Lifted Probabilistic Inference in Relational Models

Guy Van den Broeck
UCLA

Dan Suciu
U. of Washington

IJCAI Tutorial

July 10, 2016

About the Tutorial

Slides available at
http://web.cs.ucla.edu/~guyvdb/talks/IJCAI16-tutorial/
Extensive bibliography at the end.
Your speakers:

http://web.cs.ucla.edu/~guyvdb/

https://homes.cs.washington.edu/~suciu/

I work in AI

I work in DB

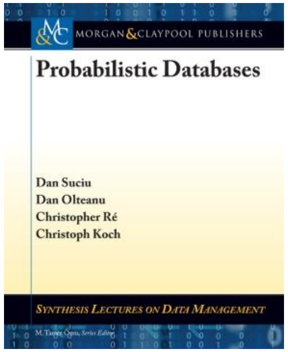
About the Tutorial

- The tutorial is about
  - deep connections between AI and DBs
  - a unified view on probabilistic reasoning
  - a logical approach to prob. reasoning

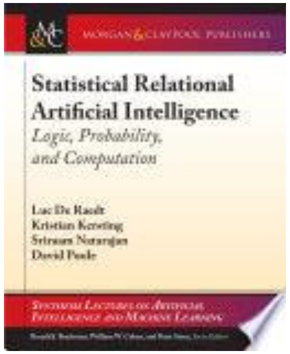
- The tutorial is NOT an exhaustive overview of lifted algorithms for graphical models (see references at the end)

If you want more…

- Books
  - Probabilistic Databases
  - Statistical Relational AI
  - (Lifted Inference Book)



[Suciu'11]



[DeRaedt'16]

- StarAI workshop on Monday
  http://www.starai.org

- Main conference papers

Outline

- Part 1: Motivation
- Part 2: Probabilistic Databases
- Part 3: Weighted Model Counting
- Part 4: Lifted Inference for WFOMC

- Part 5: Completeness of Lifted Inference
- Part 6: Query Compilation
- Part 7: Symmetric Lifted Inference Complexity
- Part 8: Open-World Probabilistic Databases
- Part 9: Discussion & Conclusions

Outline

- Part 1: Motivation
- Part 2: Probabilistic Databases
- Part 3: Weighted Model Counting
- Part 4: Lifted Inference for WFOMC
- Part 5: Completeness of Lifted Inference
- Part 6: Query Compilation
- Part 7: Symmetric Lifted Inference Complexity
- Part 8: Open-World Probabilistic Databases
- Part 9: Discussion & Conclusions

Part 1: Motivation

- Why do we need relational representations of uncertainty?
- Why do we need probabilistic queries?
- Why do we need lifted inference algorithms?

Why Relational Data?

- Our data is already relational!
  - Companies run relational databases
  - Scientific data is relational:
    - Large Hadron Collider generated 25PB in 2012
    - LSST Telescope will produce 30TB per night

- Big data is big business:
  - Oracle: $7.1BN in sales
  - IBM: $3.2BN in sales
  - Microsoft: $2.6BN in sales

≈ GDP of Iceland!

[Gartner'06]

Why Probabilistic Relational Data?

- Relational data is increasingly probabilistic
  - NELL machine reading (>50M tuples)
  - Google Knowledge Vault (>2BN tuples)
  - DeepDive (>7M tuples)

- Data is inferred from unstructured information using statistical models
  - Learned from the web, large text corpora, ontologies, etc.
  - The learned/extracted data is relational

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

Information Extraction

PhD Students Luc De Raedt

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

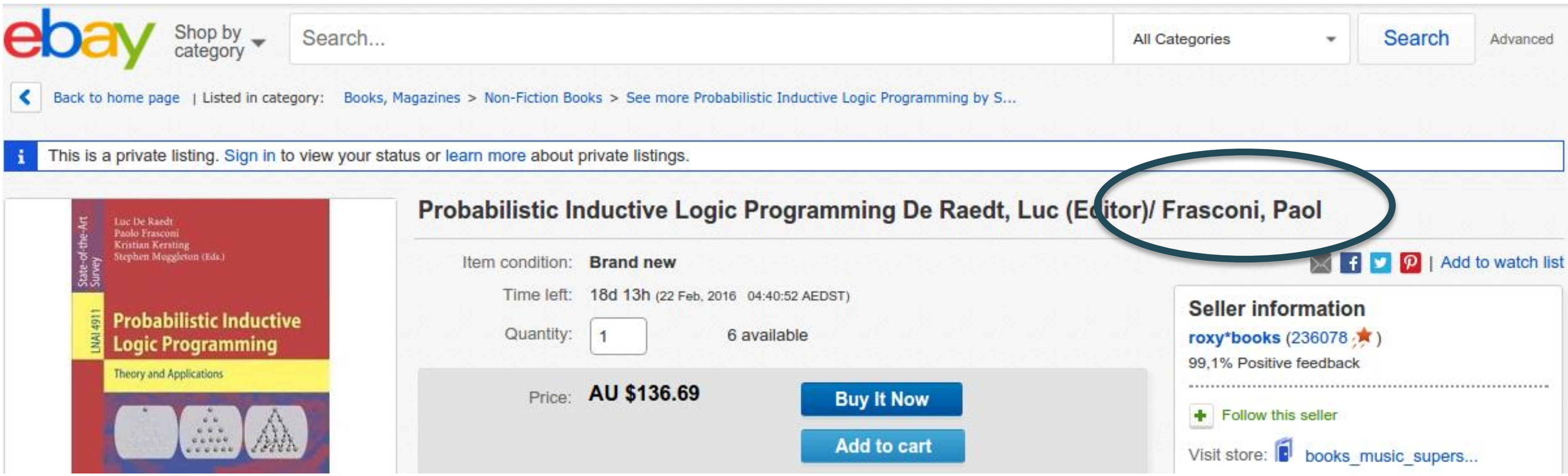
Alumni Luc De Raedt

1. Hendrik Blockeel, *Top-down induction of first order logical decision trees*, Ph.D. thesis, Department of Computer Science, K.U.Leuven, Leuven, Belgium, december 1998, 202+xv pages. (Co-promotor Maurice Bruynooghe)
2. Luc Dehaspe, *Frequent pattern discovery in first-order logic*, Ph.D. thesis, Department of Computer

PublishedWith

X	Y	P
Luc	Laura	0.7
Luc	Hendrik	0.6
Luc	Kathleen	0.3
Luc	Paol	0.3
Luc	Paolo	0.1

Extraction is so Noisy!



Representation: Probabilistic Databases

- Tuple-independent probabilistic databases

Actor

Name	Prob
Brando	0.9
Cruise	0.8
Coppola	0.1

WorkedFor

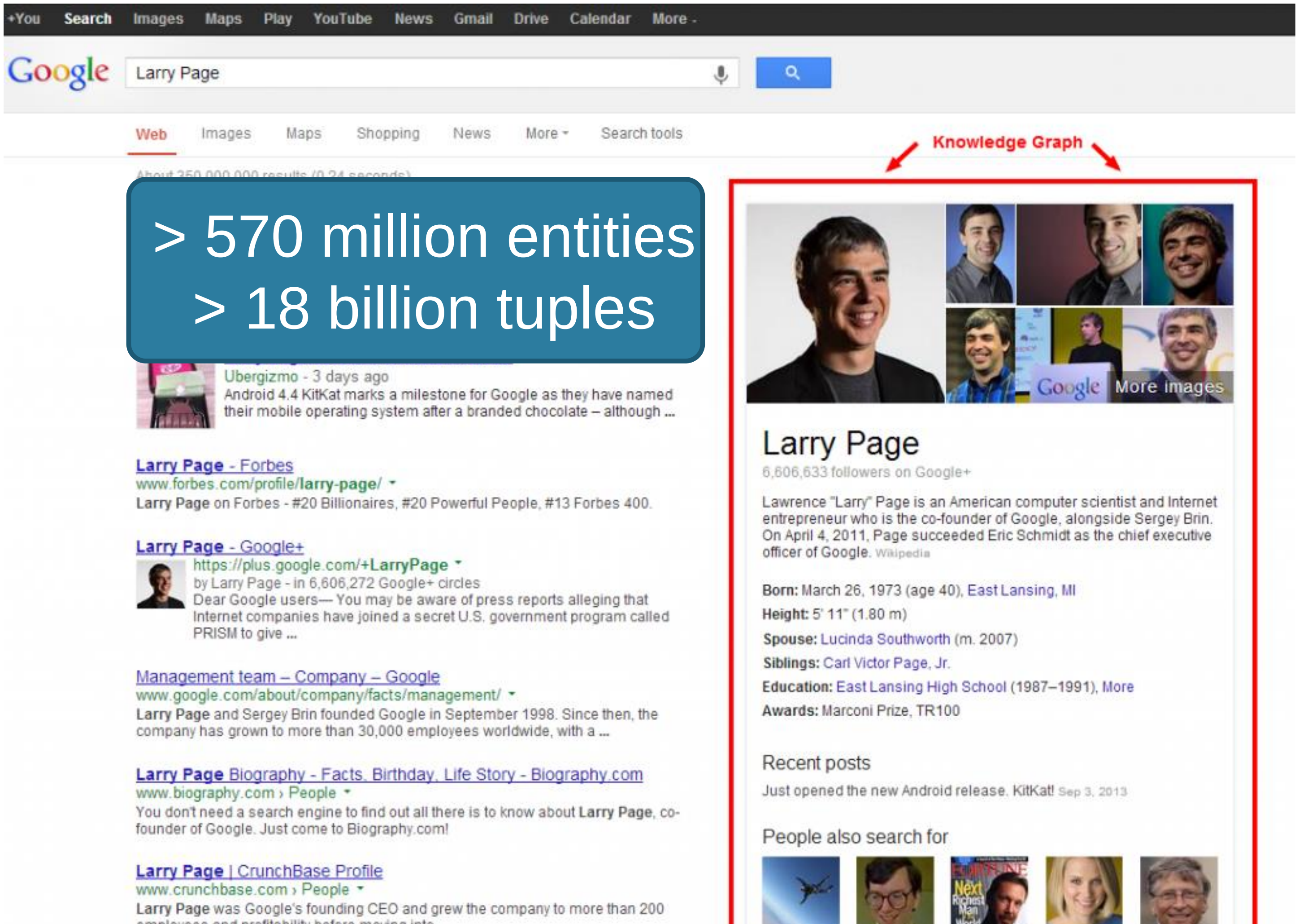
Actor	Director	Prob
Brando	Coppola	0.9
Coppola	Brando	0.2
Cruise	Coppola	0.1

- Query: SQL or First-order logic

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

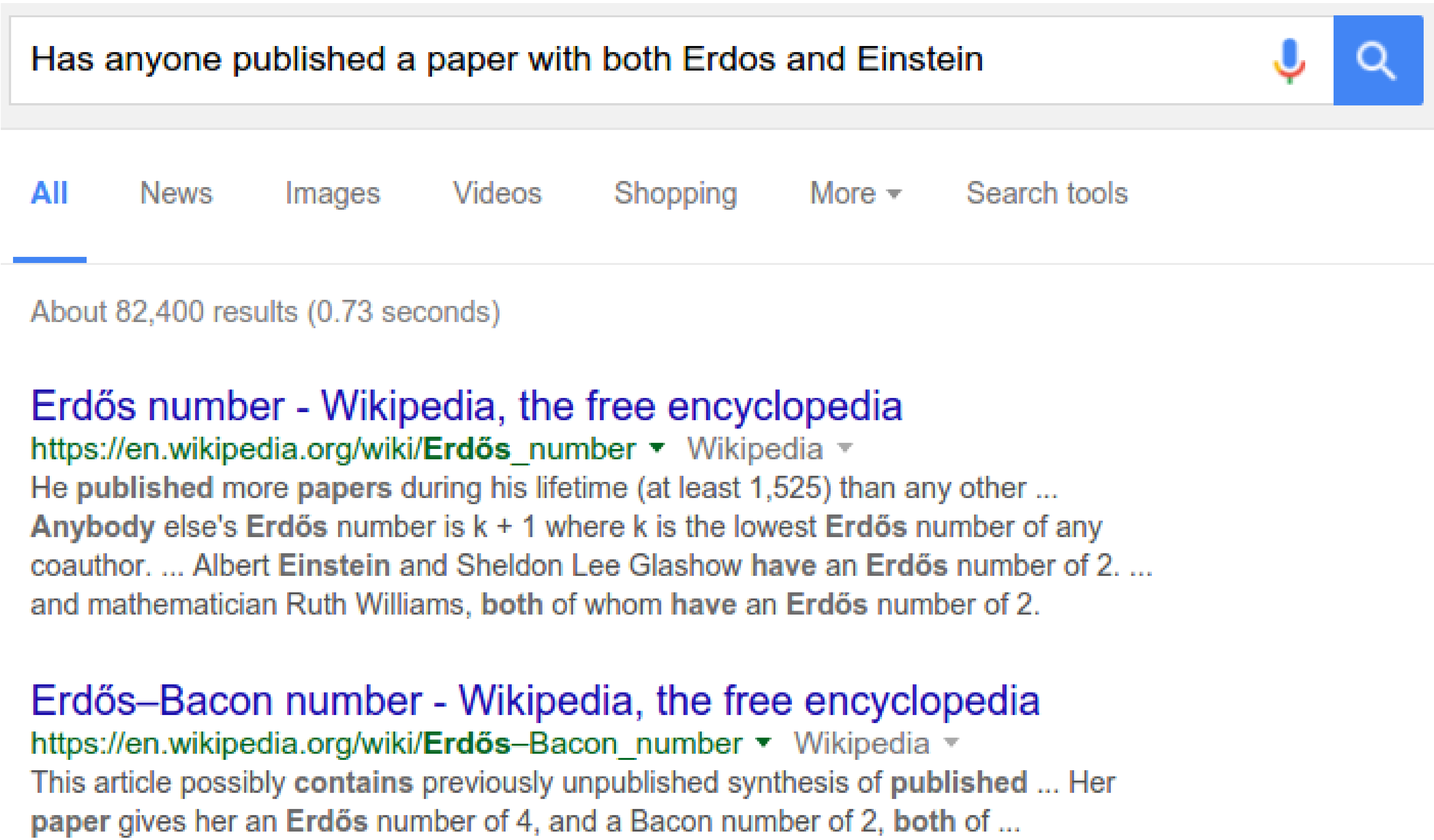
$Q(x) = \exists y\ Actor(x) \wedge WorkedFor(x,y)$

Why Probabilistic Queries?



