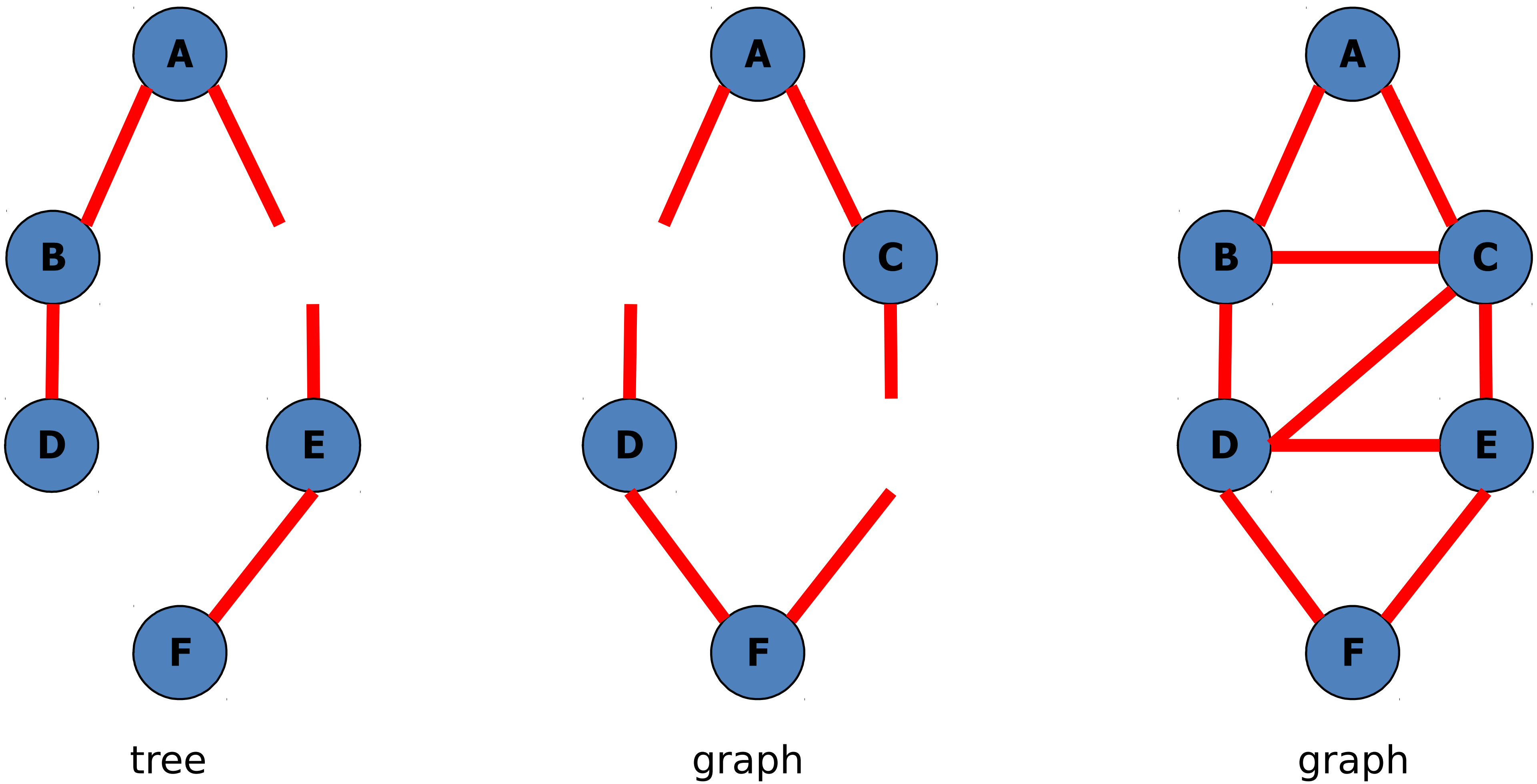
2. Probabilistic inference algorithm

Reasoning in Propositional Models



A key result: Treewidth
Why?

Reasoning in Propositional Models

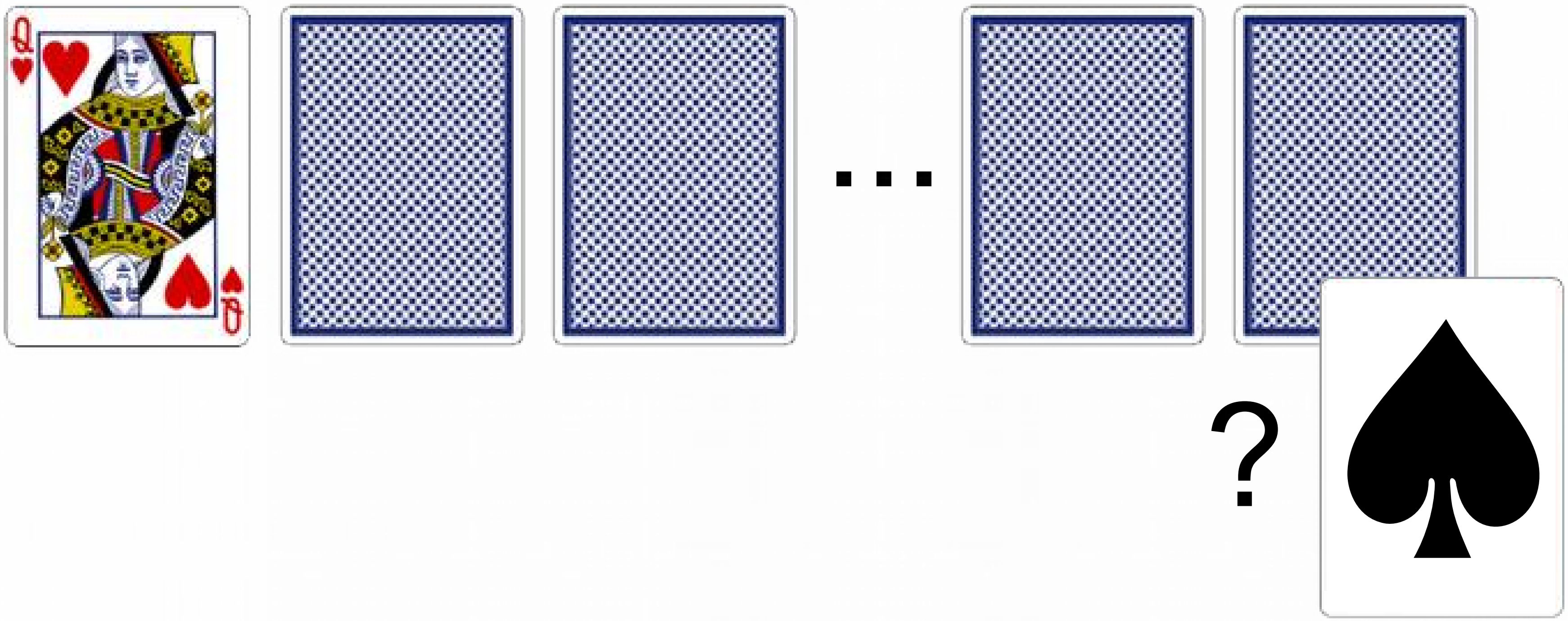


A key result: Treewidth

Why? **Conditional Independence**

Pr(A|C,E) = Pr(A|C) P(A|B,E,F) = P(A|B,E) P(A|B,E,F) ≠ P(A|B,E)

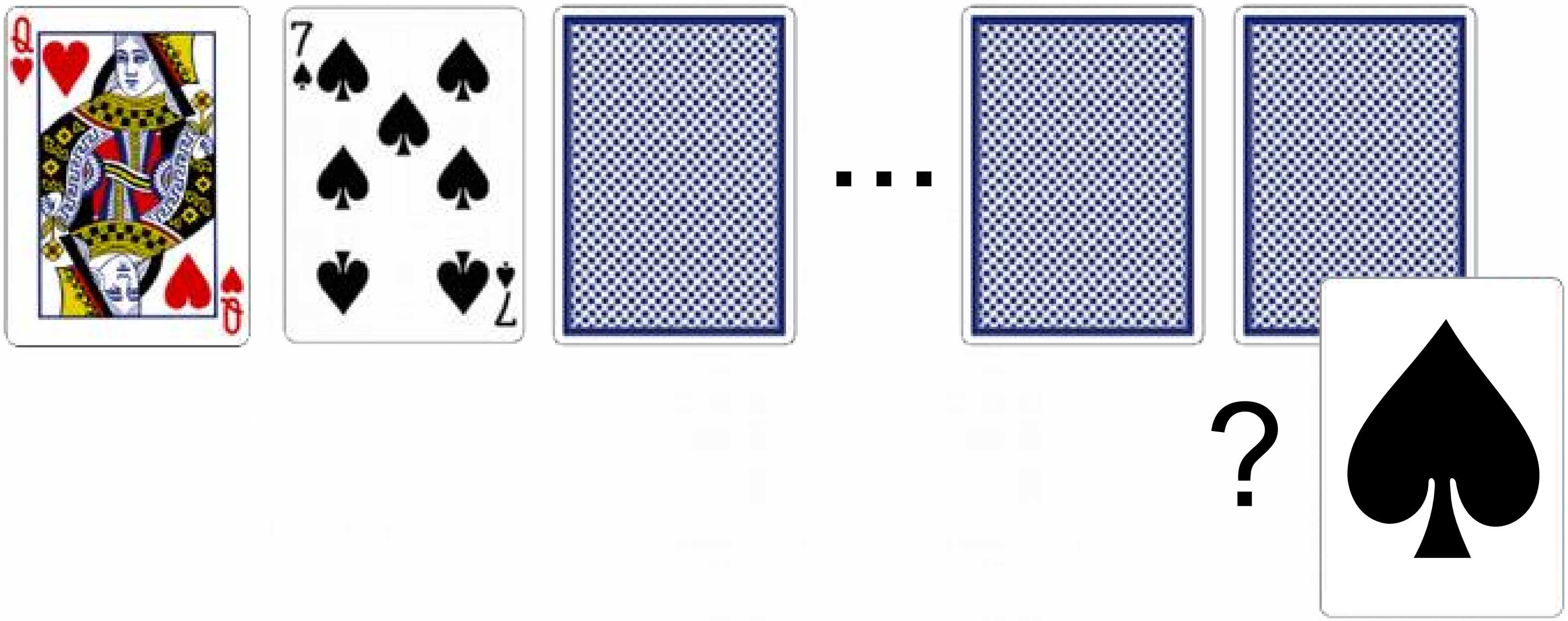
Is There Conditional Independence?



Probability 13/51

$$\Pr(\text{Card52} \mid \text{Card1}, \text{Card2}) \stackrel{?}{=} \Pr(\text{Card52} \mid \text{Card1})$$

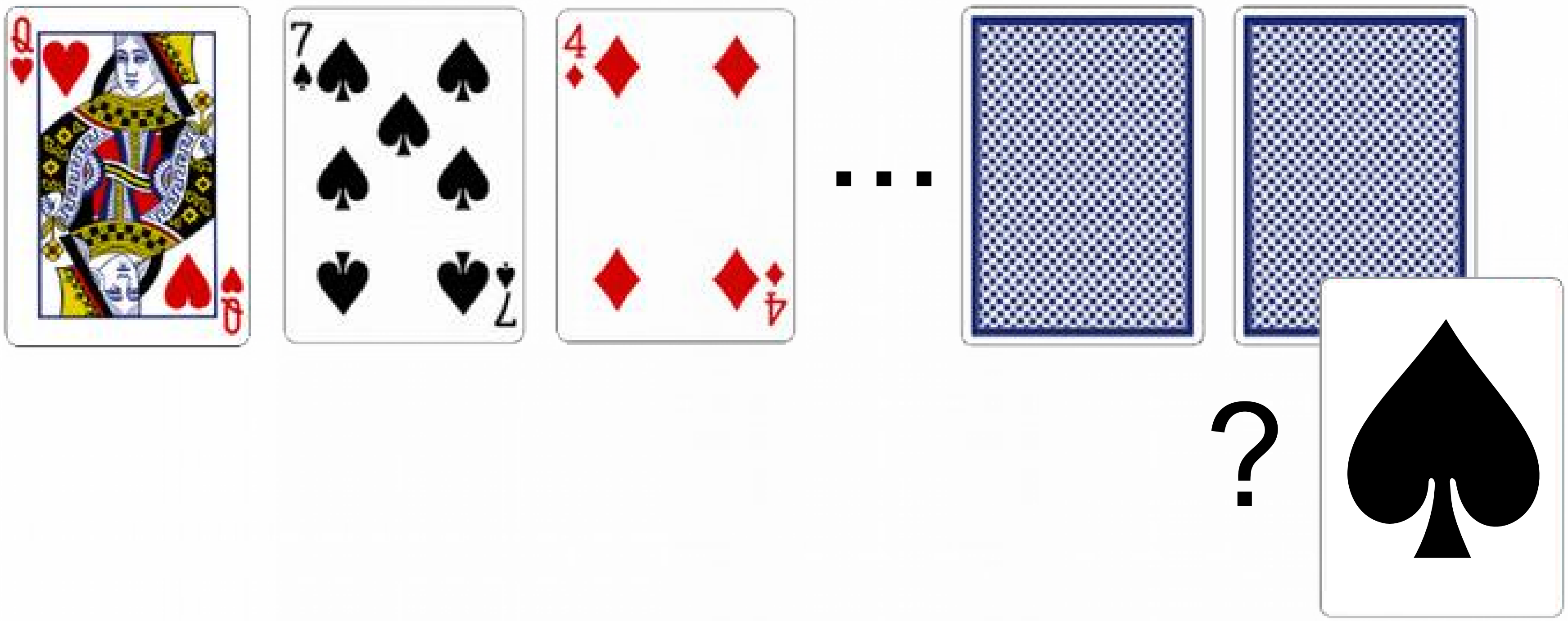
Is There Conditional Independence?



Probability 12/50

$$\Pr(\text{Card52} \mid \text{Card1, Card2, Card3}) \stackrel{?}{=} \Pr(\text{Card52} \mid \text{Card1, Card2})$$

Is There Conditional Independence?

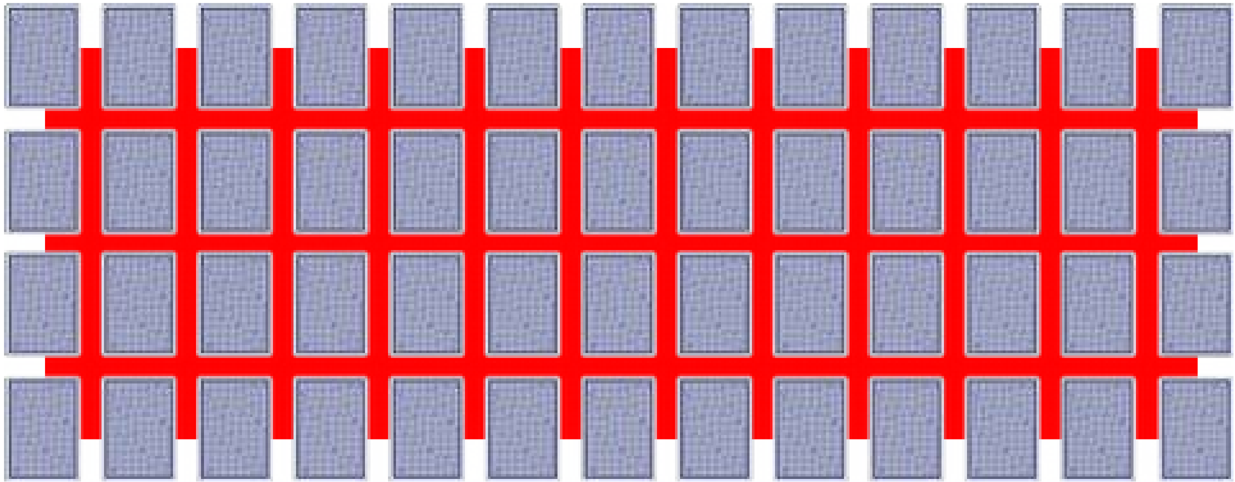


Probability 12/49

Automated Reasoning

Let us automate this:

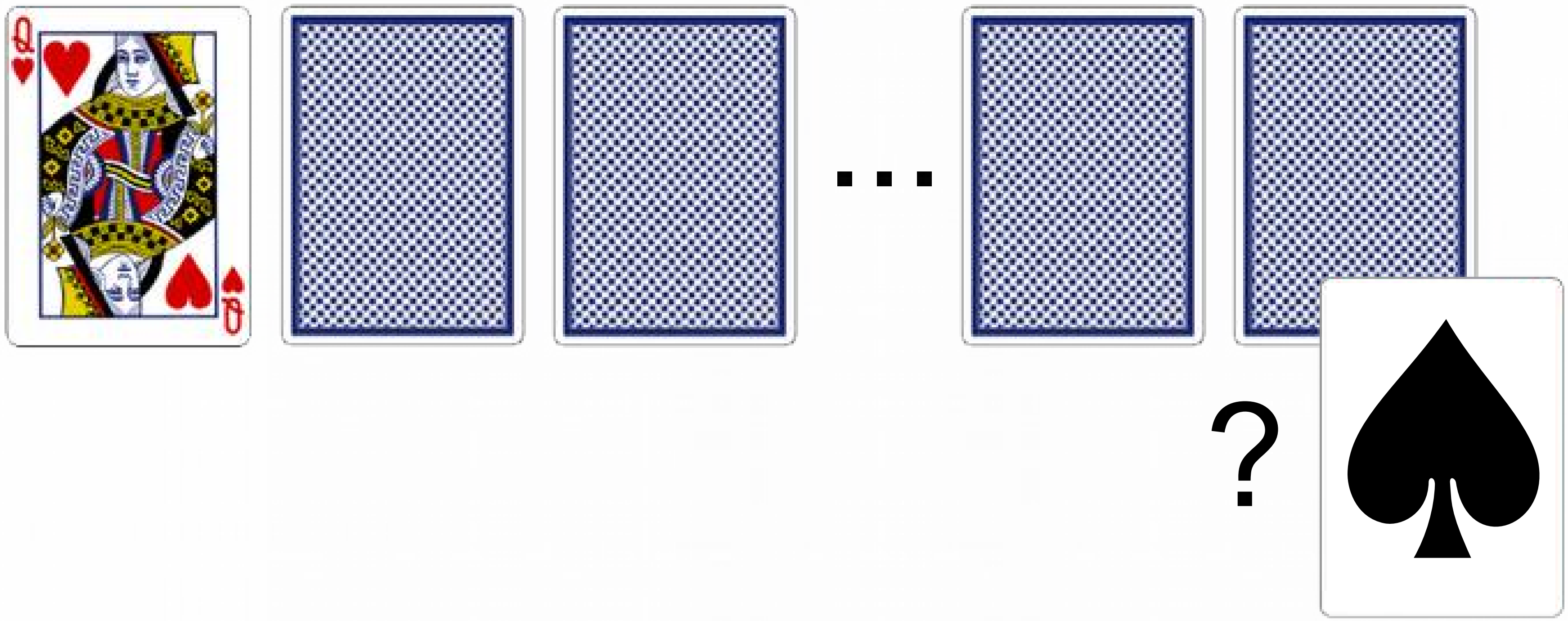
1. Probabilistic propositional model
   is fully connected!



(artist's impression)

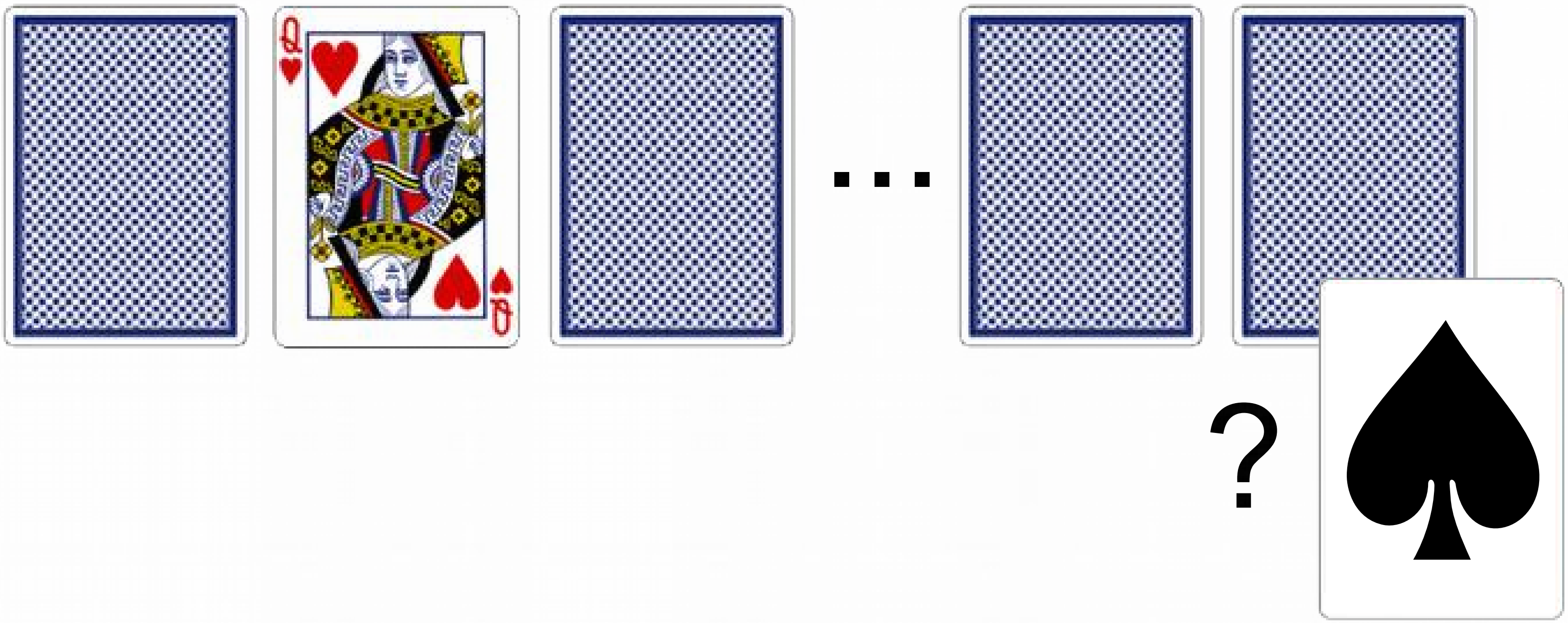
2. Probabilistic inference algorithm (VE)
   builds a table with 13^{52} rows (or equivalent)

What's Going On Here?



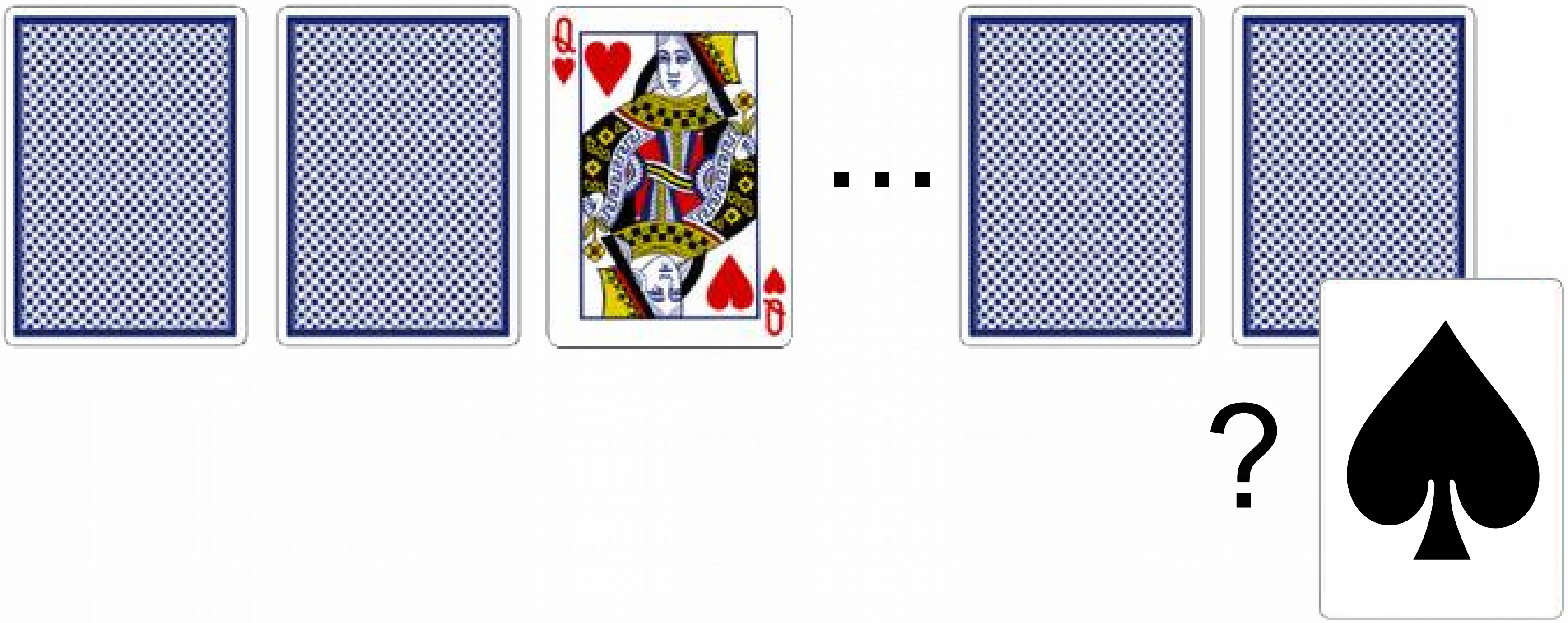
Probability 13/51

What's Going On Here?



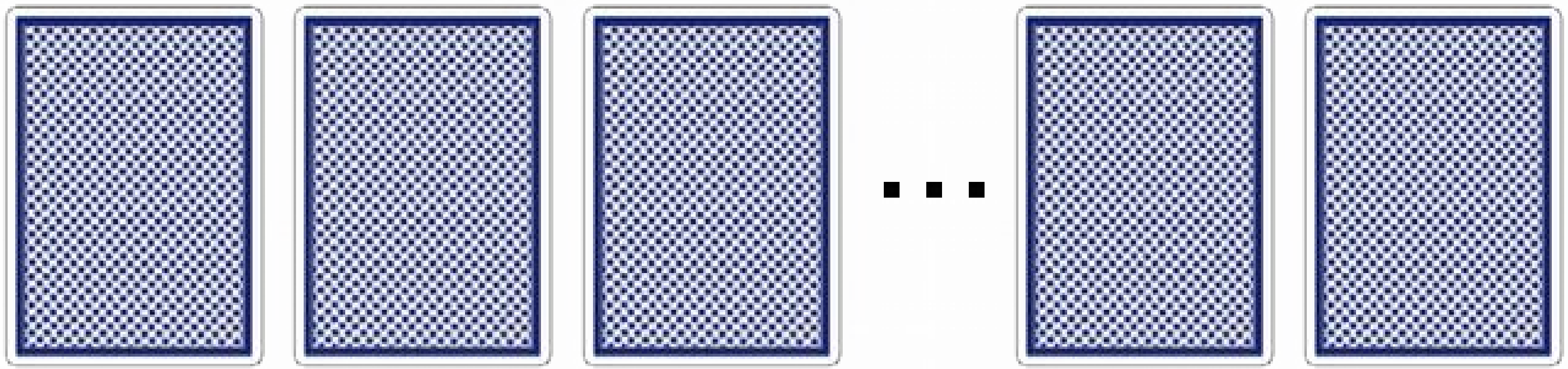
Probability 13/51

What's Going On Here?



Probability 13/51

Tractable Probabilistic Inference



Which property makes inference tractable?