> 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 ...

Erdős is in the Knowledge Graph

Paul Erdos

All   Images   Videos   Books   News   More   Search tools

About 333,000 results (0.35 seconds)

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

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

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

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

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

Paul Erdős

Mathematician

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

Born: March 26, 1913, Budapest, Hungary

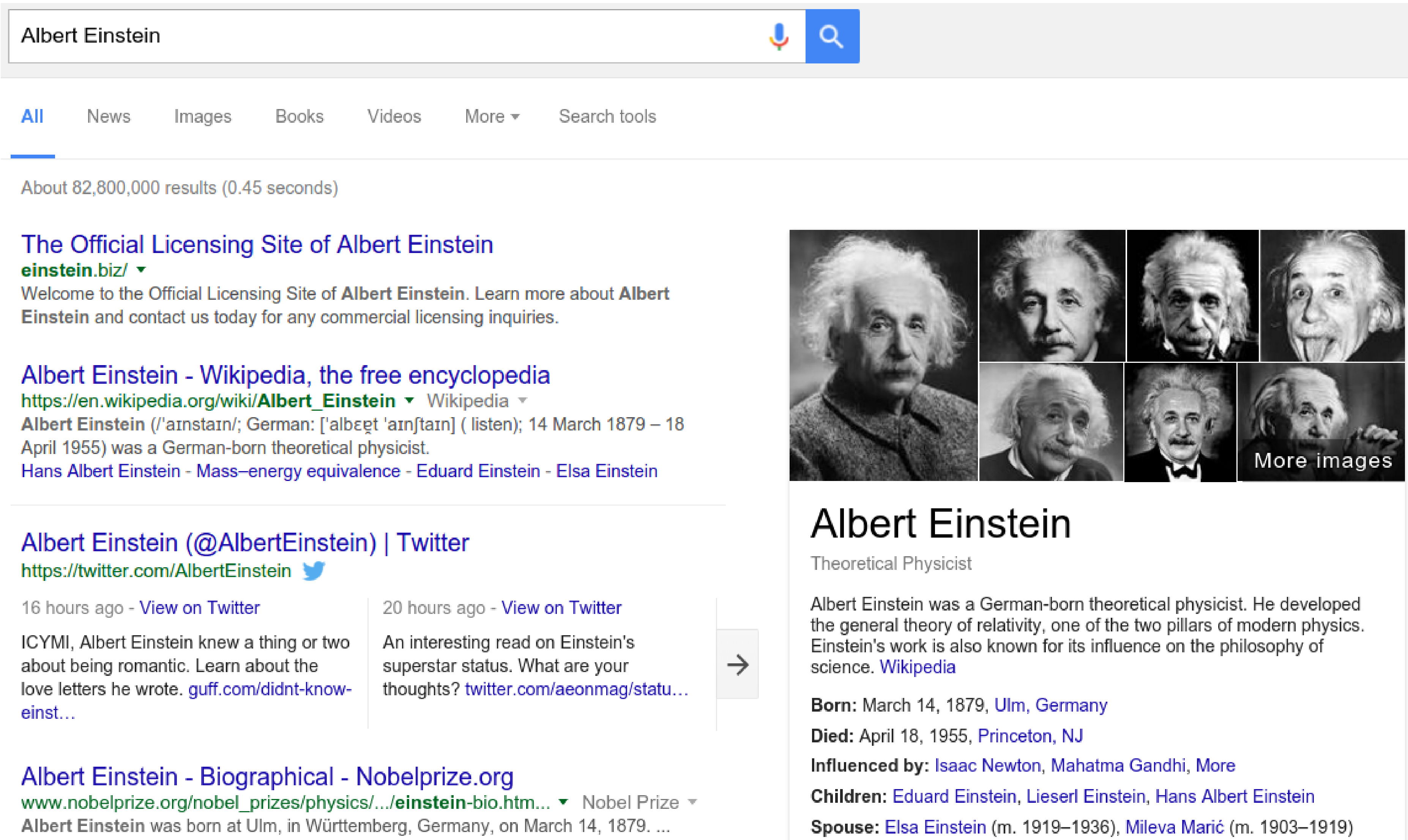
Died: September 20, 1996, Warsaw, Poland

Education: Eötvös Loránd University (1934)

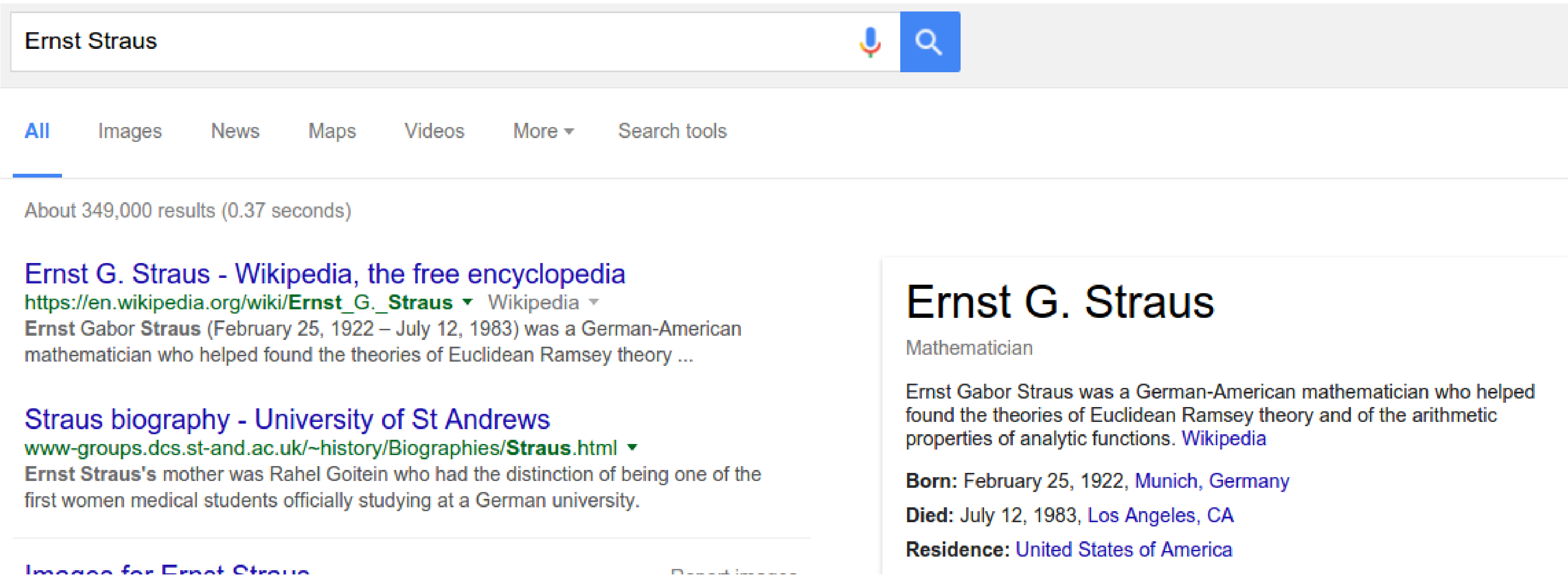
Books: Probabilistic Methods in Combinatorics, More

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

Einstein is in the Knowledge Graph

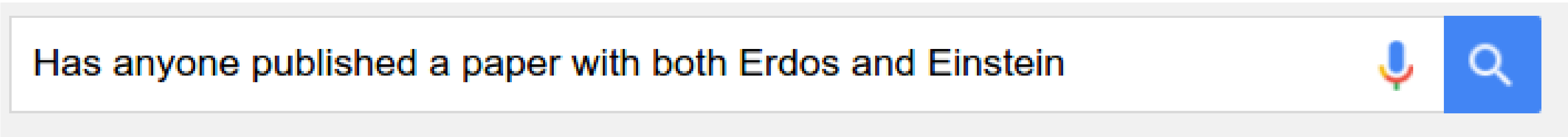


This guy is in the Knowledge Graph



… and he published with both Einstein and Erdos!

Desired Query Answer



★★★★★

Ernst Straus

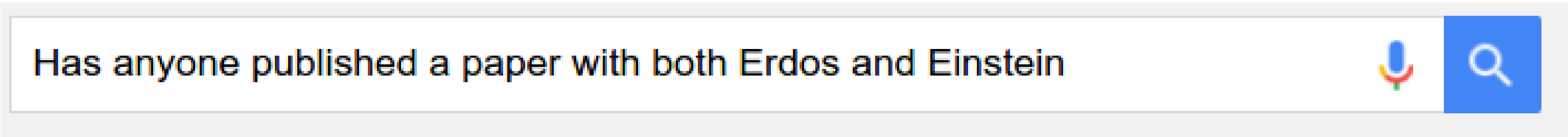
★★☆☆☆

Kristian Kersting, …

★☆☆☆☆

Justin Bieber, …

Observations



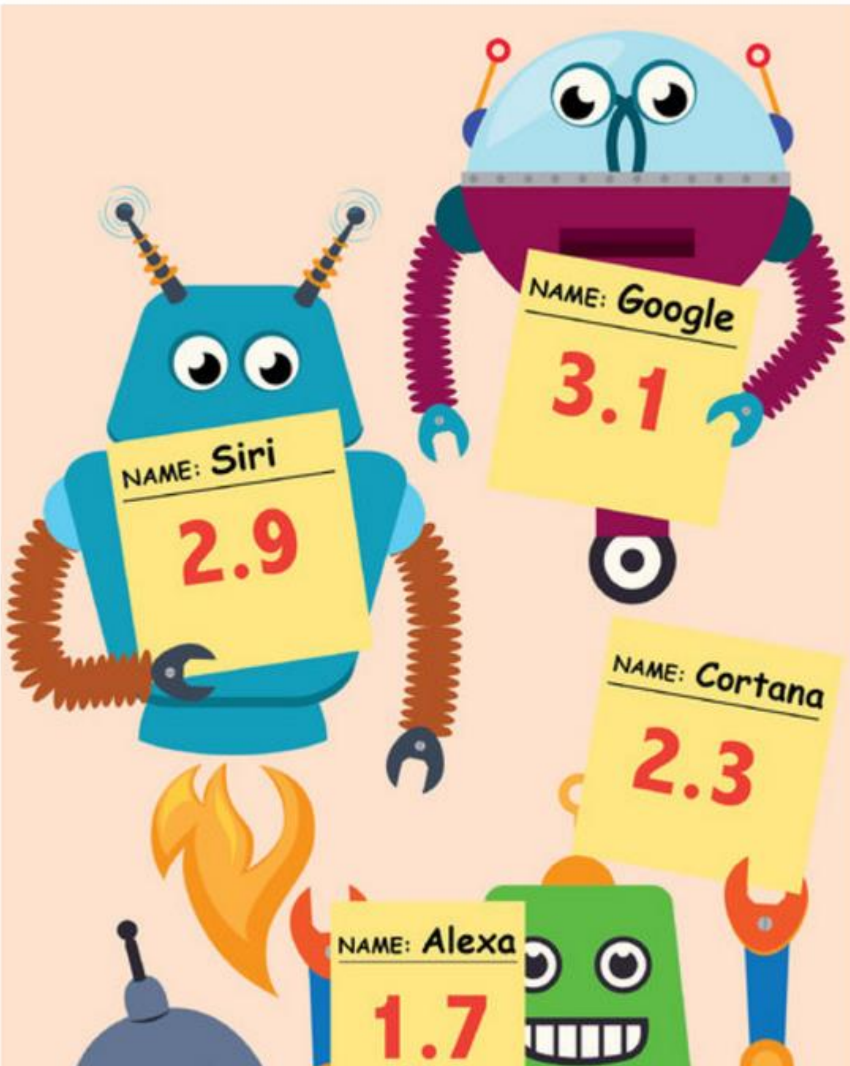
- Cannot come from labeled data
- Fuse uncertain information from many pages

- Expose uncertainty in query answers
  … and risk incorrect answers
- Embrace probability!

Siri, Alexa and Other Virtual Assistants Put to the Test

Tech Fix

By **BRIAN X. CHEN** JAN. 27, 2016



WHEN I asked Alexa earlier this week who was playing in the Super Bowl, she responded, somewhat monotonously, "Super Bowl 49's winner is New England Patriots."

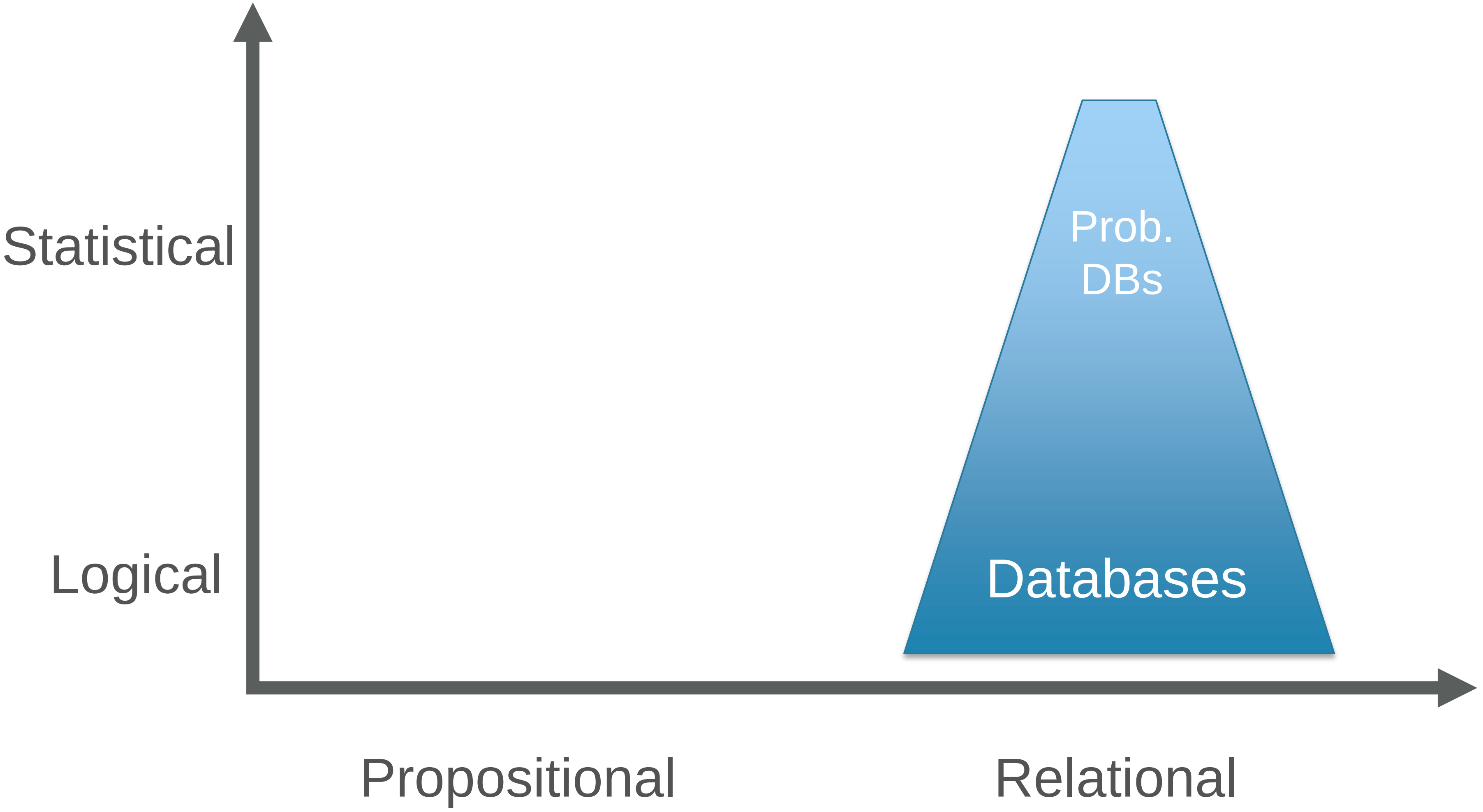
"Come on, that's last year's Super Bowl," I said. "Even I can do better than that."

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

All the so-called Frightful Five tech

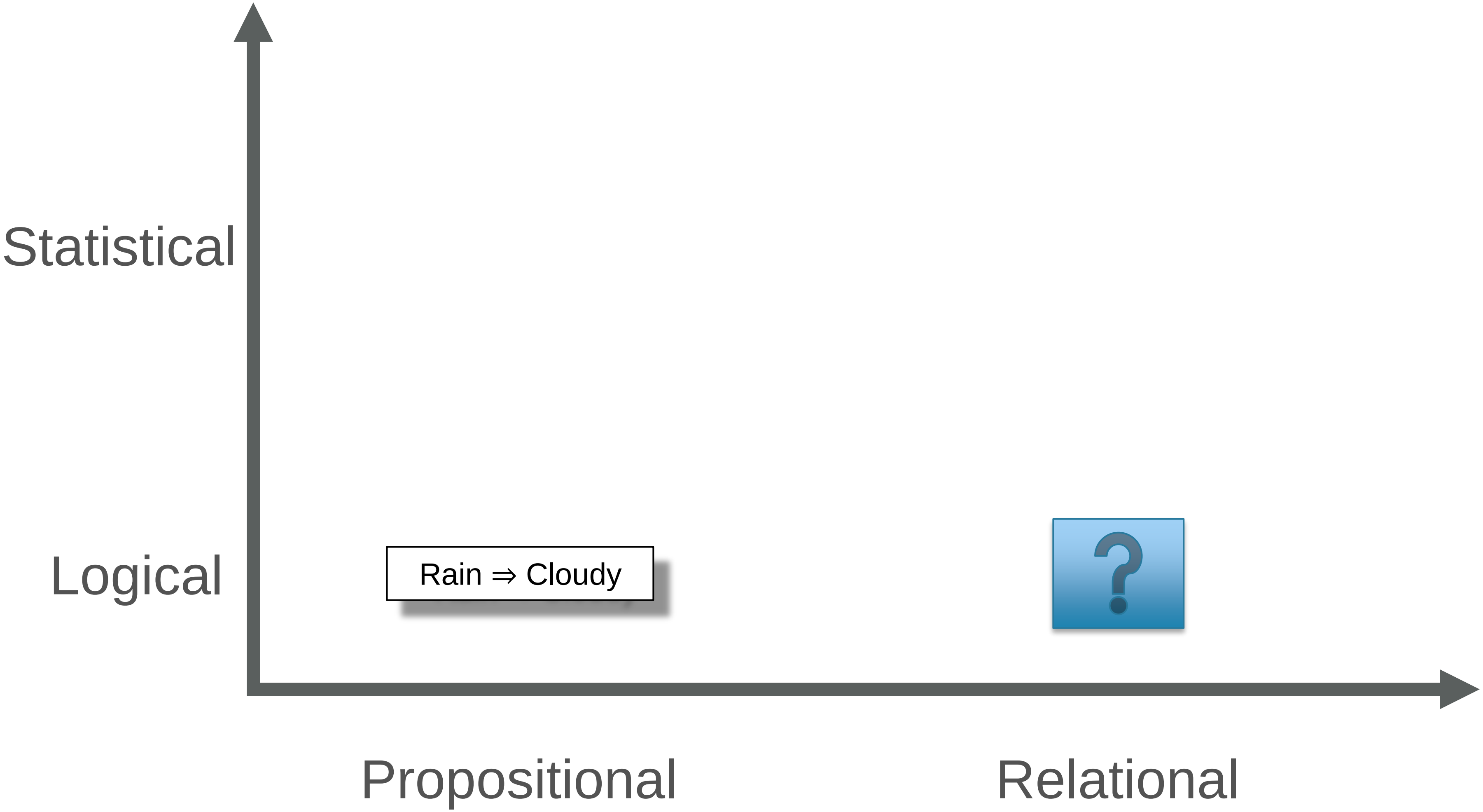
[Chen'16]
(NYTimes)

Summary



Statistical
Logical
Prob.
DBs
Databases
Propositional
Relational

Representations in AI and ML



Graphical Model Learning

Medical Records

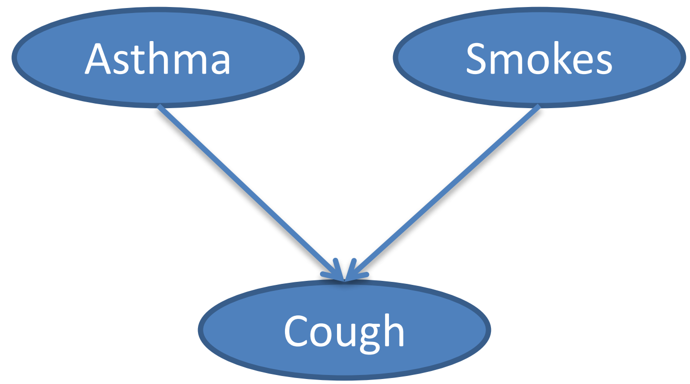
Bayesian Network

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

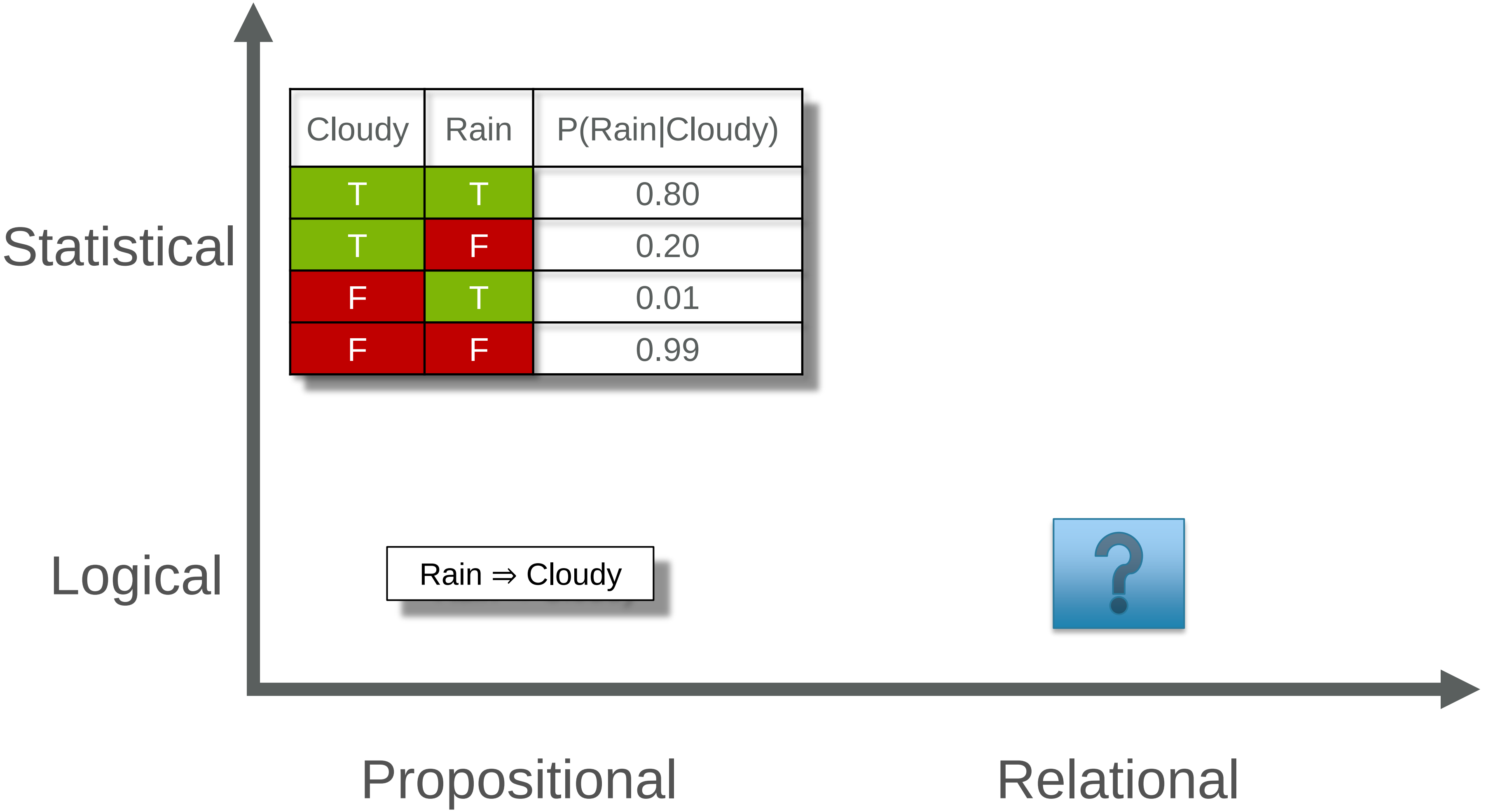
Frank	1	?	?

Big data

Frank	1	0.3	0.2



Representations in AI and ML



Cloudy	Rain	P(Rain|Cloudy)
T	T	0.80
T	F	0.20
F	T	0.01
F	F	0.99

Relational Representations

- Example: First-Order Logic

$$\overbrace{\forall x,y,\ \underbrace{Smokes(x)}_{\text{Atom}} \wedge Friends(x,\underset{\substack{\uparrow\\ \text{Logical Variable}}}{y}) \Rightarrow Smokes(y)}^{\text{Formula}}$$

- Logical variables have domain of constants

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

- Ground formula has no logical variables

  Smokes(Alice) ∧ Friends(Alice,Bob) ⇒ Smokes(Bob)

Representations in AI and ML

Statistical

Cloudy	Rain	P(Rain|Cloudy)
T	T	0.80
T	F	0.20
F	T	0.01
F	F	0.99

Logical

Rain ⇒ Cloudy

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

Propositional

Relational

Why Statistical Relational Models?

- Probabilistic graphical models
  - ✔ Quantify uncertainty and noise
  - ✖ Not very expressive
    Rules of chess in ~100,000 pages
- First-order logic
  - ✔ Very expressive
    Rules of chess in 1 page
  - ✔ Good match for abundant relational data
  - ✖ Hard to express uncertainty and noise

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

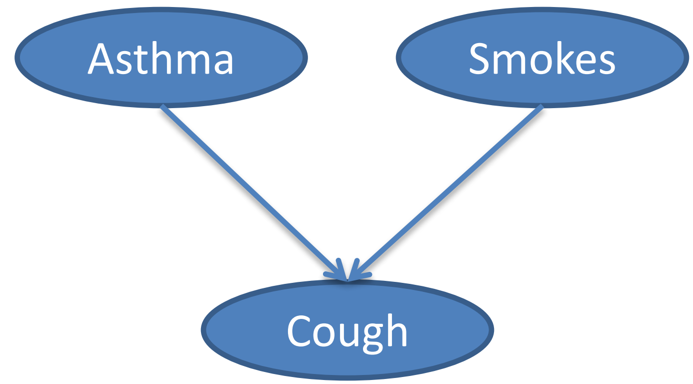
Frank	1	?	?

Friends

Brothers

Frank	1	0.3	0.2

Frank	1	0.2	0.6

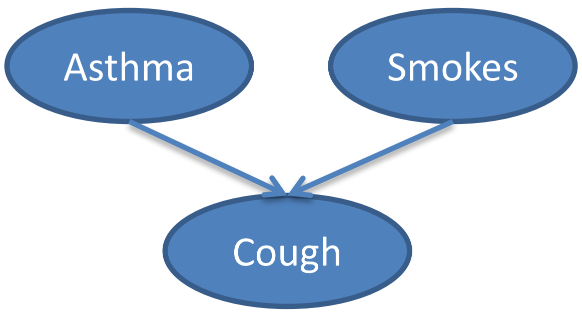


Rows are **independent** during learning and inference!

Statistical Relational Representations

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

Logical variables refer to entities

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

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

Classical Machine Learning

Purchases

Model

Name	Age	Product	Price
Dave	40	Android	€249
Alice	35	iPhone	€799
Bob	32	iPhone	€799
Charlie	22	iPhone	€699
Eve	17	Android	€299
Frank	15	Android	€199

People **older** than **27** probably buy **iPhone**.

People **younger** than **27** probably buy **Android**.

Inference: *Does Guy buy an iPhone?*
Answer: Yes, with probability 66%

Statistical Relational Learning

Purchases

Name	Age	Product	Price
Dave	40	Android	€249
Alice	35	iPhone	€799
Bob	32	iPhone	€799
Charlie	22	iPhone	€699
Eve	17	Android	€299
Frank	15	Android	€199

Relationships

Person A	Person B	Type	
Alice	Bob	Spouse	
Alice	Charlie	Mother	Family 1
Bob	Charlie	Father	
Dave	Eve	Father	
Dave	Frank	Father	Family 2
Eve	Frank	Siblings	

Model

Family members probably buy the **same type** of phone.