- Traditional belief: Independence (conditional/contextual)
- What's going on here?
  - Symmetry
  - Exchangebility

⇒ Lifted Inference

[Niepert-AAAI14]

Automated Reasoning

Let us automate this:

- **Relational** model

$$\forall p,x,y,\ Card(p,x) \wedge Card(p,y) \Rightarrow x = y$$
$$\forall c,x,y,\ Card(x,c) \wedge Card(y,c) \Rightarrow x = y$$

- **Lifted** probabilistic inference algorithm

Other Examples of Lifted Inference

- First-Order resolution

$$\forall x, Human(x) \Rightarrow Mortal(x)$$
$$\forall x, Greek(x) \Rightarrow Human(x)$$

then

$$\forall x, Greek(x) \Rightarrow Mortal(x)$$

Other Examples of Lifted Inference

- First-Order resolution
- Reasoning about populations

We are investigating a rare disease. The disease is more rare in women, presenting only in **one in every two billion women** and **one in every billion men**. Then, assuming there are **3.4 billion men** and **3.6 billion women** in the world, the probability that **more than five people** have the disease is

$$1 - \sum_{n=0}^{5} \sum_{f=0}^{n} \binom{3.6 \cdot 10^9}{f} \left(1 - 0.5 \cdot 10^{-9}\right)^{3.6 \cdot 10^9 - f} \left(0.5 \cdot 10^{-9}\right)^{f}$$

$$\times \binom{3.4 \cdot 10^9}{(n-f)} \left(1 - 10^{-9}\right)^{3.4 \cdot 10^9 - (n-f)} \left(10^{-9}\right)^{(n-f)}$$

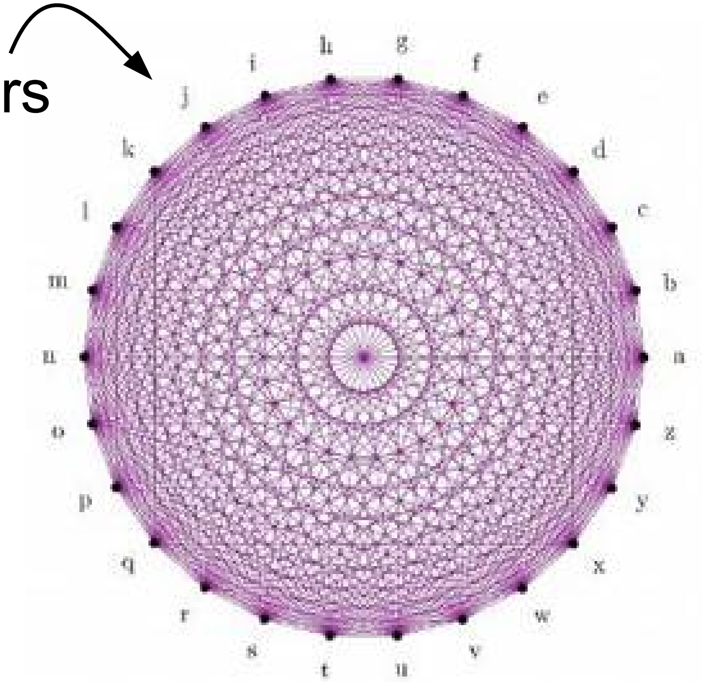
Relational Representations

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

$$3.14\ \ \mathrm{FacultyPage}(x) \wedge \mathrm{Linked}(x,y) \Rightarrow \mathrm{CoursePage}(y)$$

- As a probabilistic graphical model:
  - 26 pages, 728 random variables, 676 factors
  - 1000 pages, 1,002,000 random variables, 1,000,000 factors
- Highly intractable?

  Lifted inference in milliseconds!



A Formal Definition of Lifting

- Informal

  Exploit symmetries, Reason at first-order level, Reason about groups of objects, Scalable inference

- Formal Definition: **Domain-lifted inference**

> Probabilistic inference runs in time **polynomial** in the **number of objects** in the domain.

  - polynomial in #people, #webpages, #cards
  - <u>not</u> polynomial in #predicates, #formulas, #logical variables

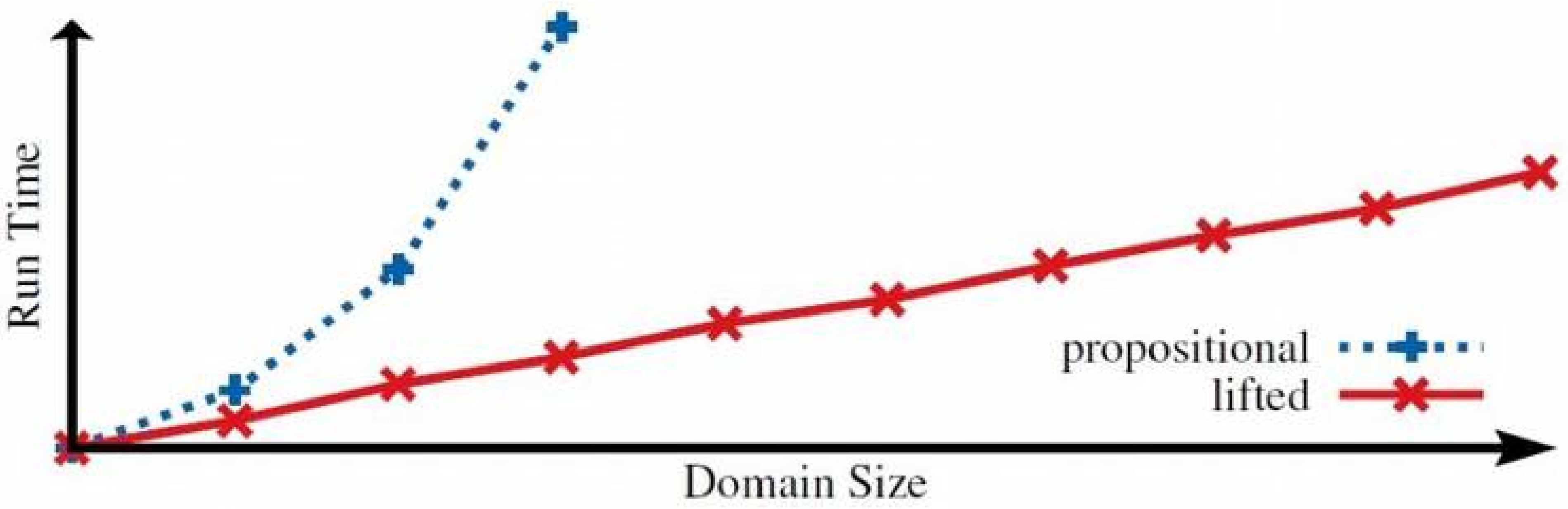
[VdB-NIPS11]

A Formal Definition of Lifting

- Informal

  Exploit symmetries, Reason at first-order level, Reason about groups of objects, Scalable inference

- Formal Definition: **Domain-lifted inference**



[VdB-NIPS11, Jaeger-StarAI12]

Overview

1. What are statistical relational models?
2. What is lifted inference?
3. **How does lifted inference work?**
4. Theoretical insights
5. Practical applications

Lifted Algorithms (in the AI community)

- Exact Probabilistic Inference
  - First-Order Variable Elimination [Poole-IJCAI03, Braz-IJCAI05, Milch-AAAI08, Taghipour-JAIR13]
  - First-Order Knowledge Compilation [VdB-IJCAI11, VdB-NIPS11, VdB-AAAI12, VdB-Thesis13]
  - Probabilistic Theorem Proving [Gogate-UAI11]
- Approximate Probabilistic Inference
  - Lifted Belief Propagation [Jaimovich-UAI07, Singla-AAAI08, Kersting-UAI09]
  - Lifted Bisimulation/Mini-buckets [Sen-VLDB08, Sen-UAI09]
  - Lifted Importance Sampling [Gogate-UAI11, Gogate-AAAI12]
  - Lifted Relax, Compensate & Recover (Generalized BP) [VdB-UAI12]
  - Lifted MCMC [Niepert-UAI12, Niepert-AAAI13, Venugopal-NIPS12]
  - Lifted Variational Inference [Choi-UAI12, Bui-StarAI12]
  - Lifted MAP-LP [Mladenov-AISTATS14, Apsel-AAAI14]
- Special-Purpose Inference:
  - Lifted Kalman Filter [Ahmadi-IJCAI11, Choi-IJCAI11]
  - Lifted Linear Programming [Mladenov-AISTATS12]

Lifted Algorithms (in the AI community)

- Exact Probabilistic Inference
  - First-Order Variable Elimination [Poole-IJCAI03, Braz-IJCAI05, Milch-AAAI08, Taghipour-JAIR13]
  - **First-Order Knowledge Compilation** [VdB-IJCAI11, VdB-NIPS11, VdB-AAAI12, VdB-Thesis13]
  - Probabilistic Theorem Proving [Gogate-UAI11]
- Approximate Probabilistic Inference
  - Lifted Belief Propagation [Jaimovich-UAI07, Singla-AAAI08, Kersting-UAI09]
  - Lifted Bisimulation/Mini-buckets [Sen-VLDB08, Sen-UAI09]
  - Lifted Importance Sampling [Gogate-UAI11, Gogate-AAAI12]
  - Lifted Relax, Compensate & Recover (Generalized BP) [VdB-UAI12]
  - Lifted MCMC [Niepert-UAI12, Niepert-AAAI13, Venugopal-NIPS12]
  - Lifted Variational Inference [Choi-UAI12, Bui-StarAI12]
  - Lifted MAP-LP [Mladenov-AISTATS14, Apsel-AAAI14]
- Special-Purpose Inference:
  - Lifted Kalman Filter [Ahmadi-IJCAI11, Choi-IJCAI11]
  - Lifted Linear Programming [Mladenov-AISTATS12]

Assembly Language for Lifted Probabilistic Inference

Computing conditional probabilities with:

- Parfactor graphs
- Markov logic networks
- Probabilistic datalog/logic programs
- Probabilistic databases
- Relational Bayesian networks

All reduces to
weighted (first-order) model counting

[VdB-IJCAI11, Gogate-UAI11, VdB-KR14, Gribkoff-UAI14]

Weighted First-Order Model Counting

A vocabulary

Smokes(Alice)	Smokes(Bob)	Friends(Alice,Bob)	Friends(Bob,Alice)
0	0	0	0
⋮	⋮	⋮	⋮
1	0	1	0
⋮	⋮	⋮	⋮
1	1	1	1

Possible worlds
Logical interpretations

Weighted First-Order Model Counting

A logical theory

$$\forall x,y,\ \mathrm{Smokes}(x) \wedge \mathrm{Friends}(x,y) \Rightarrow \mathrm{Smokes}(y)$$

Smokes(Alice)	Smokes(Bob)	Friends(Alice,Bob)	Friends(Bob,Alice)	theory
0	0	0	0	1
⋮	⋮	⋮	⋮	⋮
1	0	1	0	0
⋮	⋮	⋮	⋮	⋮
1	1	1	1	1

Interpretations that satisfy the theory

Models

Weighted First-Order Model Counting

A logical theory

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

Smokes(Alice)	Smokes(Bob)	Friends(Alice,Bob)	Friends(Bob,Alice)	theory
0	0	0	0	1
⋮	⋮	⋮	⋮	⋮
1	0	1	0	0
⋮	⋮	⋮	⋮	⋮
1	1	1	1	1

Σ

First-order model count
~ #SAT

Weighted First-Order Model Counting

A logical theory and a weight function for predicates

Smokes(Alice)	Smokes(Bob)	Friends(Alice,Bob)	Friends(Bob,Alice)	theory	weight
0	0	0	0	1	$2 \cdot 2 \cdot 1 \cdot 1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	0	1	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	1	1	1	1	$1 \cdot 1 \cdot 4 \cdot 4$

$$
\begin{aligned}
\text{Smokes} &\rightarrow 1 \\
\neg\text{Smokes} &\rightarrow 2 \\
\text{Friends} &\rightarrow 4 \\
\neg\text{Friends} &\rightarrow 1
\end{aligned}
$$

Weighted First-Order Model Counting

A logical theory and a weight function for predicates

Smokes(Alice)	Smokes(Bob)	Friends(Alice,Bob)	Friends(Bob,Alice)	theory	weight
0	0	0	0	1	$2 \cdot 2 \cdot 1 \cdot 1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	0	1	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	1	1	1	1	$1 \cdot 1 \cdot 4 \cdot 4$

Smokes $\rightarrow$ 1
¬Smokes $\rightarrow$ 2
Friends $\rightarrow$ 4
¬Friends $\rightarrow$ 1

Σ

Weighted first-order model count
~ Partition function

Example:
First-Order Model Counting

1. Logical sentence

$$\text{Stress}(\text{Alice}) \Rightarrow \text{Smokes}(\text{Alice})$$

Domain

Alice

Example:
First-Order Model Counting

1. Logical sentence

Stress(Alice) ⇒ Smokes(Alice)

Domain

Alice

Stress(Alice)	Smokes(Alice)	Formula
0	0	1
0	1	1
1	0	0
1	1	1

Example:
First-Order Model Counting

1. Logical sentence

$$\text{Stress}(\text{Alice}) \Rightarrow \text{Smokes}(\text{Alice})$$

→ 3 models

Domain

Alice

Example:
First-Order Model Counting

1. Logical sentence

Stress(Alice) ⇒ Smokes(Alice)

→ 3 models

Domain

Alice

2. Logical sentence

∀x, Stress(x) ⇒ Smokes(x)

Domain

Alice

Example: First-Order Model Counting

1. Logical sentence

$\text{Stress}(\text{Alice}) \Rightarrow \text{Smokes}(\text{Alice})$

Domain

Alice

→ 3 models

2. Logical sentence

$\forall x, \text{Stress}(x) \Rightarrow \text{Smokes}(x)$

Domain

Alice

→ 3 models

Example:
First-Order Model Counting

2. Logical sentence

$$\boxed{\forall x, \text{Stress}(x) \Rightarrow \text{Smokes}(x)}$$

Domain

$$\boxed{\text{Alice}}$$

$\rightarrow$ 3 models

Example:
First-Order Model Counting

2. Logical sentence

$\forall x, Stress(x) \Rightarrow Smokes(x)$

Domain

Alice

$\rightarrow$ 3 models

3. Logical sentence

$\forall x, Stress(x) \Rightarrow Smokes(x)$

Domain

n people

Example:
First-Order Model Counting

2. Logical sentence

$\forall x, \text{Stress}(x) \Rightarrow \text{Smokes}(x)$

Domain

Alice

$\rightarrow$ 3 models

3. Logical sentence

$\forall x, \text{Stress}(x) \Rightarrow \text{Smokes}(x)$

Domain

n people

$\rightarrow 3^n$ models

Example: First-Order Model Counting

3. Logical sentence

$\forall x, \text{Stress}(x) \Rightarrow \text{Smokes}(x)$

Domain

n people

$\rightarrow 3^n$ models

Example:
First-Order Model Counting

3. Logical sentence

$\forall x, \mathrm{Stress}(x) \Rightarrow \mathrm{Smokes}(x)$

Domain

n people

$\rightarrow 3^n$ models

4. Logical sentence

$\forall y, \mathrm{ParentOf}(y) \wedge \mathrm{Female} \Rightarrow \mathrm{MotherOf}(y)$

Domain

n people

Example:
First-Order Model Counting

3. Logical sentence

$\forall x, \text{Stress}(x) \Rightarrow \text{Smokes}(x)$

Domain

n people

$\rightarrow 3^n$ models

4. Logical sentence

$\forall y, \text{ParentOf}(y) \wedge \text{Female} \Rightarrow \text{MotherOf}(y)$

Domain

n people

if Female: $\forall y, \text{ParentOf}(y) \Rightarrow \text{MotherOf}(y)$

Example: First-Order Model Counting

3. Logical sentence

$\forall x, Stress(x) \Rightarrow Smokes(x)$

Domain

n people

$\rightarrow 3^n$ models

4. Logical sentence

$\forall y, ParentOf(y) \wedge Female \Rightarrow MotherOf(y)$

Domain

n people

if not Female: True

Example: First-Order Model Counting

3. Logical sentence

$$\boxed{\forall x, \mathrm{Stress}(x) \Rightarrow \mathrm{Smokes}(x)}$$

Domain

$\boxed{\text{n people}}$

$\rightarrow 3^n$ models

4. Logical sentence

$$\boxed{\forall y, \mathrm{ParentOf}(y) \wedge \mathrm{Female} \Rightarrow \mathrm{MotherOf}(y)}$$

Domain

$\boxed{\text{n people}}$

$\rightarrow (3^n+4^n)$ models

Example: First-Order Model Counting

4. Logical sentence

$\boxed{\forall y, \text{ParentOf}(y) \wedge \text{Female} \Rightarrow \text{MotherOf}(y)}$

Domain

$\boxed{\text{n people}}$

$\rightarrow (3^n+4^n)$ models

Example:
First-Order Model Counting

4. Logical sentence

$\forall y$, ParentOf(y) $\wedge$ Female $\Rightarrow$ MotherOf(y)

Domain

n people

$\rightarrow$ (3^n+4^n) models

5. Logical sentence

$\forall x,y$, ParentOf(x,y) $\wedge$ Female(x) $\Rightarrow$ MotherOf(x,y)

Domain

n people

Example:
First-Order Model Counting

4. Logical sentence

$\boxed{\forall y,\ \mathrm{ParentOf}(y) \wedge \mathrm{Female} \Rightarrow \mathrm{MotherOf}(y)}$

Domain

$\boxed{n \text{ people}}$

$\rightarrow (3^n+4^n)$ models

5. Logical sentence

$\boxed{\forall x,y,\ \mathrm{ParentOf}(x,y) \wedge \mathrm{Female}(x) \Rightarrow \mathrm{MotherOf}(x,y)}$

Domain

$\boxed{n \text{ people}}$

$\rightarrow (3^n+4^n)^n$ models

Example: First-Order Model Counting

6. Logical sentence

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

Domain

n people

Example:
First-Order Model Counting

6. Logical sentence

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

Domain

n people

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

Example:
First-Order Model Counting

6. Logical sentence

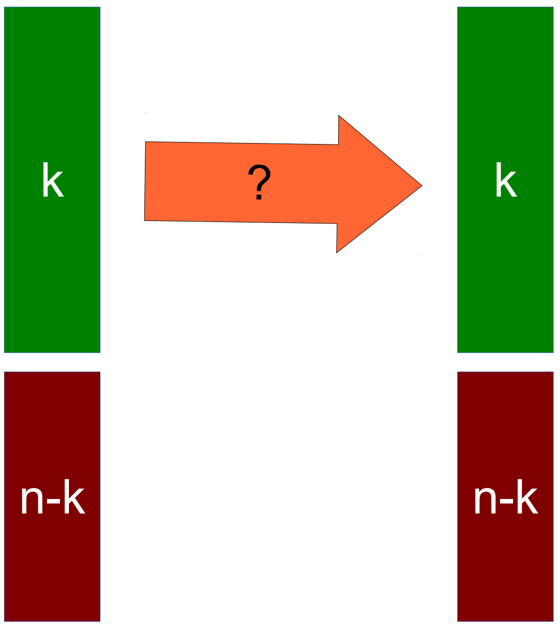
$\boxed{\forall x,y,\ Smokes(x) \wedge Friends(x,y) \Rightarrow Smokes(y)}$

Domain

$\boxed{n \text{ people}}$

- 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
...



Example: First-Order Model Counting

6. Logical sentence

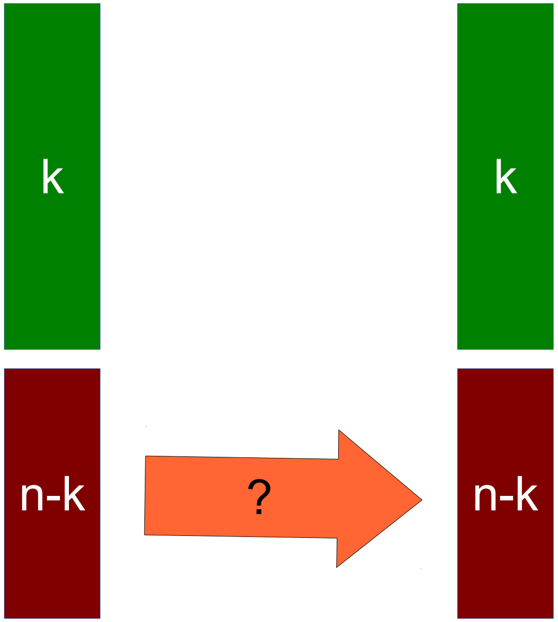
$\forall x,y,\ \mathrm{Smokes}(x) \wedge \mathrm{Friends}(x,y) \Rightarrow \mathrm{Smokes}(y)$

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
...



Example:
First-Order Model Counting

6. Logical sentence

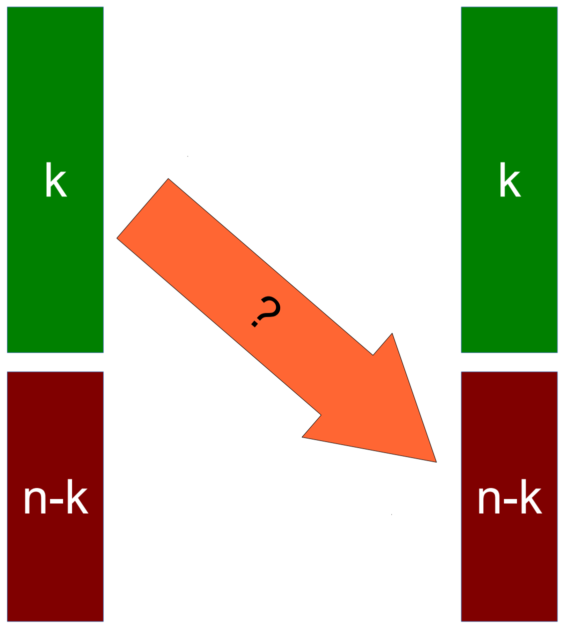
$\forall x,y,\ Smokes(x) \land Friends(x,y) \Rightarrow Smokes(y)$

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
...



Example:
First-Order Model Counting

6. Logical sentence

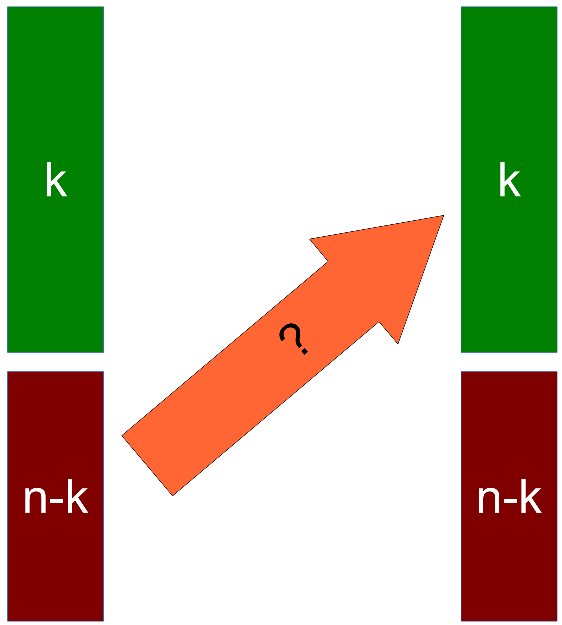
$\forall$x,y, Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

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
...



Example:
First-Order Model Counting

6. Logical sentence

$\forall$x,y, Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

Domain

n people

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

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

Example:
First-Order Model Counting

6. Logical sentence

Domain

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

n people

- If we know precisely who smokes, and there are *k* smokers
  $\rightarrow$ $2^{n^2-k(n-k)}$ models
- If we know that there are *k* smokers

Example: First-Order Model Counting

6. Logical sentence

$\forall$x,y, Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

Domain

n people

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

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

- If we know that there are *k* smokers

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

Example:
First-Order Model Counting

6. Logical sentence | Domain

∀x,y, Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y) | n people

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

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

- If we know that there are *k* smokers

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

- In total

Example:
First-Order Model Counting

6. Logical sentence | Domain

$\forall$x,y, Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

n people

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

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

The Full Pipeline

MLN $\quad$ 3.14 $\quad \mathrm{Smokes}(x) \wedge \mathrm{Friends}(x,y) \Rightarrow \mathrm{Smokes}(y)$

The Full Pipeline

MLN

$3.14 \quad Smokes(x) \wedge Friends(x,y) \Rightarrow Smokes(y)$

$\forall x,y,\ F(x,y) \Leftrightarrow [\ Smokes(x) \wedge Friends(x,y) \Rightarrow Smokes(y)\ ]$

Relational Logic

The Full Pipeline

MLN

$3.14 \quad \mathrm{Smokes}(x) \wedge \mathrm{Friends}(x,y) \Rightarrow \mathrm{Smokes}(y)$

$\forall x,y,\ F(x,y) \Leftrightarrow [\ \mathrm{Smokes}(x) \wedge \mathrm{Friends}(x,y) \Rightarrow \mathrm{Smokes}(y)\ ]$

Relational Logic

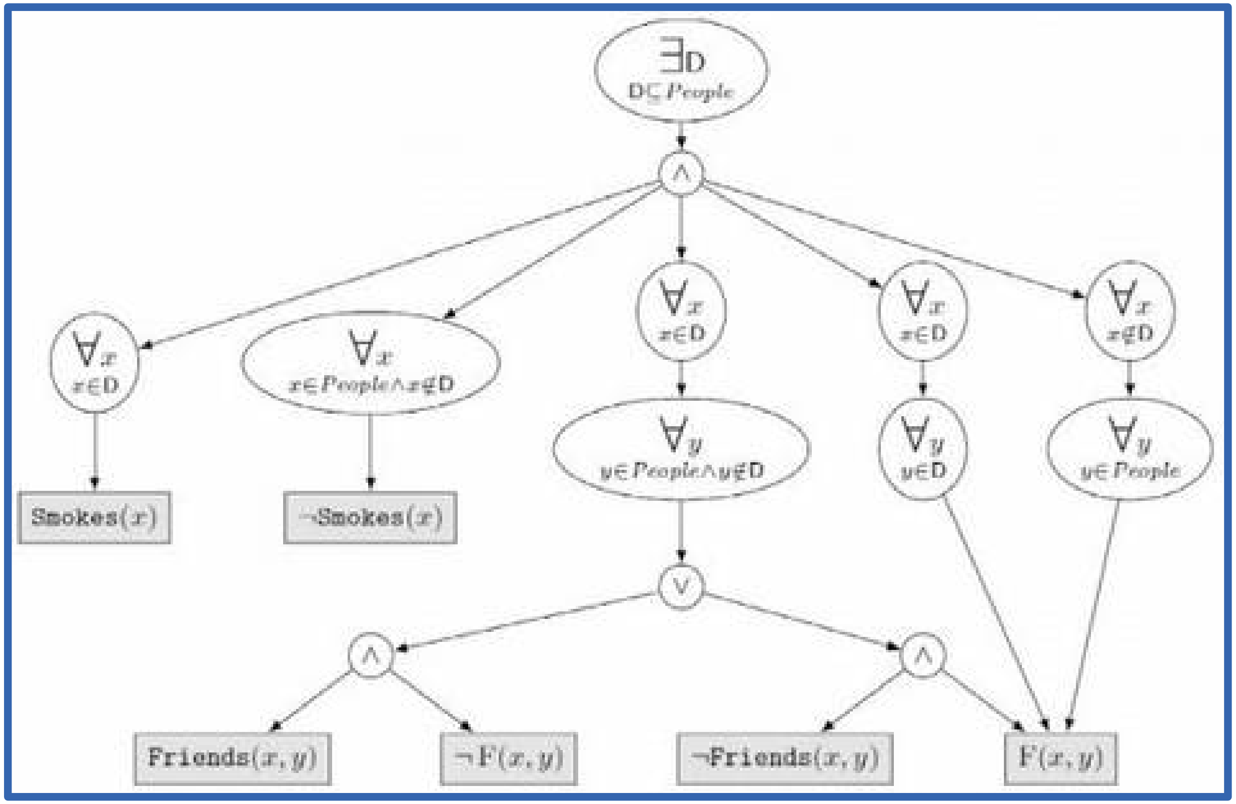
$\mathrm{Smokes} \longrightarrow 1$
$\neg\mathrm{Smokes} \longrightarrow 1$
$\mathrm{Friends} \longrightarrow 1$
$\neg\mathrm{Friends} \longrightarrow 1$
$F \longrightarrow \exp(3.14)$
$\neg F \longrightarrow 1$