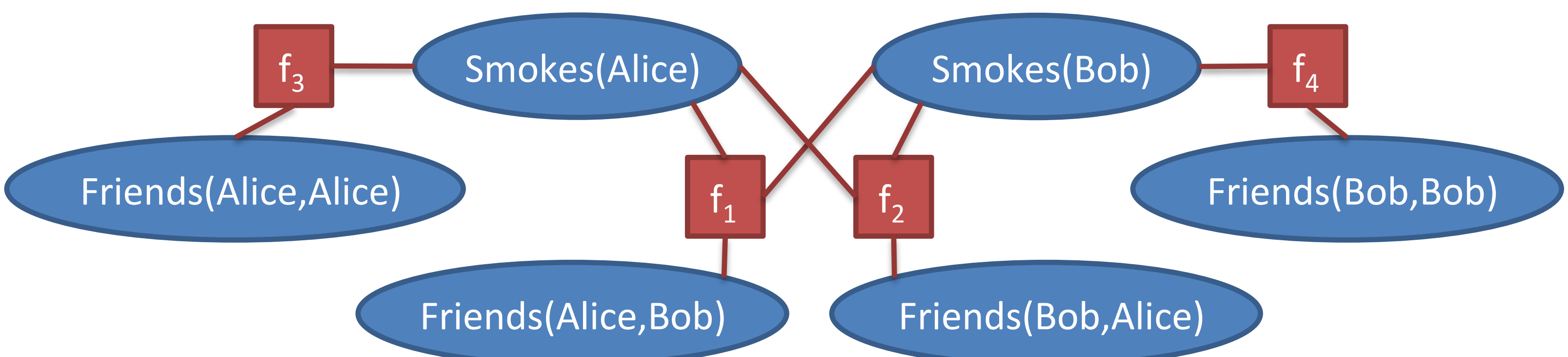
Example: Markov Logic

- Weighted First-Order Logic

Weight or Probability

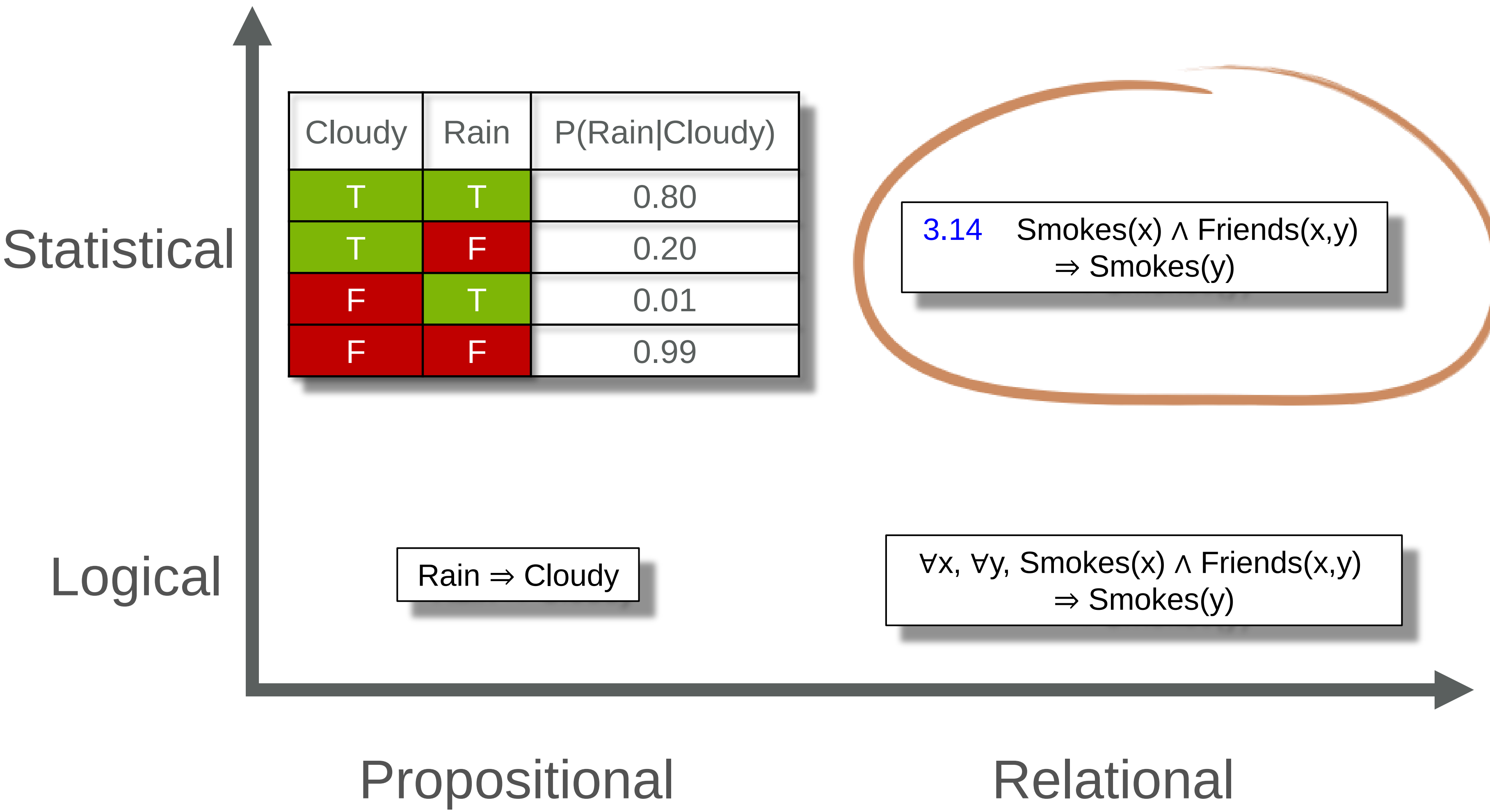
$$3.14 \quad \overbrace{\text{Smokes}(x) \wedge \text{Friends}(x,y) \Rightarrow \text{Smokes}(y)}^{\text{FOL Formula}}$$

- Ground atom/tuple = random variable in {true,false}
  e.g., Smokes(Alice), Friends(Alice,Bob), etc.
- Ground formula = factor in propositional factor graph



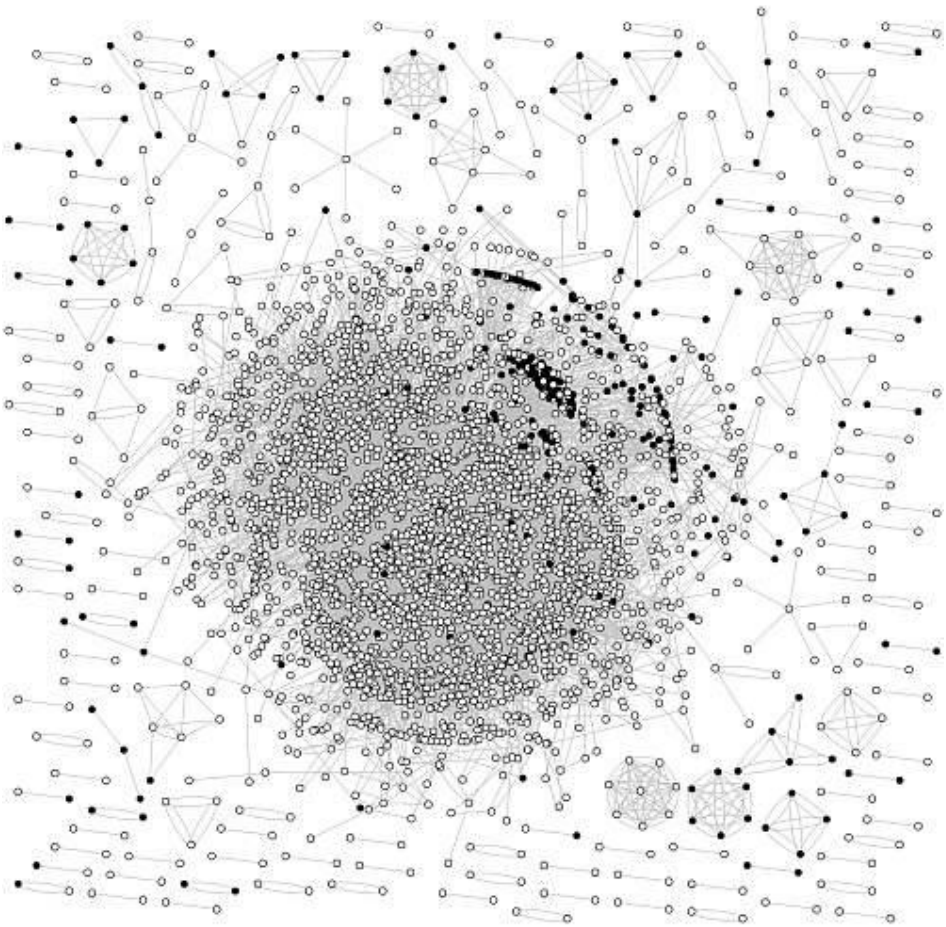
[Richardson'06]

Representations in AI and ML



Cloudy	Rain	P(Rain|Cloudy)
T	T	0.80
T	F	0.20
F	T	0.01
F	F	0.99

Collective Classification

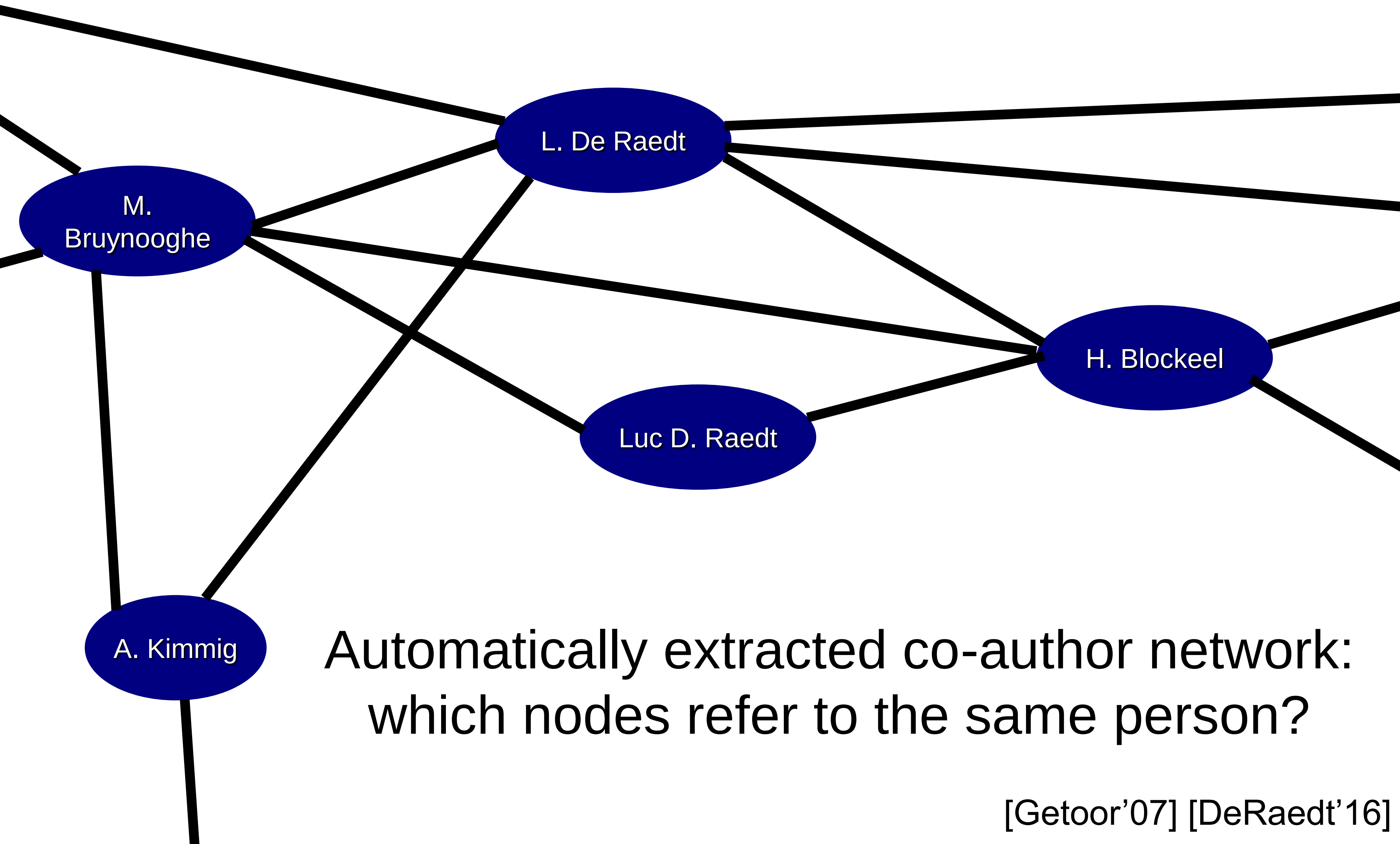


Can we predict the type of the nodes given information on its links and attributes?

E.g., the type of a webpage given its links and the words on the page?

[Getoor'07] [DeRaedt'16]

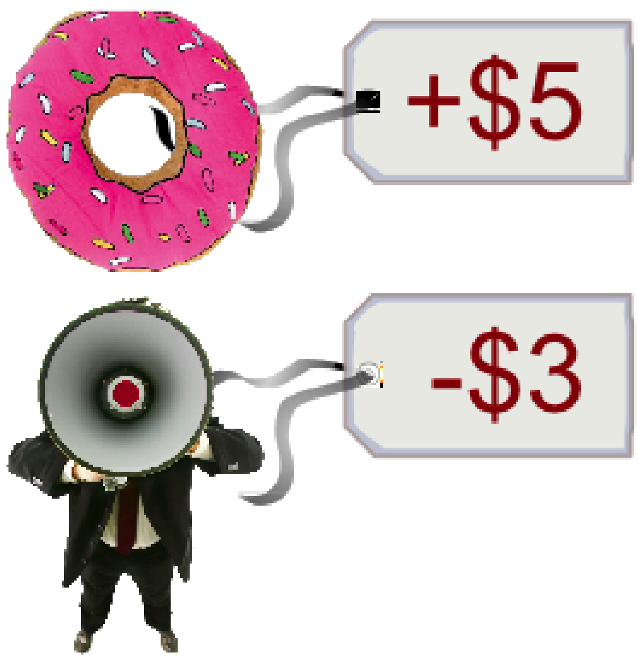
Entity Resolution



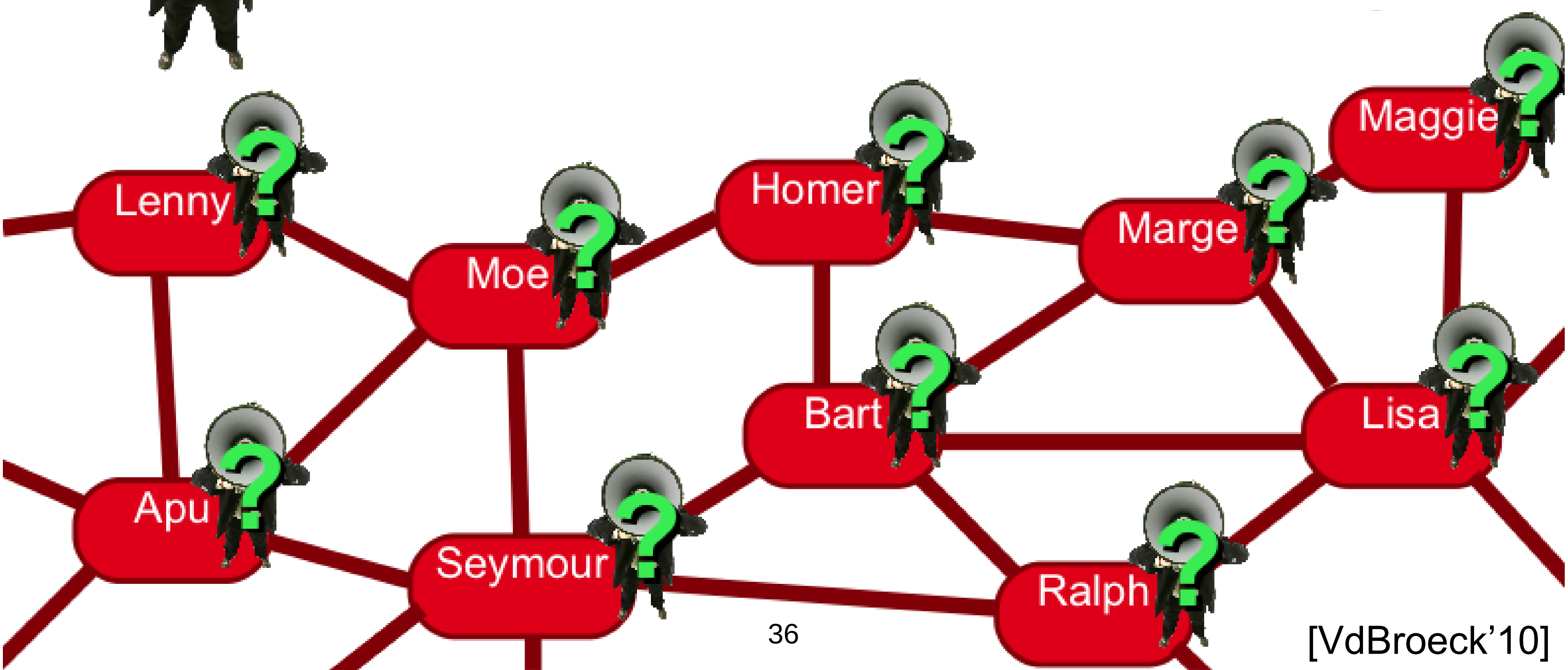
Automatically extracted co-author network:
which nodes refer to the same person?

[Getoor'07] [DeRaedt'16]

Viral Marketing

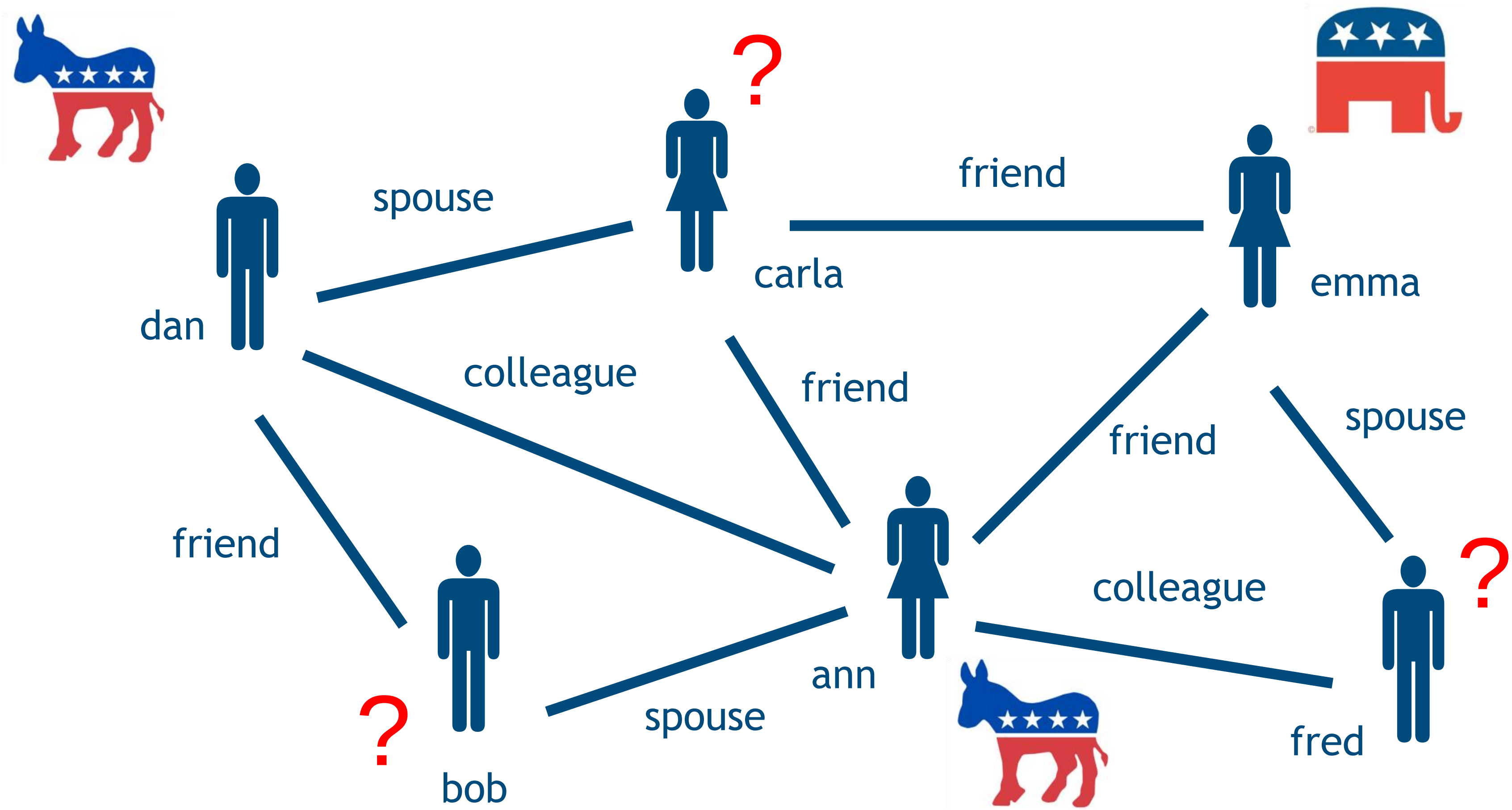


Which advertising strategy maximizes expected profit?



[VdBroeck'10]

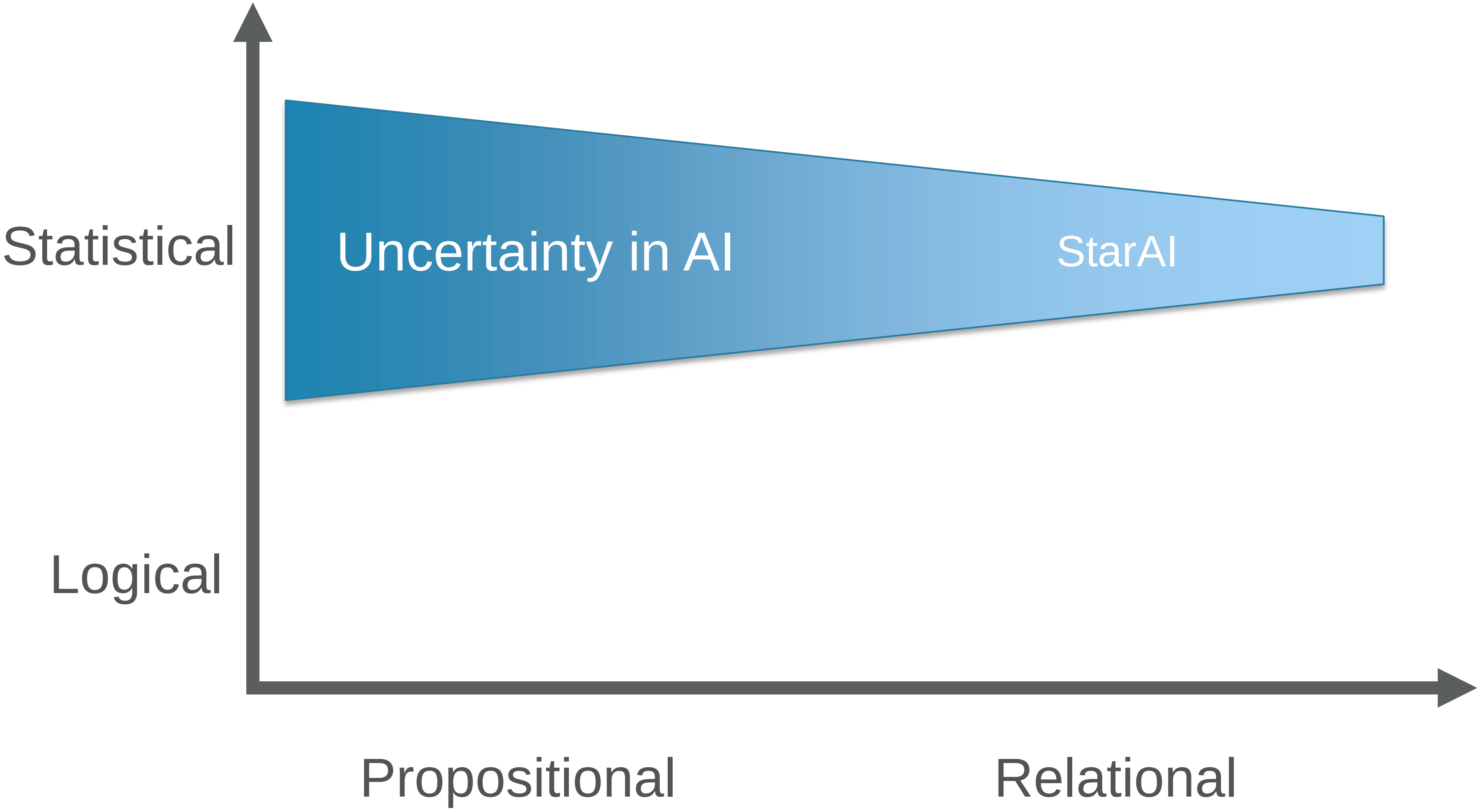
Voter Opinion Modeling



Can we predict preferences?

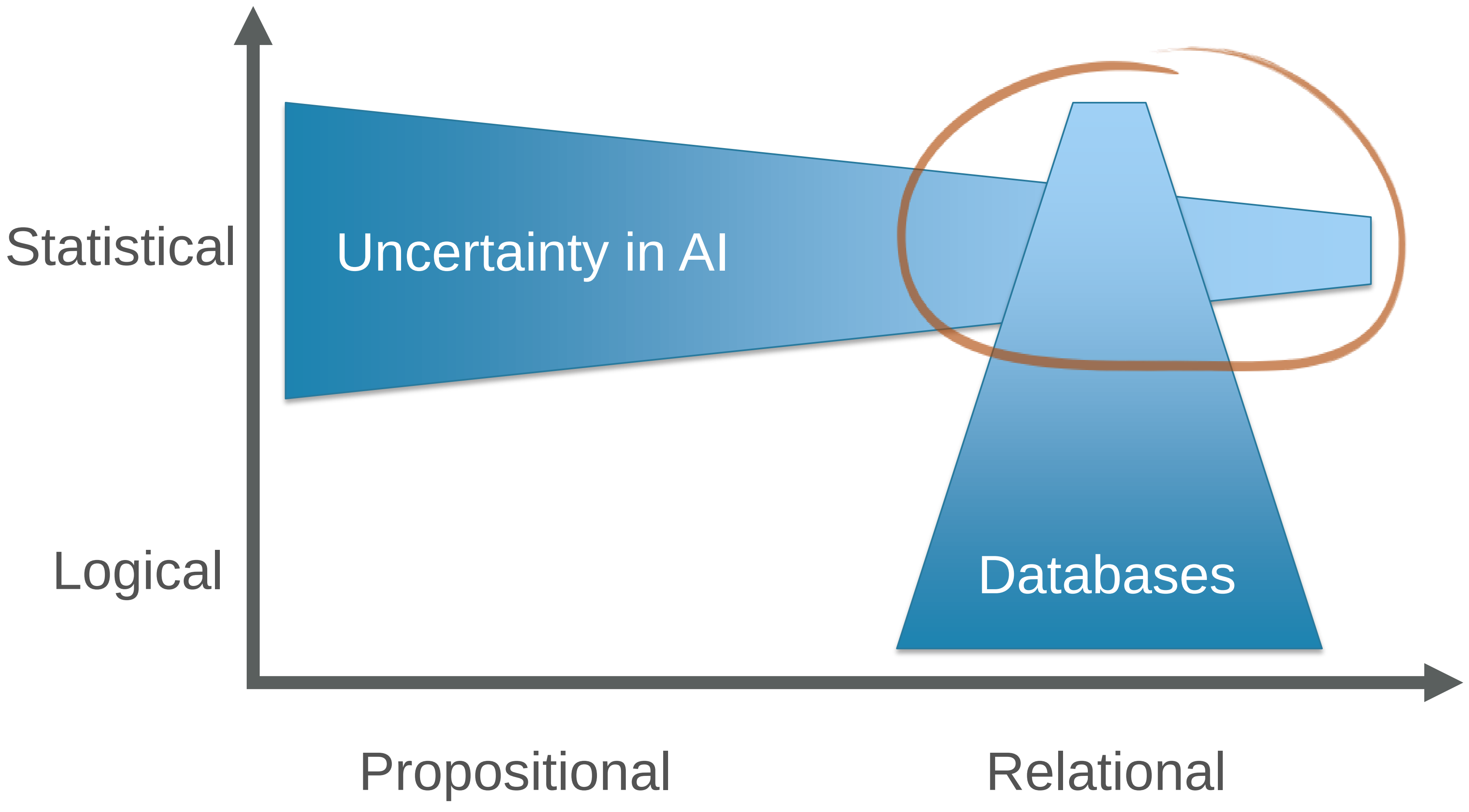
[Bach'15]