Weight Function

The Full Pipeline

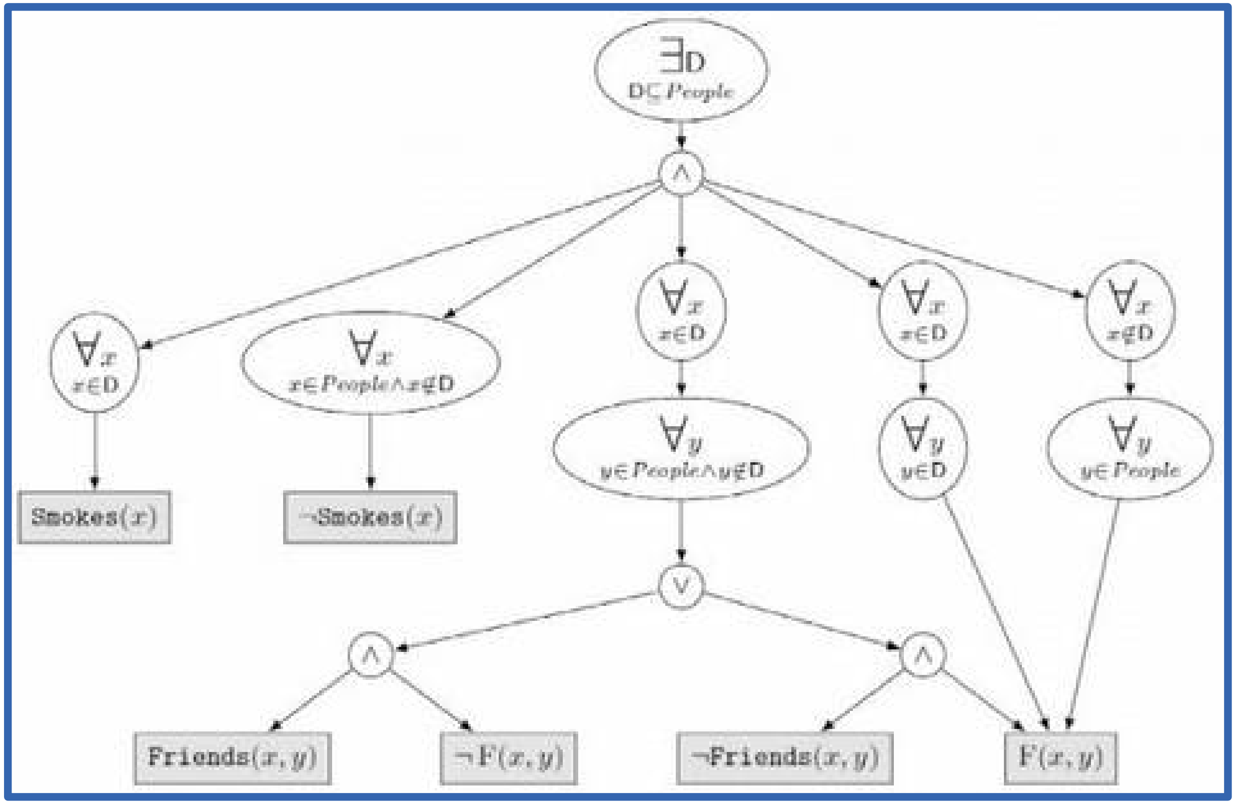
$\forall x,y,\ F(x,y) \Leftrightarrow [\ Smokes(x) \wedge Friends(x,y) \Rightarrow Smokes(y)\ ]$

Relational Logic



First-Order
d-DNNF Circuit

The Full Pipeline



First-Order d-DNNF Circuit

Smokes → 1
¬Smokes → 1
Friends → 1
¬Friends → 1
F → exp(3.14)
¬F → 1

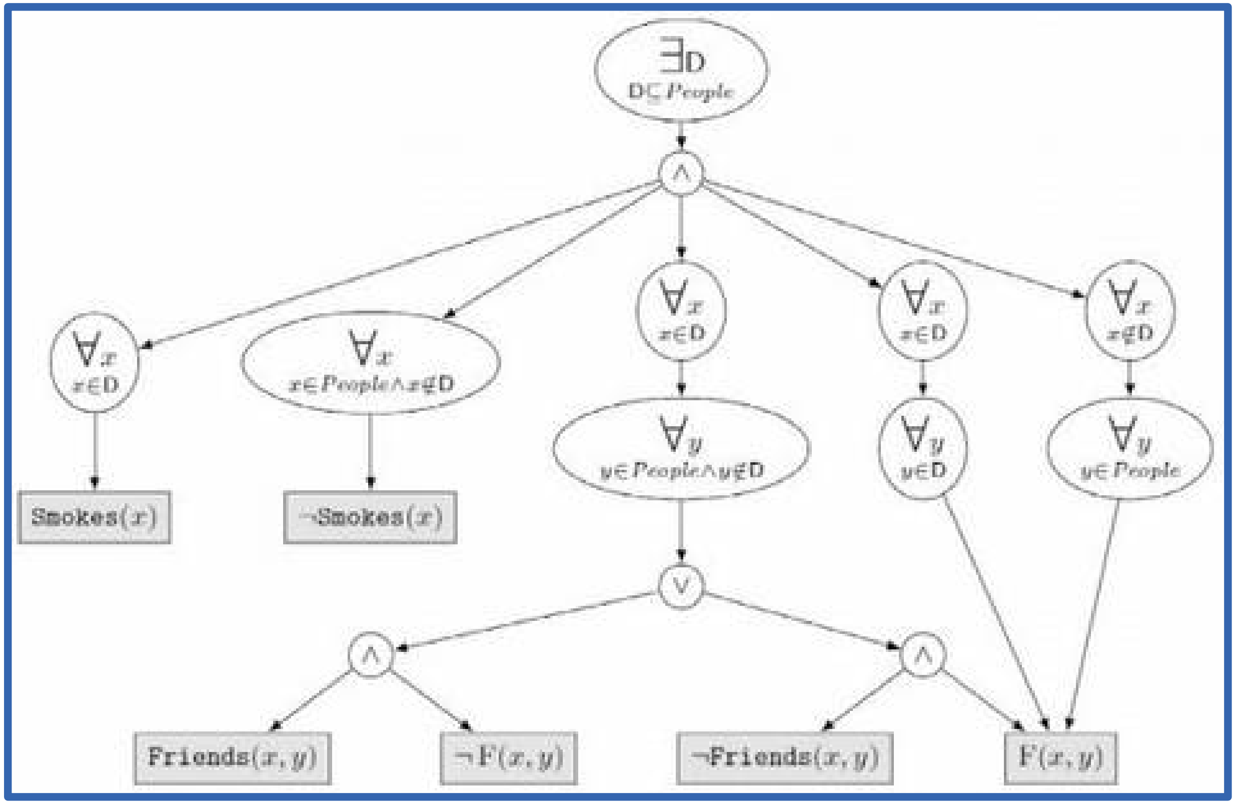
Weight Function

Alice
Bob
Charlie

Domain

Weighted First-Order Model Count is 1479.85

The Full Pipeline



First-Order d-DNNF Circuit

Smokes $\rightarrow$ 1
¬Smokes $\rightarrow$ 1
Friends $\rightarrow$ 1
¬Friends $\rightarrow$ 1
F $\rightarrow$ exp(3.14)
¬F $\rightarrow$ 1

Weight Function

Alice
Bob
Charlie

Domain

Weighted First-Order Model Count is 1479.85

Circuit evaluation is polynomial in domain size!

Assembly Language for Lifted Probabilistic Inference

Computing conditional probabilities with:

- Parfactor graphs
- Markov logic networks
- Probabilistic datalog/logic programs
- Probabilistic databases
- Relational Bayesian networks

All reduces to
weighted (first-order) model counting

[VdB-IJCAI11, Gogate-UAI11, VdB-KR14, Gribkoff-UAI14]

Overview

1. What are statistical relational models?
2. What is lifted inference?
3. How does lifted inference work?
4. **Theoretical insights**
5. Practical applications

Liftability Framework

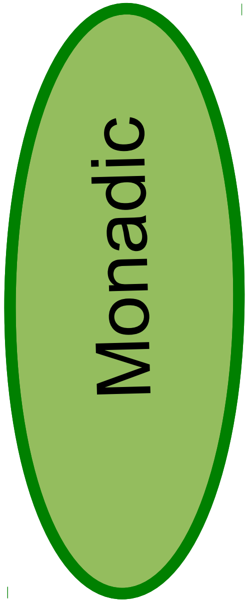
- **Domain-lifted** algorithms run in time polynomial in the domain size (~data complexity).
- A class of inference tasks C is **liftable** iff there *exists* an algorithm that
  - is domain-lifted and
  - solves all problems in C.
- Such an algorithm is **complete** for C.
- Liftability depends on the type of task.

[VdB-NIPS11, Jaeger-StarAI12]

Liftable Classes

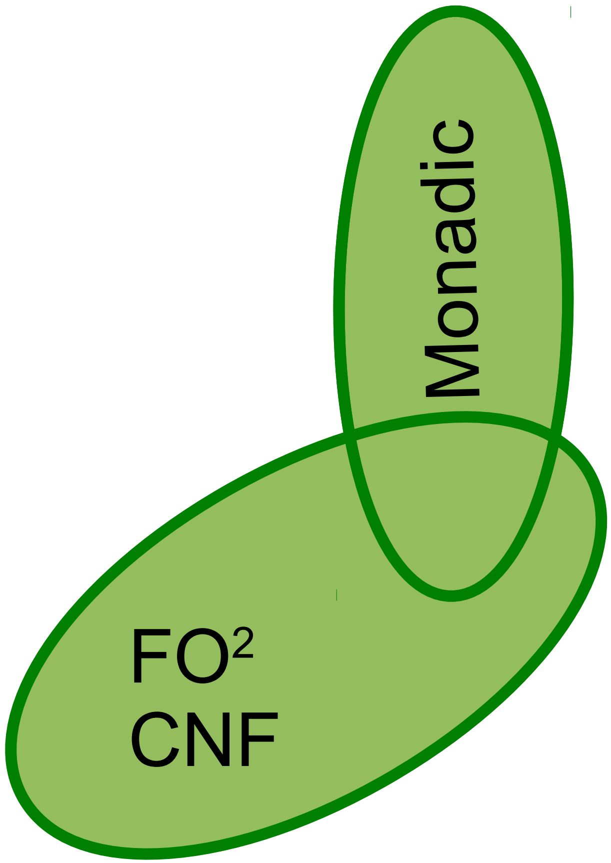
(of model counting problems)

Liftable Classes



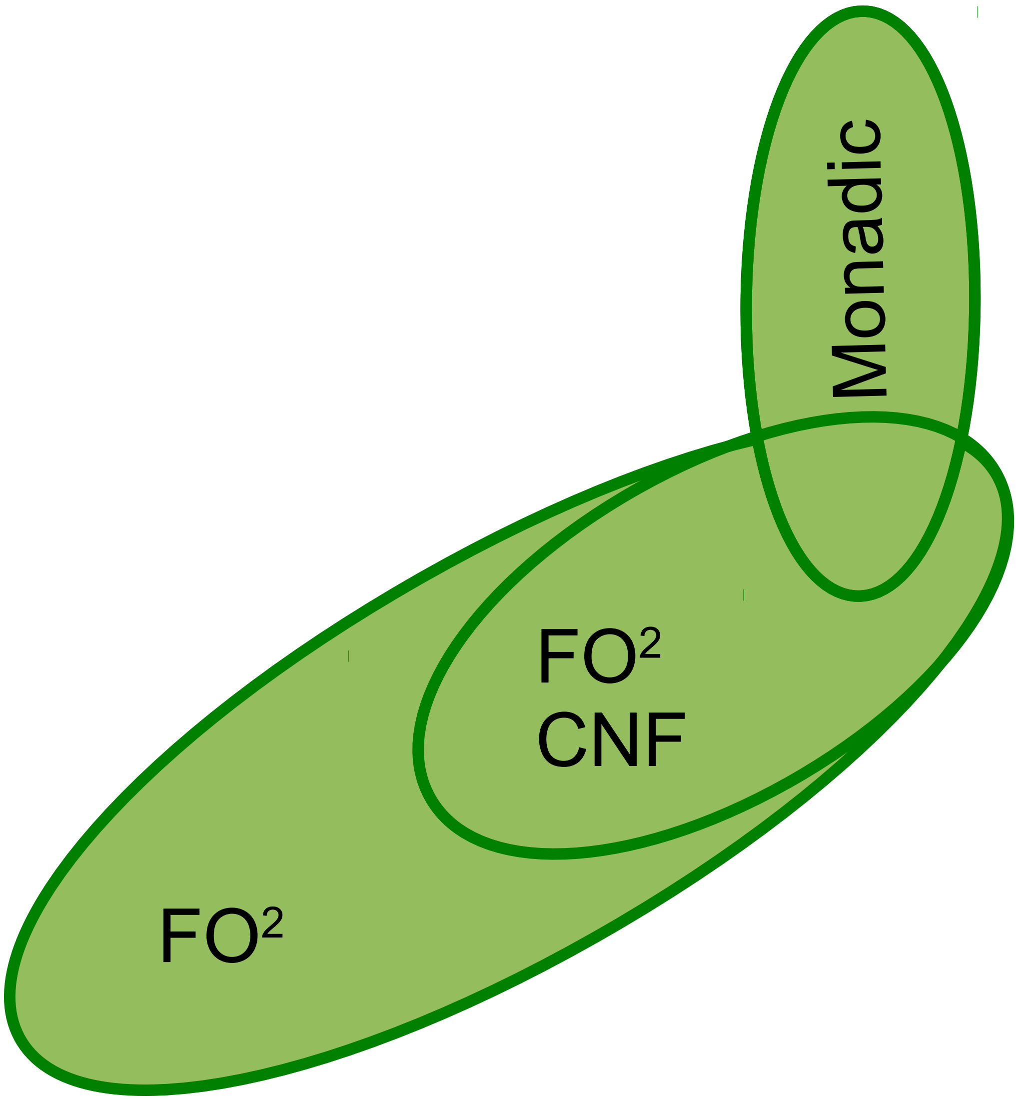
[VdB-NIPS11]