Summary



Statistical
Logical
Uncertainty in AI
StarAI
Propositional
Relational

Summary

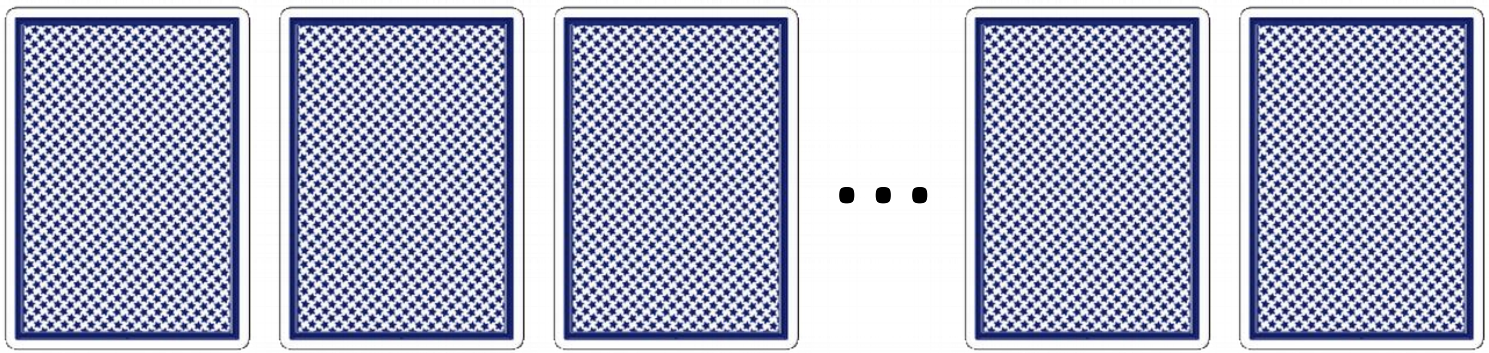

Statistical
Logical
Uncertainty in AI
Databases
Propositional
Relational

Why Lifted Inference?

- Main idea: exploit high level relational representation to speed up reasoning

- Let’s see an example…

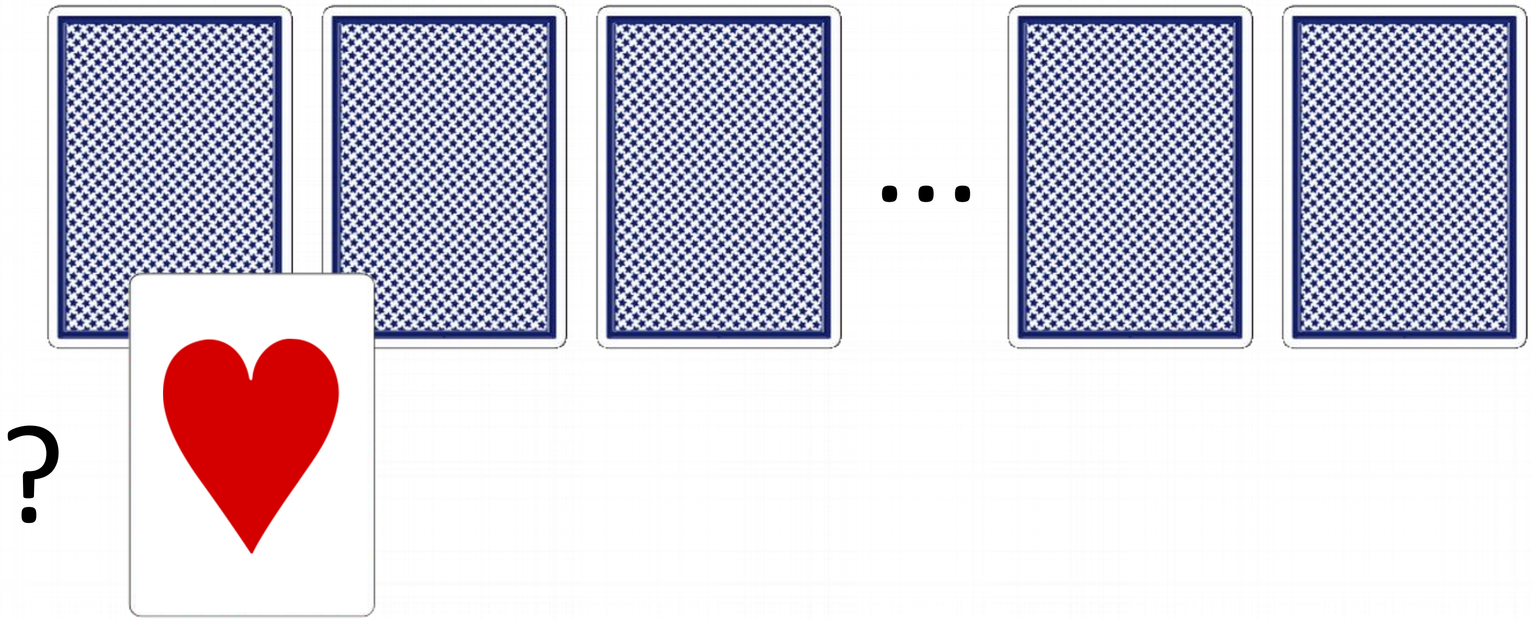
A Simple Reasoning Problem



- 52 playing cards
- Let us ask some simple questions

[VdB'15]

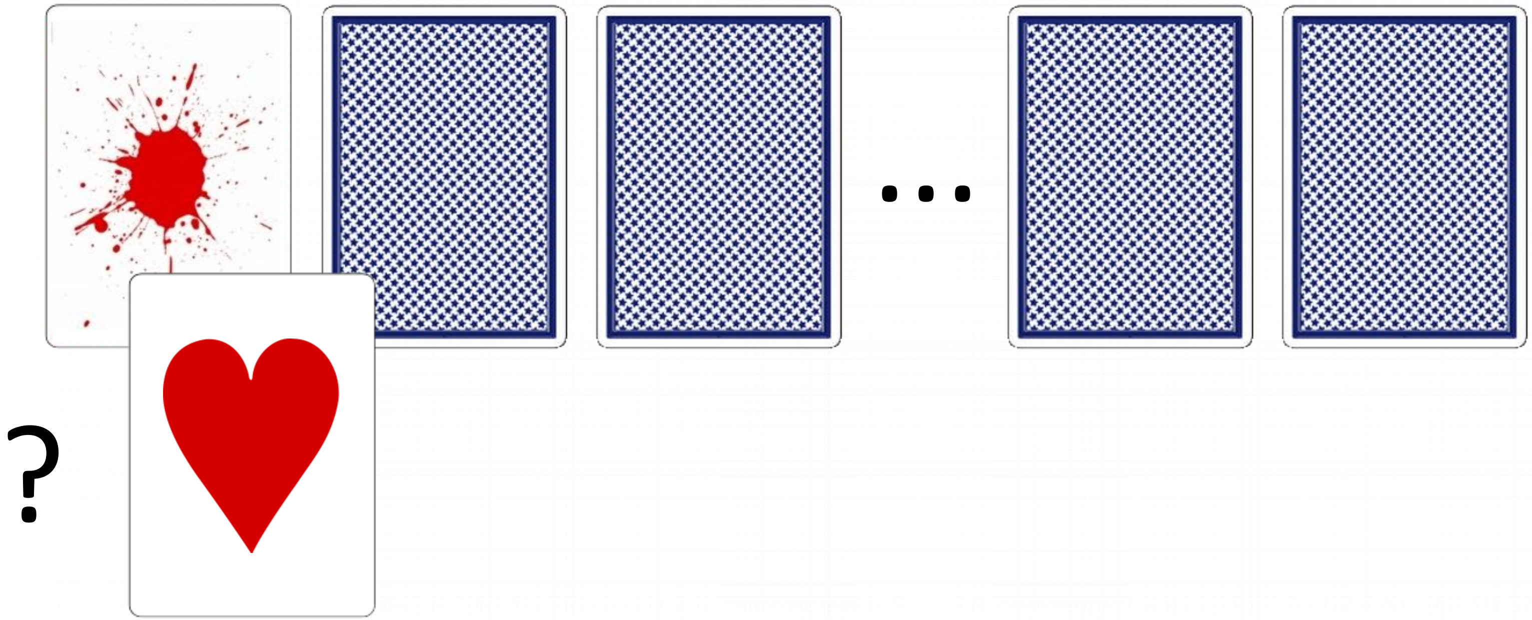
A Simple Reasoning Problem



Probability that Card1 is Hearts? 1/4

[VdB'15]

A Simple Reasoning Problem

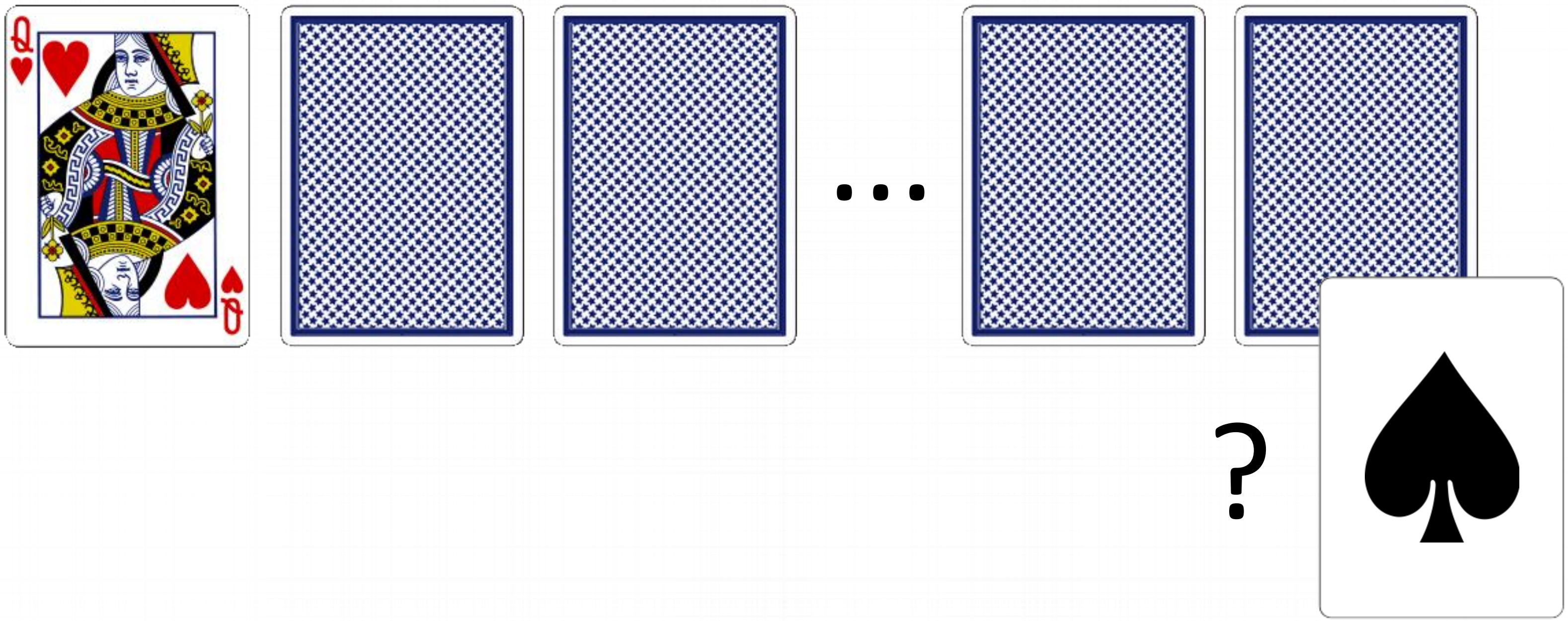


Probability that Card1 is Hearts given that Card1 is red?

1/2

[VdB'15]

A Simple Reasoning Problem

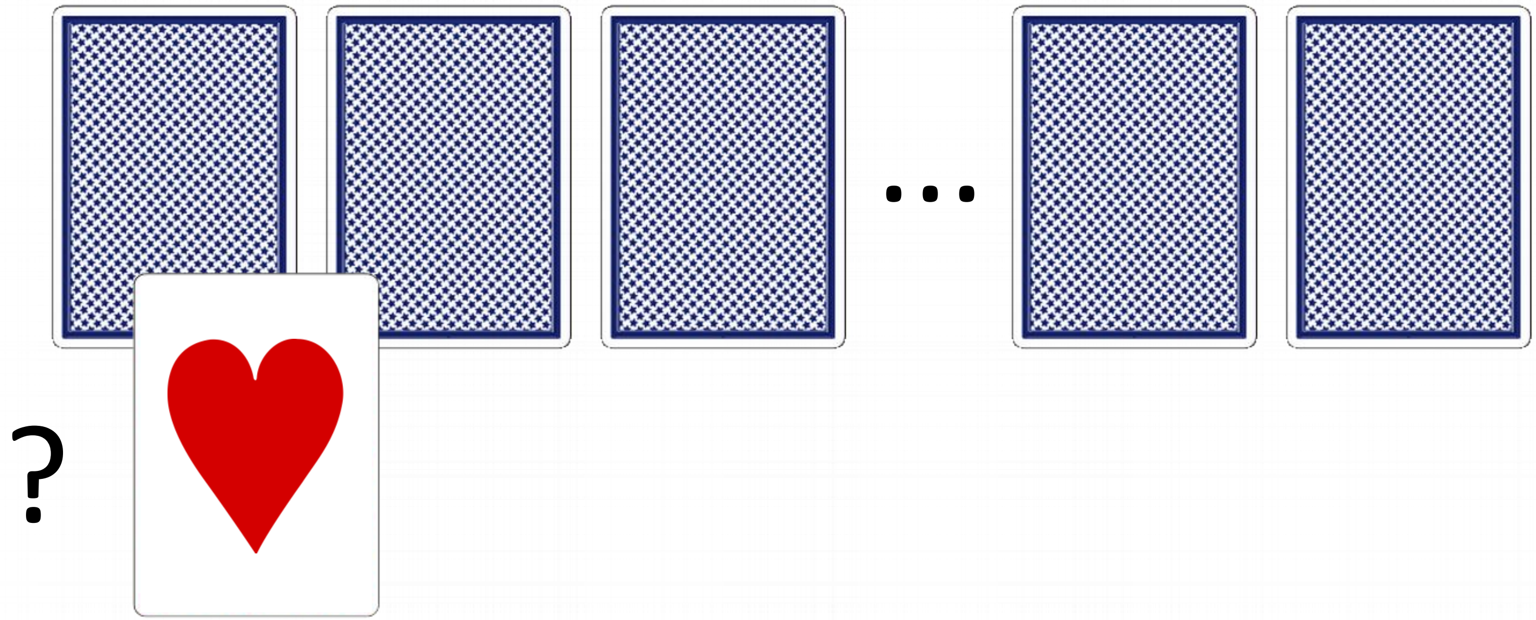


Probability that Card52 is Spades given that Card1 is QH?