Liftable Classes



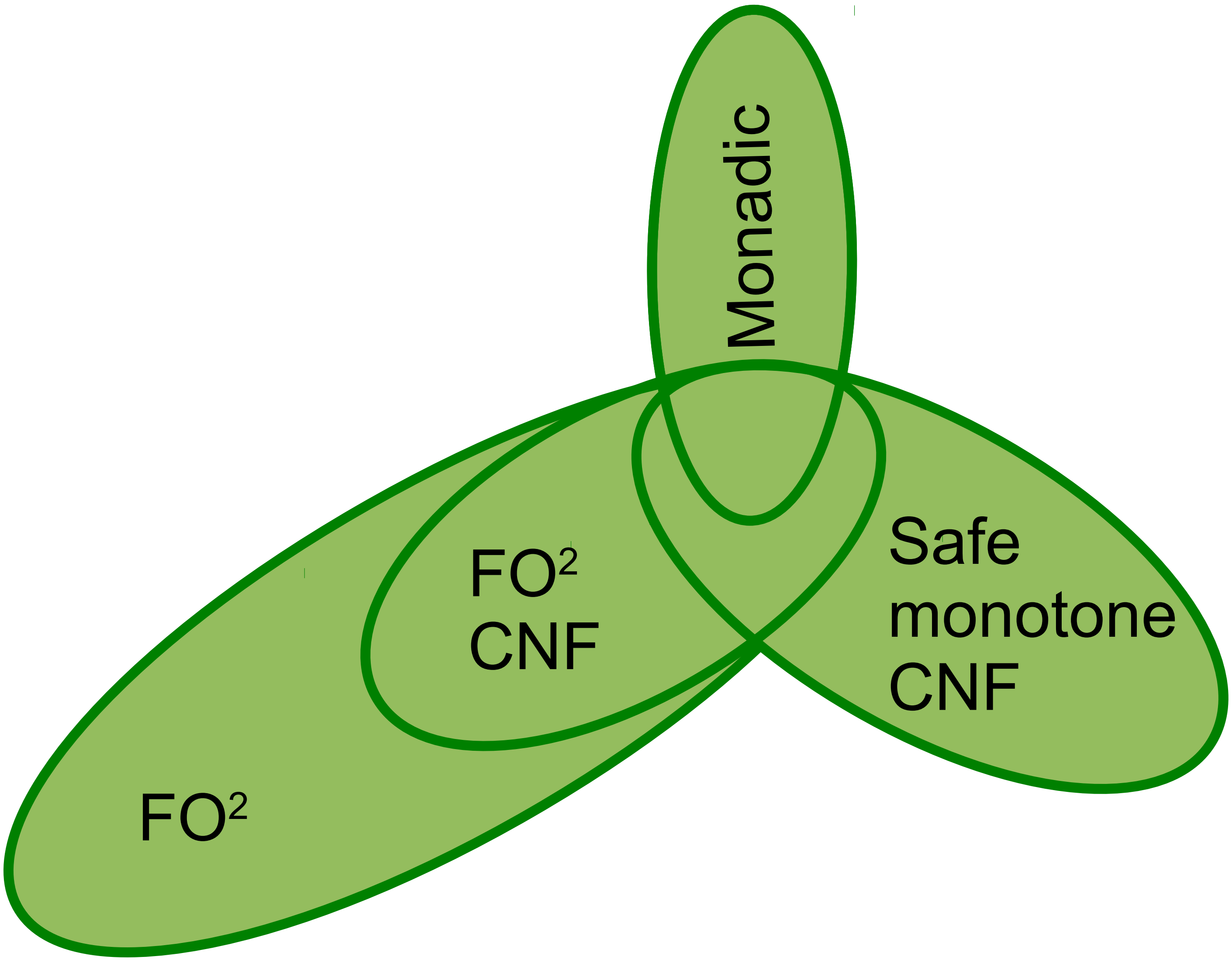
[VdB-NIPS11]

Liftable Classes



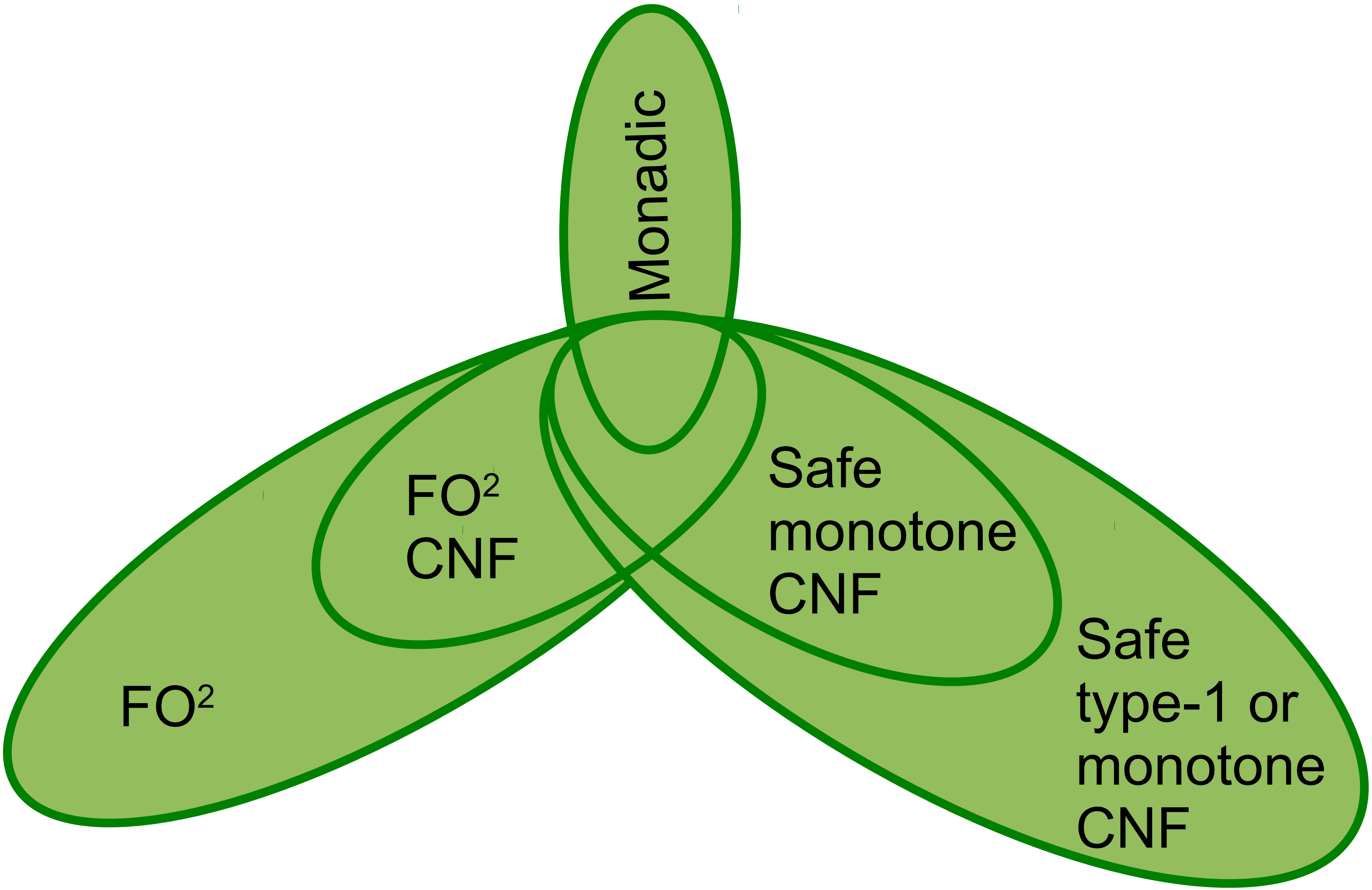
[VdB-KR14]

Liftable Classes



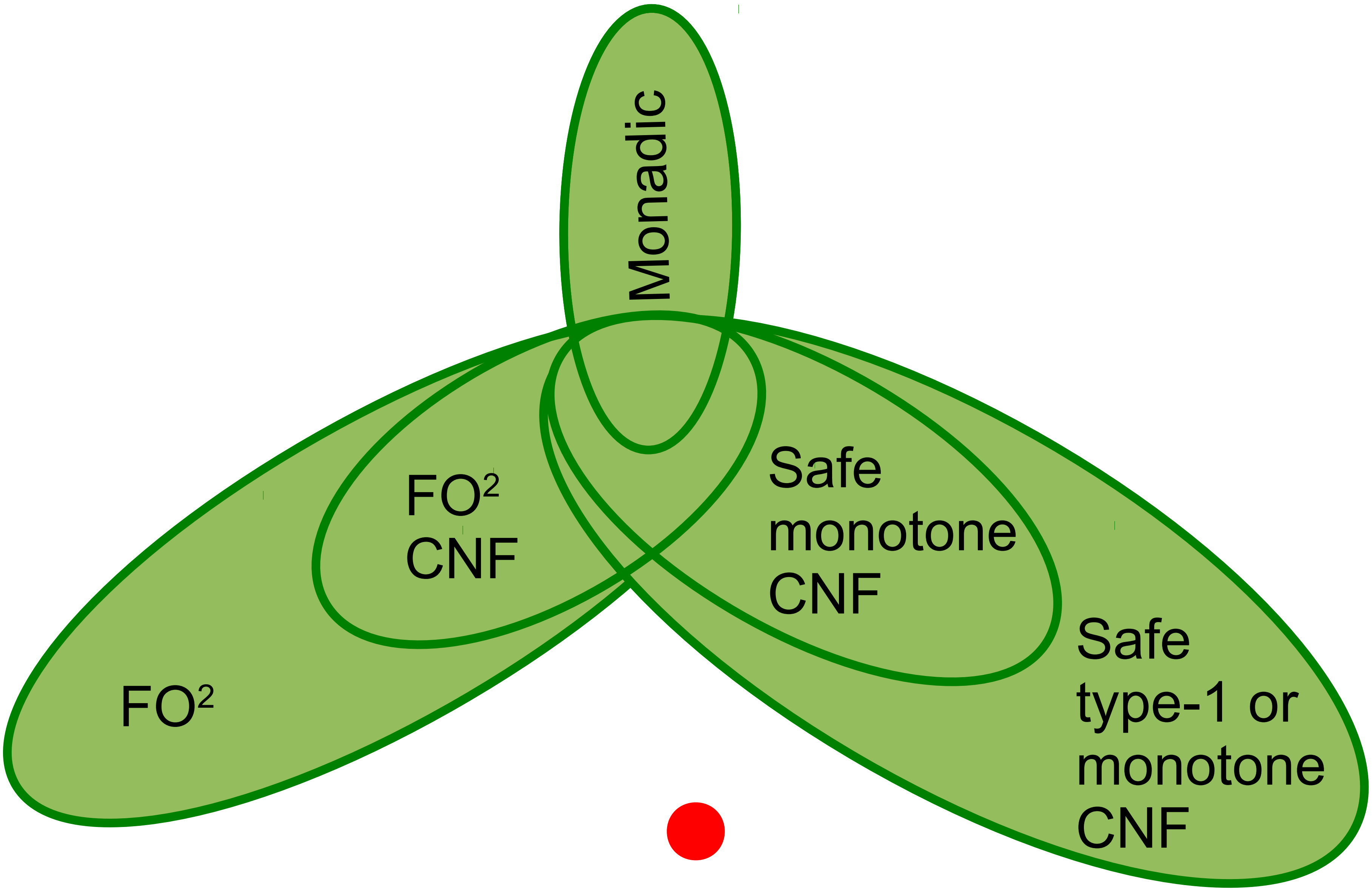
[Dalvi-JACM12]

Liftable Classes



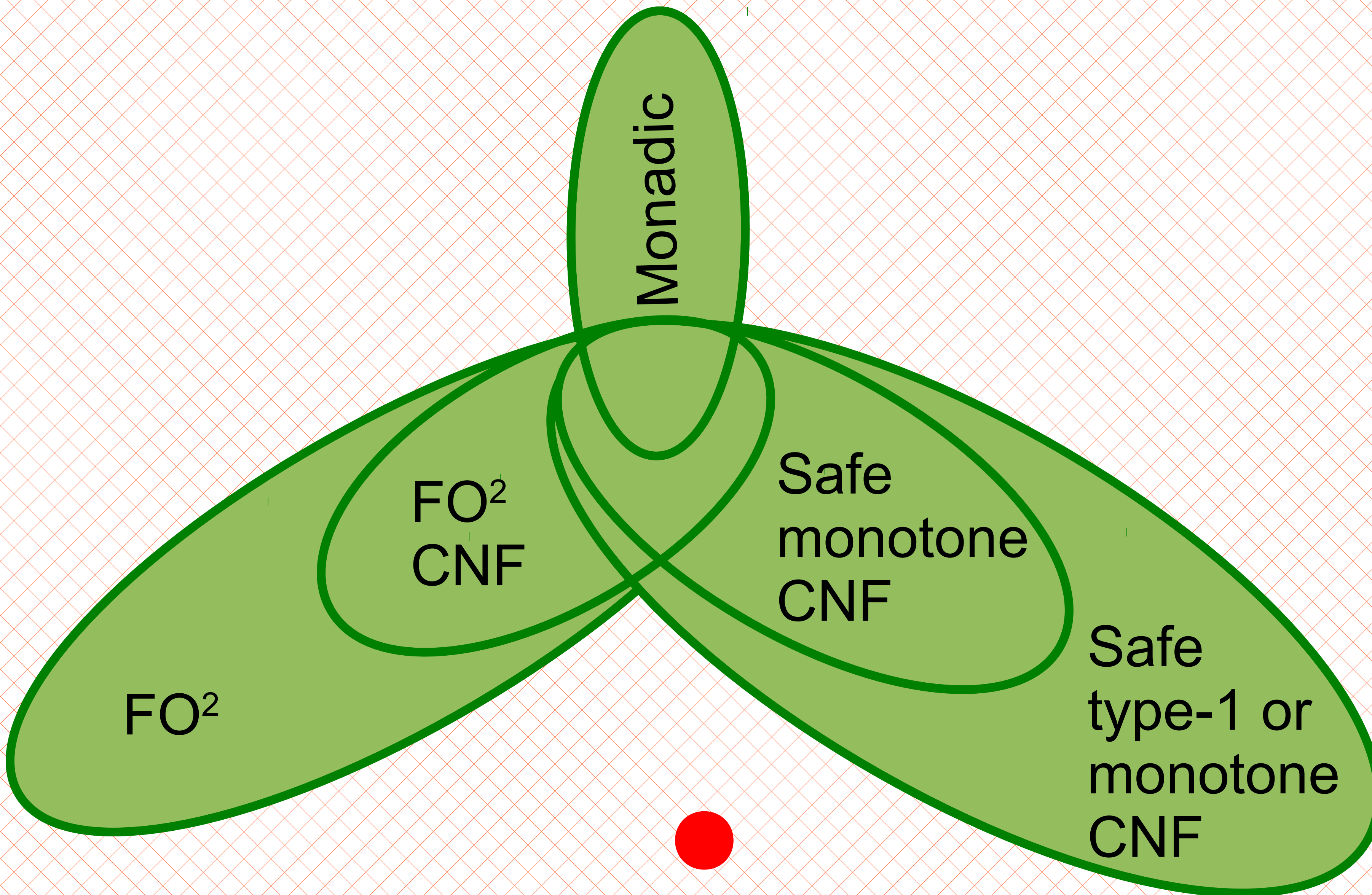
[Gribkoff-UAI14]

Liftable Classes



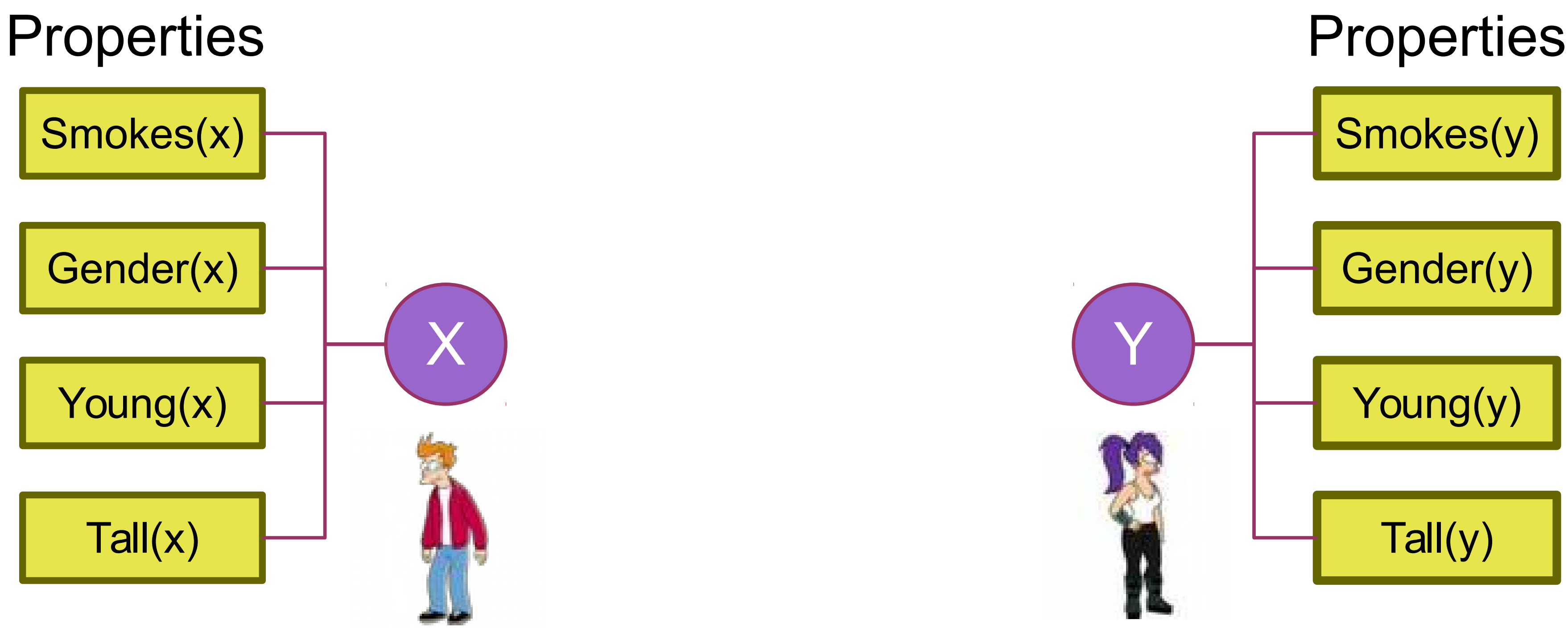
[Jaeger-StarAI12,Jaeger-TPLP12]

Liftable Classes

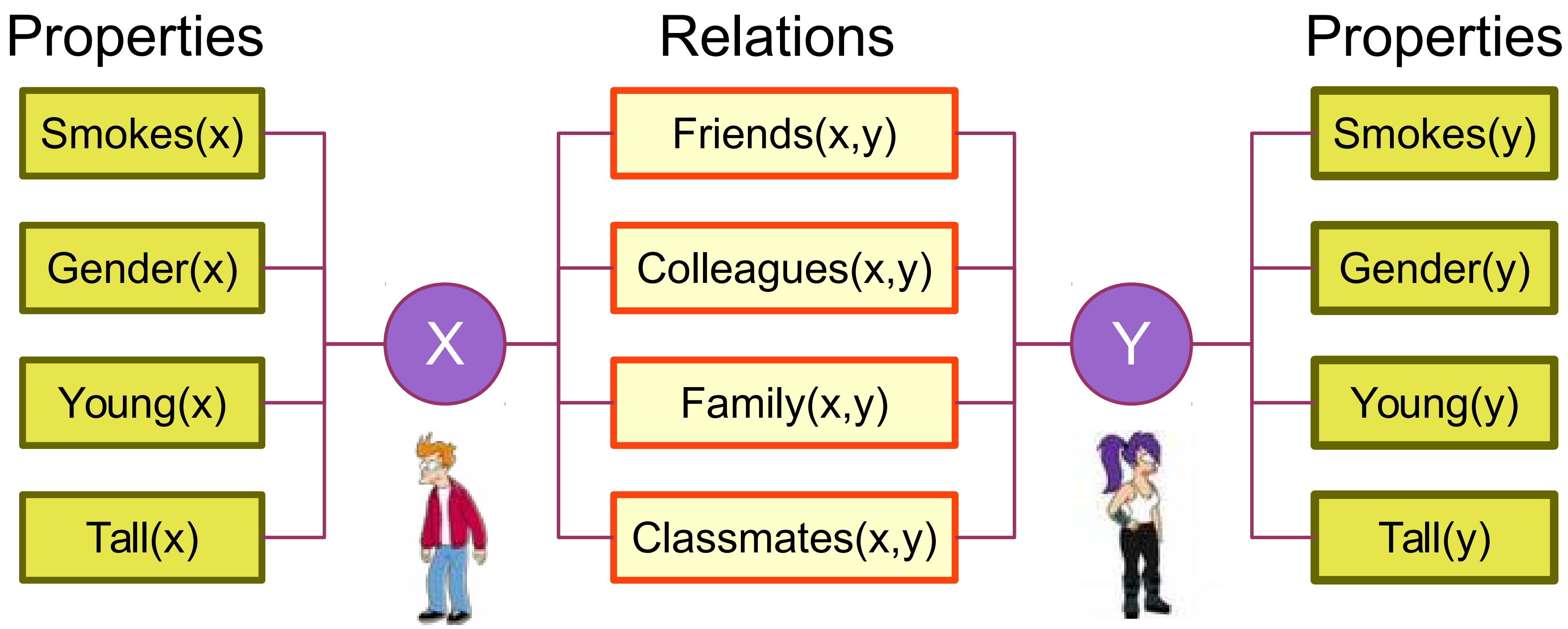


Positive Liftability Result

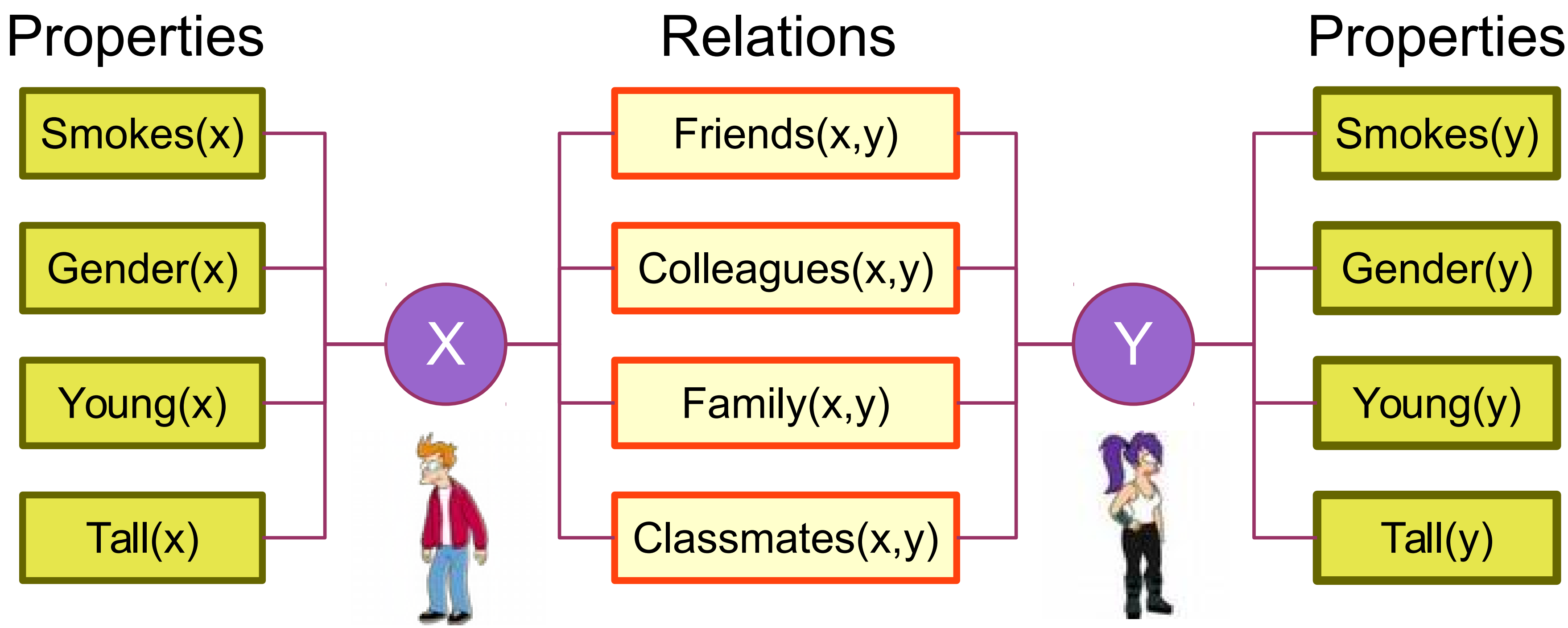
Positive Liftability Result



Positive Liftability Result

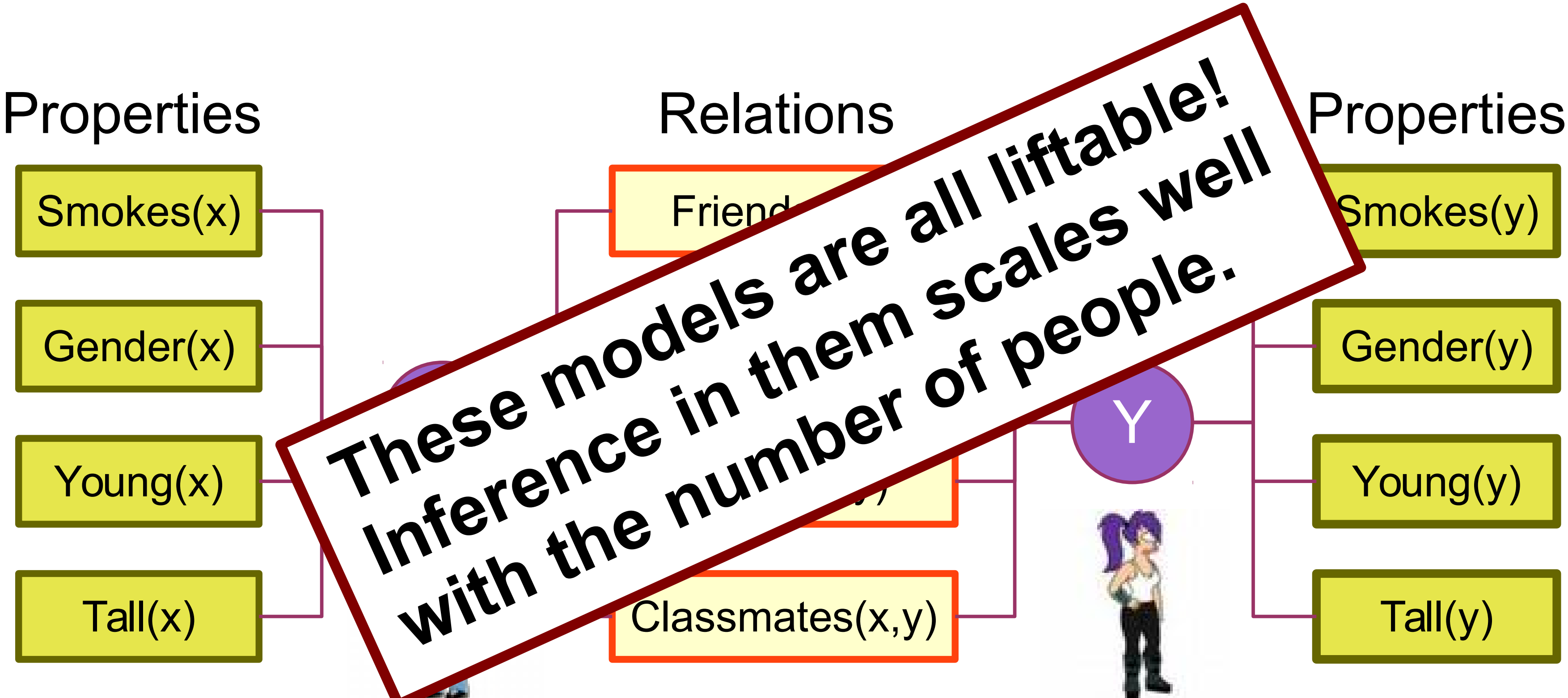


Positive Liftability Result



"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."

Positive Liftability Result



"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."