13/51

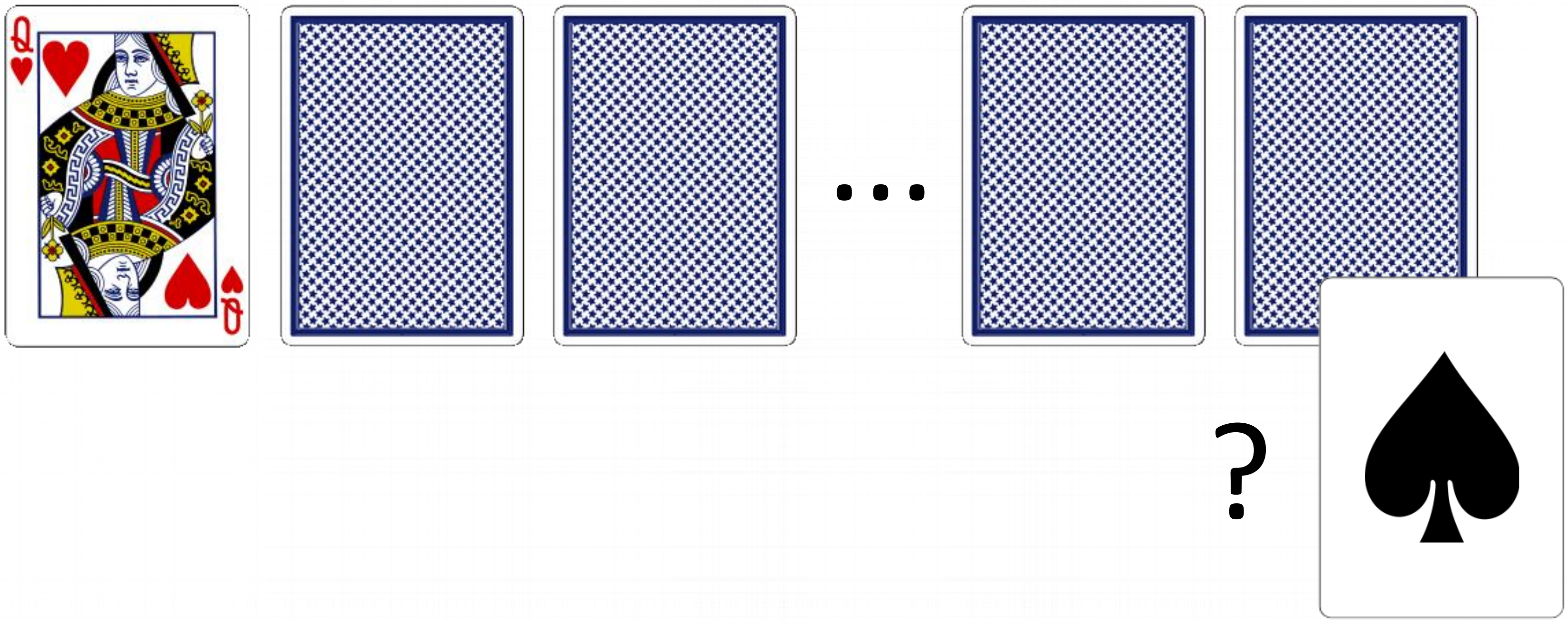
[VdB'15]

A Simple Reasoning Problem



Probability that Card1 is Hearts? 1/4

A Simple Reasoning Problem



Probability that Card52 is Spades given that Card1 is QH?

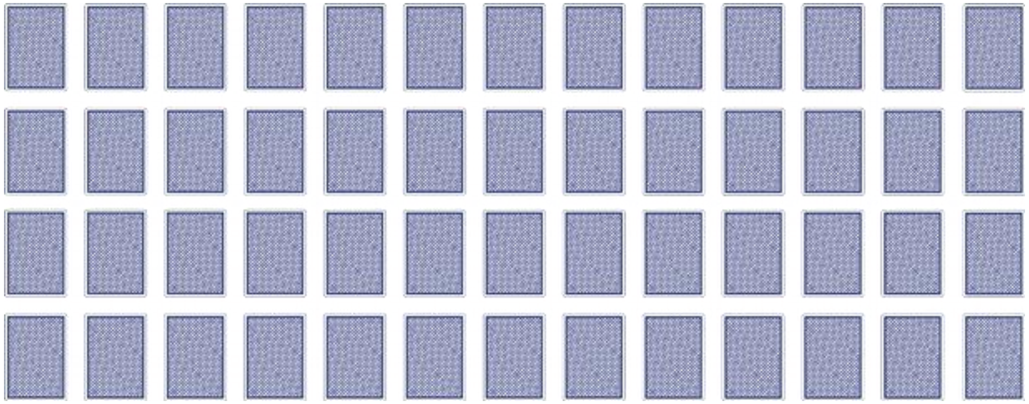
13/51

[VdB'15]

Automated Reasoning

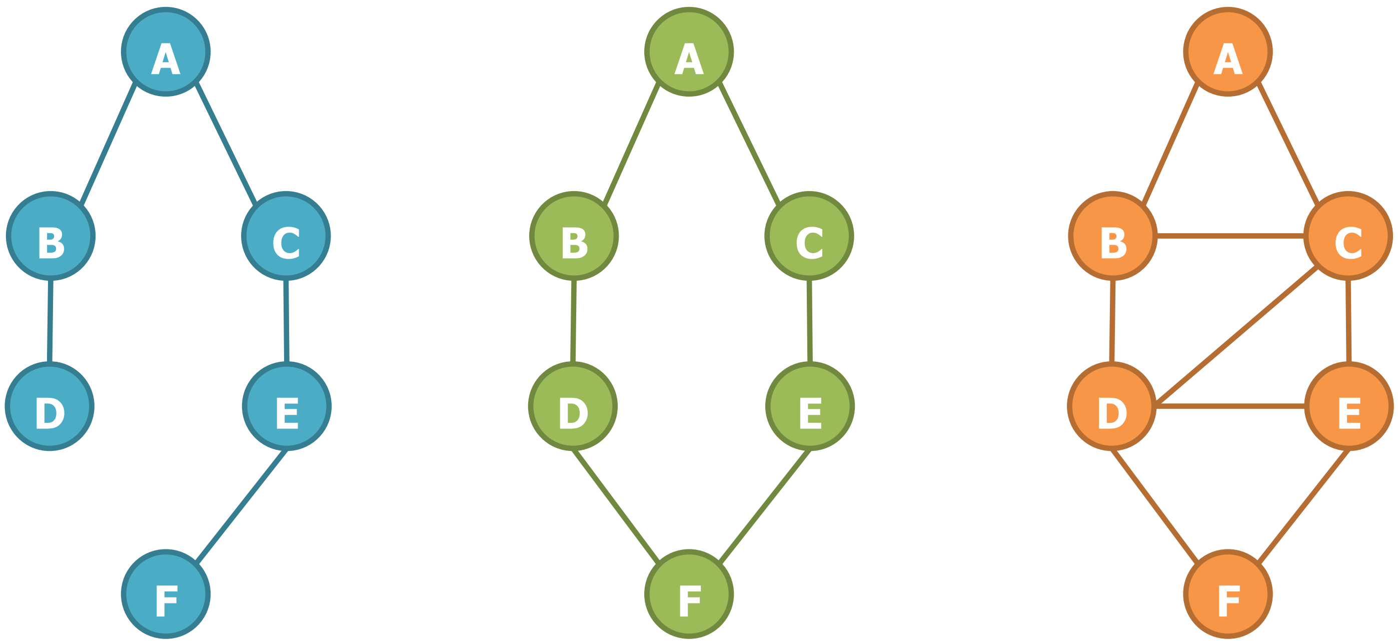
Let us automate this:

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



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

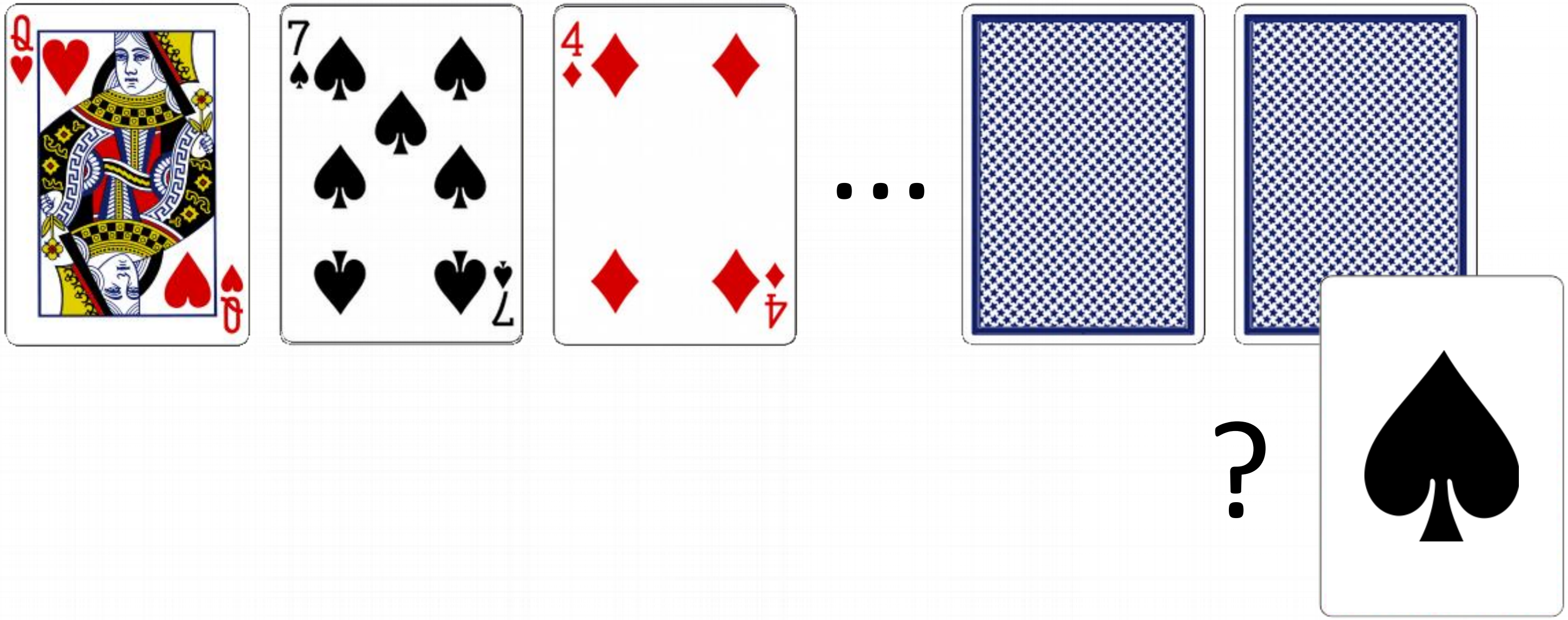
Classical Reasoning



Tree *Sparse Graph* *Dense Graph*

- Higher treewidth
- Fewer conditional independencies
- Slower inference

Is There Conditional Independence?