Complexity in Size of "Evidence"

- Consider a model liftable for model counting:

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

- Given database DB, compute P(Q|DB). Complexity in DB size?
  - Evidence on unary relations: Efficient

    FacultyPage("google.com")=0, CoursePage("coursera.org")=1, …

  - Evidence on binary relations: #P-hard

    Linked("google.com","gmail.com")=1, Linked("google.com","coursera.org")=0

    Intuition: Binary evidence breaks symmetries

  - Evidence on binary relations of Boolean rank < k: Efficient
  - Safe monotone or type-1 CNFs: Any evidence is Efficient

[VdB-AAAI12, Bui-AAAI12, VdB-NIPS13, Dalvi-JACM12, Gribkoff-UAI14]

Overview

1. What are statistical relational models?
2. What is lifted inference?
3. How does lifted inference work?
4. Theoretical insights
5. **Practical applications**

Applications of Lifted Inference

- Many applications of SRL
    - Computational biology
    - Social network analysis
    - Robot mapping
    - Activity recognition
    - Personal assistants
    - Natural language processing
    - Information extraction
    - Entity resolution
    - Link prediction
    - Collective classification
    - Web mining
    - etc.
- Plug in (approximate) lifted inference algorithm
- Notable examples in lifted inference literature
    - Content distribution [Kersting-AAAI10]
    - Groundwater analysis [Choi-UAI12]
    - Video segmentation [Nath-StarAI10]

Lifted Weight Learning

Given: a set of first-order logic **formulas**
a set of training **databases**

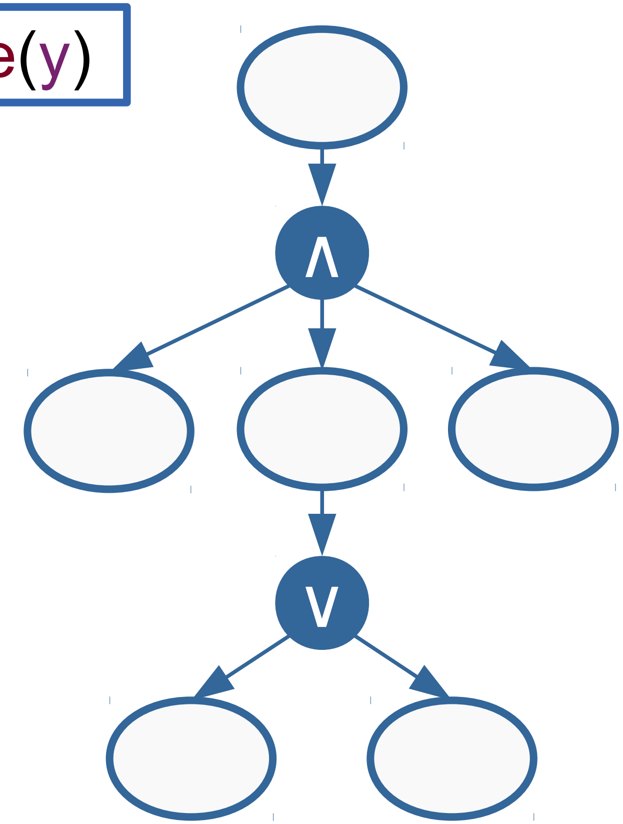
Learn: the associated maximum likelihood **weights**

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

1 Compile formula into circuit

2 Compute maximum likelihood weight W

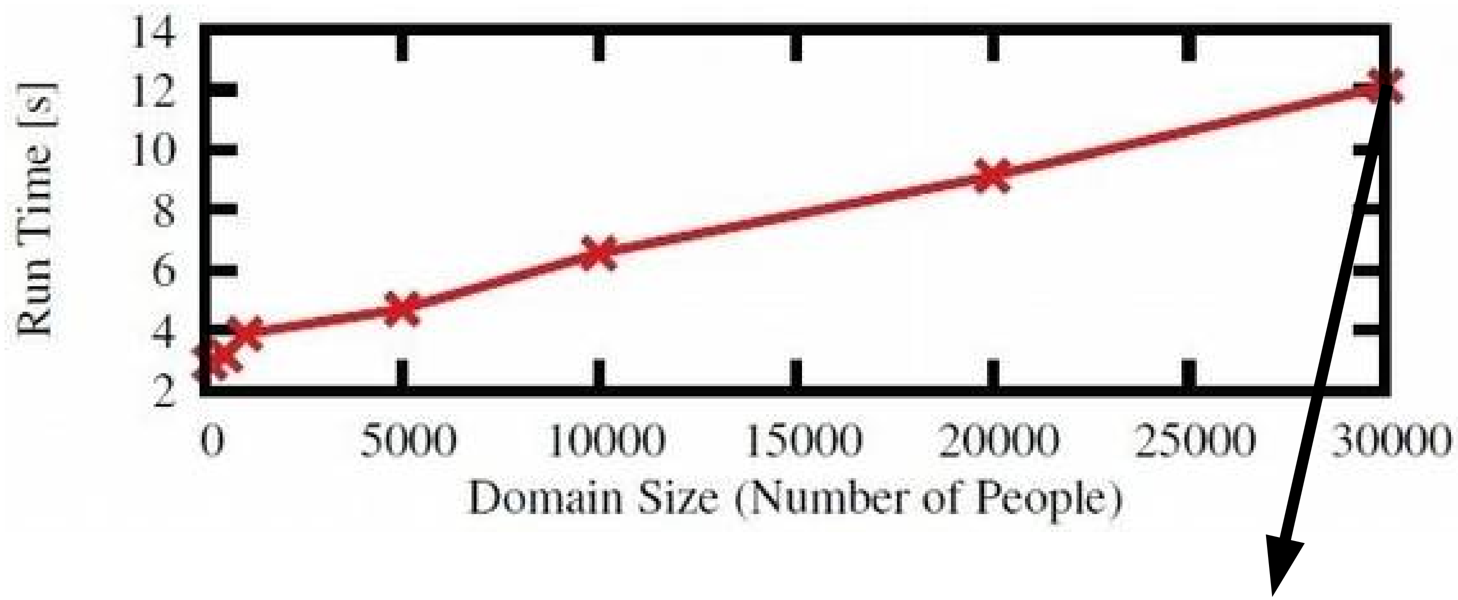
3 Compute exact likelihood of the model



[Jaimovich-UAI07, Ahmadi-ECML12, VdB-StarAI13]

Learning Time - Synthetic

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



Learns a model over 900,030,000 random variables

Lifted Structure Learning

Given: a set of training **databases**

Learn: a set of first-order logic **formulas**
the associated maximum likelihood **weights**

	IMDb			UWCSE		
	B+PLL	B+LWL	LSL	B+PLL	B+LWL	LSL
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**

[Jaimovich-UAI07, VanHaaren-LTPM14]

“But my data has no symmetries?”

1. All statistical relational models have **abundant symmetries**
2. Some **tasks** do not require symmetries in data

   Weight learning, partition functions, single marginals, etc.
3. Symmetries of **computation** are not symmetries of data

   Belief propagation and MAP-LP require weaker automorphisms
4. Over-symmetric **evidence approximation**
   - Approximate Pr(Q|DB) by Pr(Q|DB')
   - DB' has more symmetries than DB, is more liftable
   - Remove weak asymmetries, e.g. Low-rank matrix factorization

   ➔ Very high speed improvements

   ➔ Low approximation error

[Kersting-UAI09, Mladenov-AISTATS14, VdB-NIPS13]

Overview

1. What are statistical relational models?
2. What is lifted inference?
3. How does lifted inference work?
4. Theoretical insights
5. Practical applications

Conclusions

- Lifted inference is frontier of AI, AR, ML and databases

  A radically new reasoning paradigm
- No question that we need
  - relational databases and logic
  - probabilistic models and learning
- Many theoretical open problems – fertile ground
- It works in practice
- Long-term outlook: probabilistic inference exploits
  - ~1988: conditional independence
  - ~2000: contextual independence (local structure)
  - ~201?: symmetries

References

[Richardson-MLJ06]
Richardson, M., & Domingos, P. (2006). Markov logic networks. Machine learning, 62(1-2), 107-136.

[Suciu-Book11]
Suciu, D., Olteanu, D., Ré, C., & Koch, C. (2011). Probabilistic databases. Synthesis Lectures on Data Management, 3(2), 1-180.

[Jha-TCS13]
Jha, A., & Suciu, D. (2013). Knowledge compilation meets database theory: compiling queries to decision diagrams. Theory of Computing Systems, 52(3), 403-440.

[Olteanu-SUM08]
Olteanu, D., & Huang, J. (2008). Using OBDDs for efficient query evaluation on probabilistic databases. In Scalable Uncertainty Management (pp. 326-340). Springer Berlin Heidelberg.

[Gribkoff-UAI14]
Gribkoff, E., Van den Broeck, G., & Suciu, D. (2014). Understanding the Complexity of Lifted Inference and Asymmetric Weighted Model Counting. Proceedings of Uncertainty in AI.

[Gogate-UAI11]
Gogate, V., & Domingos, P. (2012). Probabilistic theorem proving. Proceedings of Uncertainty in AI.

[VdB-IJCAI11]
Van den Broeck, G., Taghipour, N., Meert, W., Davis, J., & De Raedt, L. (2011, July). 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.

References

[Niepert-AAAI14]
Niepert, M., & Van den Broeck, G. (2014). Tractability through exchangeability: A new perspective on efficient probabilistic inference. Proceedings of AAAI.

[VdB-NIPS11]
Van den Broeck, G. (2011). On the completeness of first-order knowledge compilation for lifted probabilistic inference. In Advances in Neural Information Processing Systems (pp. 1386-1394).

[Jaeger-StarAI12]
Jaeger, M., & Van den Broeck, G. (2012, August). Liftability of probabilistic inference: Upper and lower bounds. In Proceedings of the 2nd International Workshop on Statistical Relational AI.

[Poole-IJCAI03]
Poole, D. (2003, August). First-order probabilistic inference. In IJCAI (Vol. 3, pp. 985-991).

[Braz-IJCAI05]
Braz, R., Amir, E., & Roth, D. (2005, July). Lifted first-order probabilistic inference. In Proceedings of the 19th international joint conference on Artificial intelligence (pp. 1319-1325).

[Milch-AAAI08]
Milch, B., Zettlemoyer, L. S., Kersting, K., Haimes, M., & Kaelbling, L. P. (2008, July). Lifted Probabilistic Inference with Counting Formulas. In AAAI (Vol. 8, pp. 1062-1068).

[Taghipour-JAIR13]
Taghipour, N., Fierens, D., Davis, J., & Blockeel, H. (2014). Lifted variable elimination: Decoupling the operators from the constraint language. JAIR

References

[VdB-AAAI12]
Van den Broeck, G., & Davis, J. (2012, July). Conditioning in First-Order Knowledge Compilation and Lifted Probabilistic Inference. In AAAI.

[VdB-Thesis13]
Van den Broeck, G. (2013). Lifted Inference and Learning in Statistical Relational Models (Doctoral dissertation, Ph. D. Dissertation, KU Leuven).

[Jaimovich-UAI07]
Jaimovich, A., Meshi, O., & Friedman, N. (2007). Template based inference in symmetric relational Markov random fields. Proceedings of Uncertainty in AI

[Singla-AAAI08]
Singla, P., & Domingos, P. (2008, July). Lifted First-Order Belief Propagation. In AAAI (Vol. 8, pp. 1094-1099).

[Kersting-UAI09]
Kersting, K., Ahmadi, B., & Natarajan, S. (2009, June). Counting belief propagation. In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (pp. 277-284). AUAI Press.

[Sen-VLDB08]
Sen, P., Deshpande, A., & Getoor, L. (2008). Exploiting shared correlations in probabilistic databases. Proceedings of the VLDB Endowment, 1(1), 809-820.

[Sen-UAI09]
Sen, P., Deshpande, A., & Getoor, L. (2009, June). Bisimulation-based approximate lifted inference. In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (pp. 496-505). AUAI Press.

References

[Gogate-AAAI12]
Gogate, V., Jha, A. K., & Venugopal, D. (2012, July). Advances in Lifted Importance Sampling. In AAAI.

[VdB-UAI12]
Van den Broeck, G., Choi, A., & Darwiche, A. (2012). Lifted relax, compensate and then recover: From approximate to exact lifted probabilistic inference. Proceedings of Uncertainty in AI

[Niepert-UAI12]
Niepert, M. (2012). Markov chains on orbits of permutation groups. Proceedings of Uncertainty in AI

[Niepert-AAAI13]
Niepert, M. (2013). Symmetry-Aware Marginal Density Estimation. Proceedings of AAAI.

[Venugopal-NIPS12]
Venugopal, D., & Gogate, V. (2012). On lifting the gibbs sampling algorithm. In Advances in Neural Information Processing Systems (pp. 1655-1663).

[Bui-StarAI12]
Bui, H. H., Huynh, T. N., & Riedel, S. (2012). Automorphism groups of graphical models and lifted variational inference. StarAI

[Choi-UAI12]
Choi, J., & Amir, E. (2012). Lifted relational variational inference. Proceedings of Uncertainty in AI

[Mladenov-AISTATS14]
Mladenov, M., Kersting, K., & Globerson, A. (2014). Efficient Lifting of MAP LP Relaxations Using k-Locality. In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics (pp. 623-632).

References

[Apsel-AAAI14]
Apsel, U., Kersting, K., & Mladenov, M. (2014). Lifting Relational MAP-LPs using Cluster Signatures. Proceedings of AAAI

[Ahmadi-IJCAI11]
Ahmadi, B., Kersting, K., & Sanner, S. (2011, July). Multi-evidence lifted message passing, with application to pagerank and the kalman filter. In IJCAI Proceedings-International Joint Conference on Artificial Intelligence (Vol. 22, No. 1, p. 1152).

[Choi-IJCAI11]
Choi, J., Guzman-Rivera, A., & Amir, E. (2011, June). Lifted Relational Kalman Filtering. In IJCAI (pp. 2092-2099).

[Mladenov-AISTATS12]
Mladenov, M., Ahmadi, B., & Kersting, K. (2012). Lifted linear programming. In International Conference on Artificial Intelligence and Statistics (pp. 788-797).

[VdB-KR14]
Van den Broeck, G., Meert, W., & Darwiche, A. (2013). Skolemization for weighted first-order model counting. Proceedings of KR.

[Dalvi-JACM12]
Dalvi, N., & Suciu, D. (2012). The dichotomy of probabilistic inference for unions of conjunctive queries. Journal of the ACM (JACM), 59(6), 30.

[Jaeger-TPLP12]
Jaeger, M. (2012). Lower complexity bounds for lifted inference. Theory and Practice of Logic Programming

References

[Bui-AAAI12]
Bui, H. B., Huynh, T. N., & de Salvo Braz, R. (2012). Exact Lifted Inference with Distinct Soft Evidence on Every Object. Proceedings of AAAI.

[VdB-NIPS13]
Van den Broeck, G., & Darwiche, A. (2013). On the complexity and approximation of binary evidence in lifted inference. In Advances in Neural Information Processing Systems (pp. 2868-2876).

[Kersting-AAAI10]
Kersting, K., El Massaoudi, Y., Hadiji, F., & Ahmadi, B. (2010). Informed Lifting for Message-Passing. Proceedings of AAAI.

[Nath-StarAI10]
Nath, A., & Domingos, P. (2010). Efficient Lifting for Online Probabilistic Inference. In Statistical Relational Artificial Intelligence.

[Ahmadi-ECML12]
Ahmadi, B., Kersting, K., & Natarajan, S. (2012). Lifted online training of relational models with stochastic gradient methods. In Machine Learning and Knowledge Discovery in Databases (pp. 585-600). Springer Berlin Heidelberg.

[VdB-StarAI13]
Van den Broeck, G., Meert, W., & Davis, J. (2013). Lifted Generative Parameter Learning. In AAAI Workshop: Statistical Relational Artificial Intelligence.

[VanHaaren-LTPM14]
Van Haaren, J., Van den Broeck, G., Meert, W., & Davis, J. (2014). Tractable Learning of Liftable Markob Logic Networks. In Learning Tractable Probabilistic Models.

Thanks!