$P(\text{Card52} \mid \text{Card1}) \neq P(\text{Card52} \mid \text{Card1}, \text{Card2})$

$13/51 \neq 12/50$

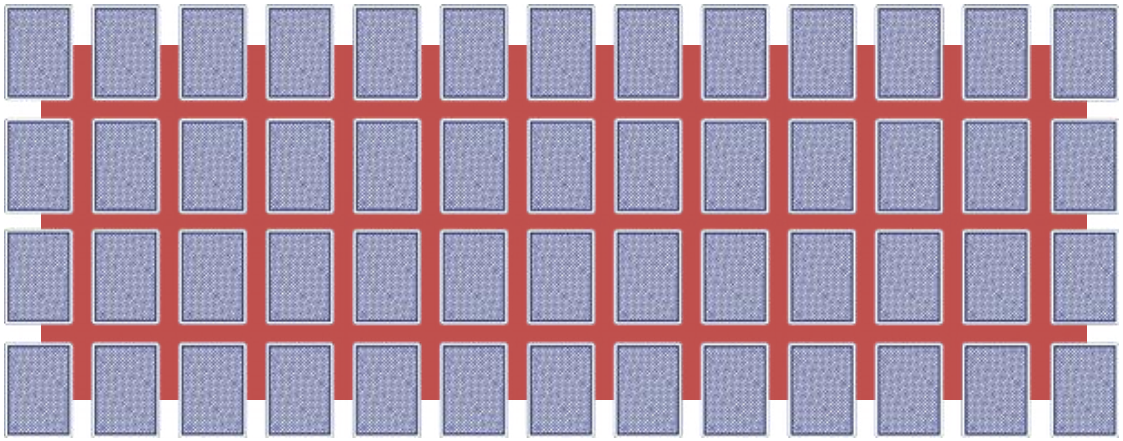
$P(\text{Card52} \mid \text{Card1}, \text{Card2}) \neq P(\text{Card52} \mid \text{Card1}, \text{Card2}, \text{Card3})$

$12/50 \neq 12/49$

Automated Reasoning

Let us automate this:

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

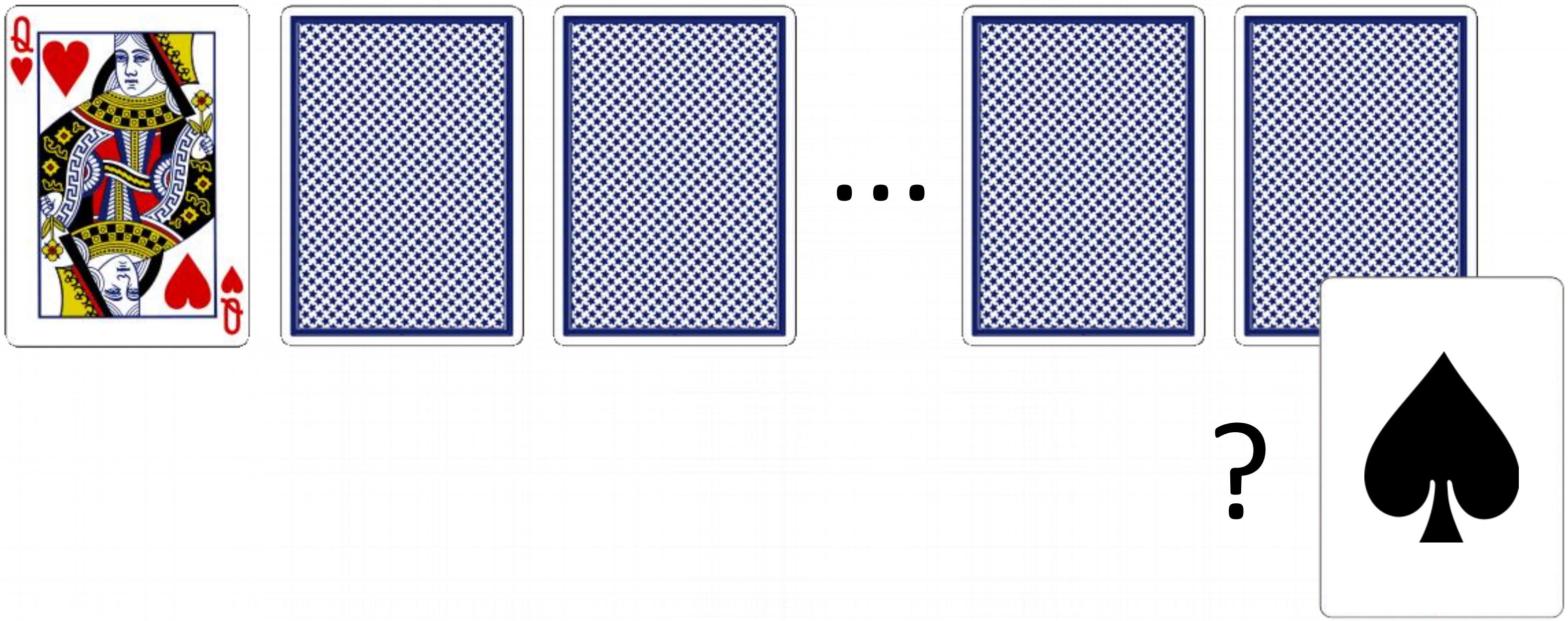


(artist's impression)

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

[VdB'15]

What's Going On Here?

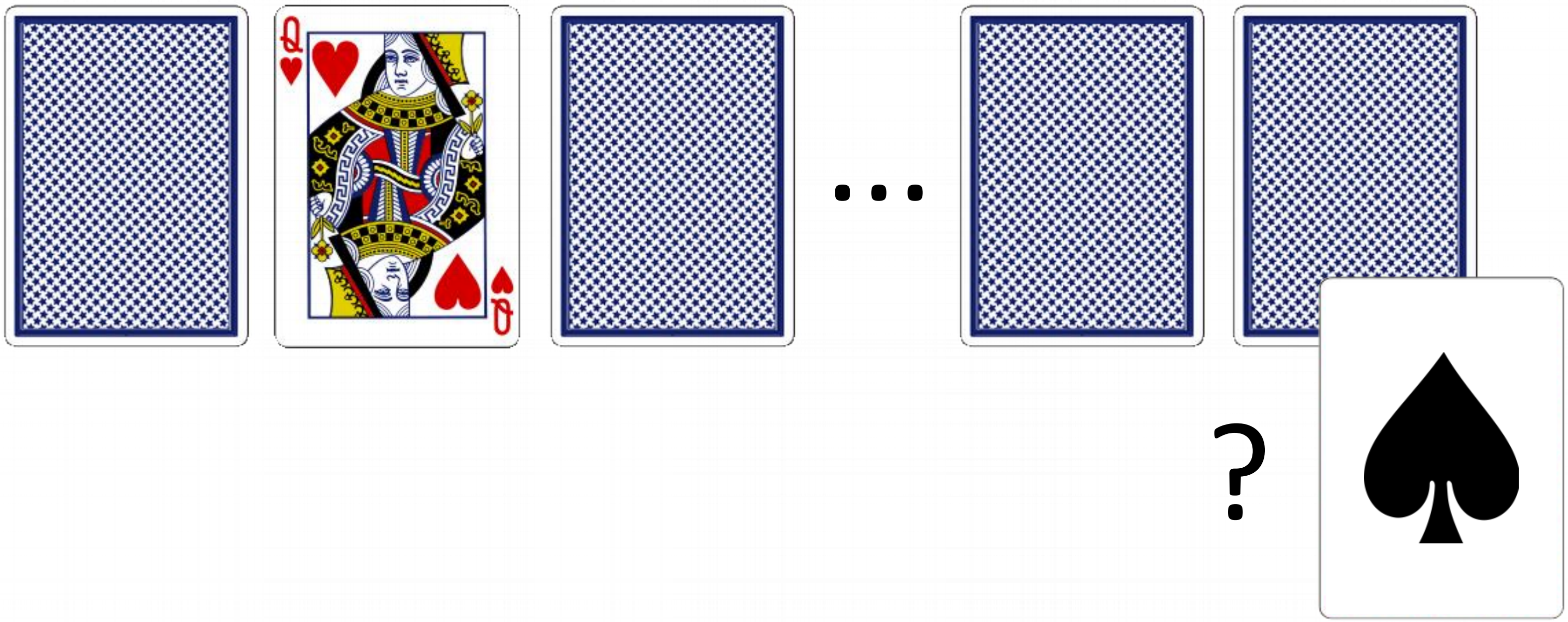


Probability that Card52 is Spades given that Card1 is QH?

13/51

[VdB' 15]

What's Going On Here?

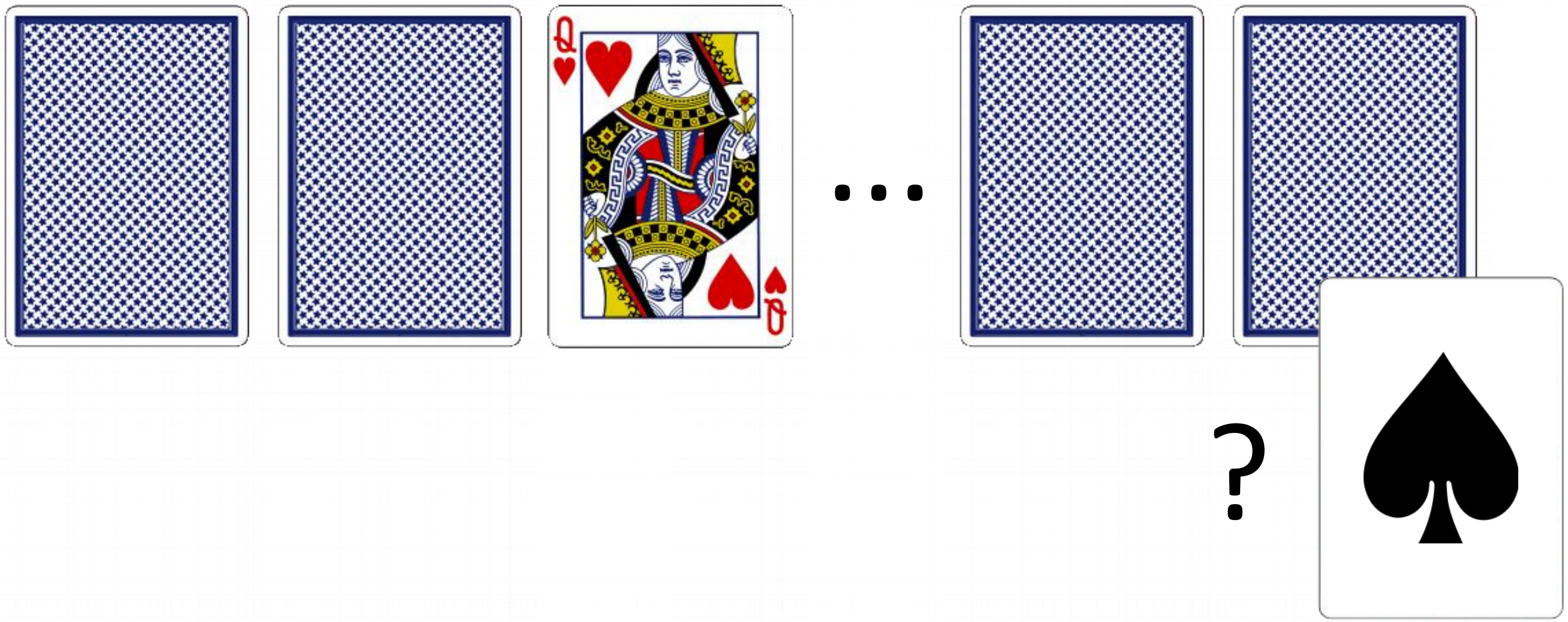


Probability that Card52 is Spades given that Card2 is QH?

13/51

[VdB' 15]

What's Going On Here?

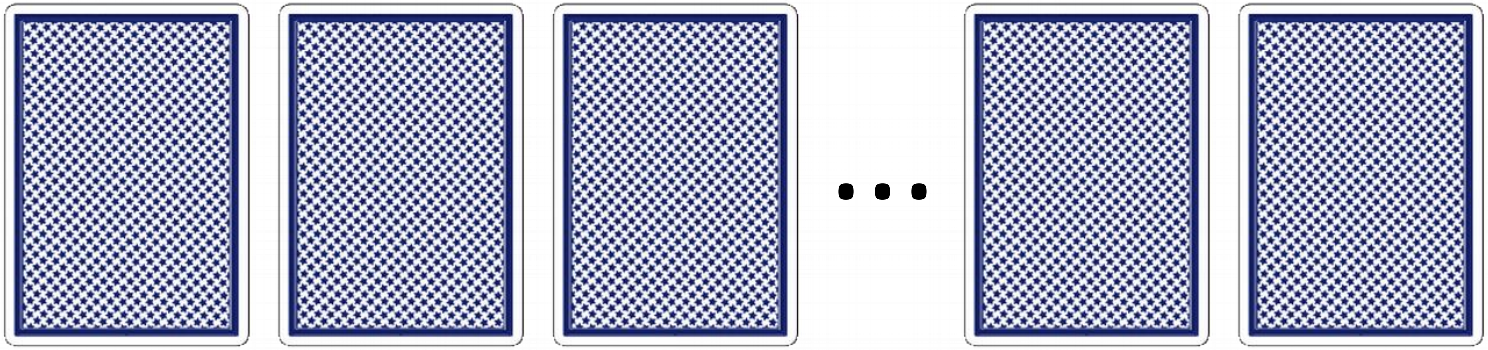


Probability that Card52 is Spades given that Card3 is QH?

13/51

[VdB' 15]

Tractable Reasoning



What's going on here?

Which property makes reasoning tractable?

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

⇒ **Lifted Inference**

[Niepert'14]

Automated Reasoning

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

Other Examples of Lifted Inference

- First-order resolution

$$\forall x,\ \mathrm{Human}(x) \Rightarrow \mathrm{Mortal}(x)$$
$$\forall x,\ \mathrm{Greek}(x) \Rightarrow \mathrm{Human}(x)$$

implies

$$\forall x,\ \mathrm{Greek}(x) \Rightarrow \mathrm{Mortal}(x)$$

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

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

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

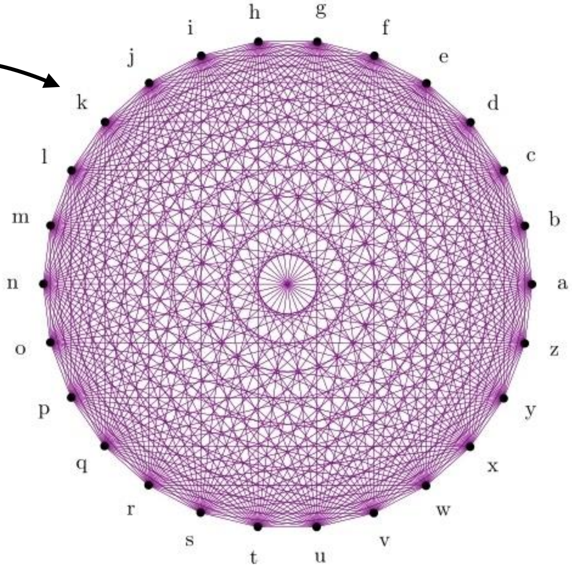
[VdB' 13]

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!



Statistical Properties

1. Independence

P(

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

) = P(

Alice	1	1	0

) x P(

Bob	0	0	0

) x P(

Charlie	0	1	0

)

2. Partial Exchangeability

P(

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

) = P(

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

)

3. Independent and identically distributed (i.i.d.)
   = Independence + Partial Exchangeability

Statistical Properties for Tractability

- Tractable classes independent of representation
- Traditionally:
  - Tractable learning from **i.i.d. data**
  - Tractable inference when **cond. independence**
- New understanding:
  - Tractable learning from **exchangeable data**
  - Tractable inference when
    - **Conditional independence**
    - **Conditional exchangeability**
    - **A combination**

[Niepert'14]

Summary of Motivation

- Relational data is everywhere:
  - Databases in industry and sciences
  - Knowledge bases
  - Probabilistically extracted/learned/queried
- Lifted inference:
  - Use relational structure during reasoning
  - Very efficient where traditional methods break

This tutorial: Lifted Inference in Relational Models