Probabilistic Reasoning and Machine Learning

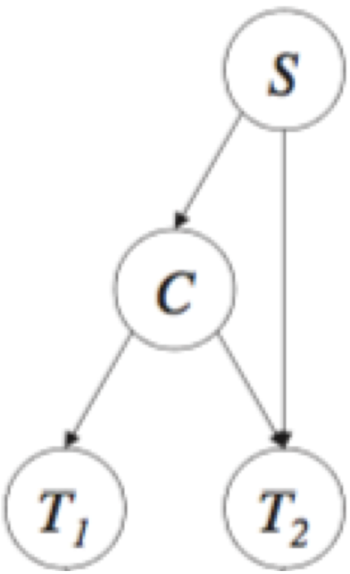
Adnan Darwiche
Computer Science Department, UCLA

Probabilistic Inference

- Prior & Posterior Marginals
  (most common)
- MPE: Most probable explanation
  (also called "MAP")
- MAP: Maximum a Posteriori Hypothesis
  (also called "partial or maginal MAP")

- SDP: Same-Decision Probability
  (relatively new ~ 2010)

Bayesian Network



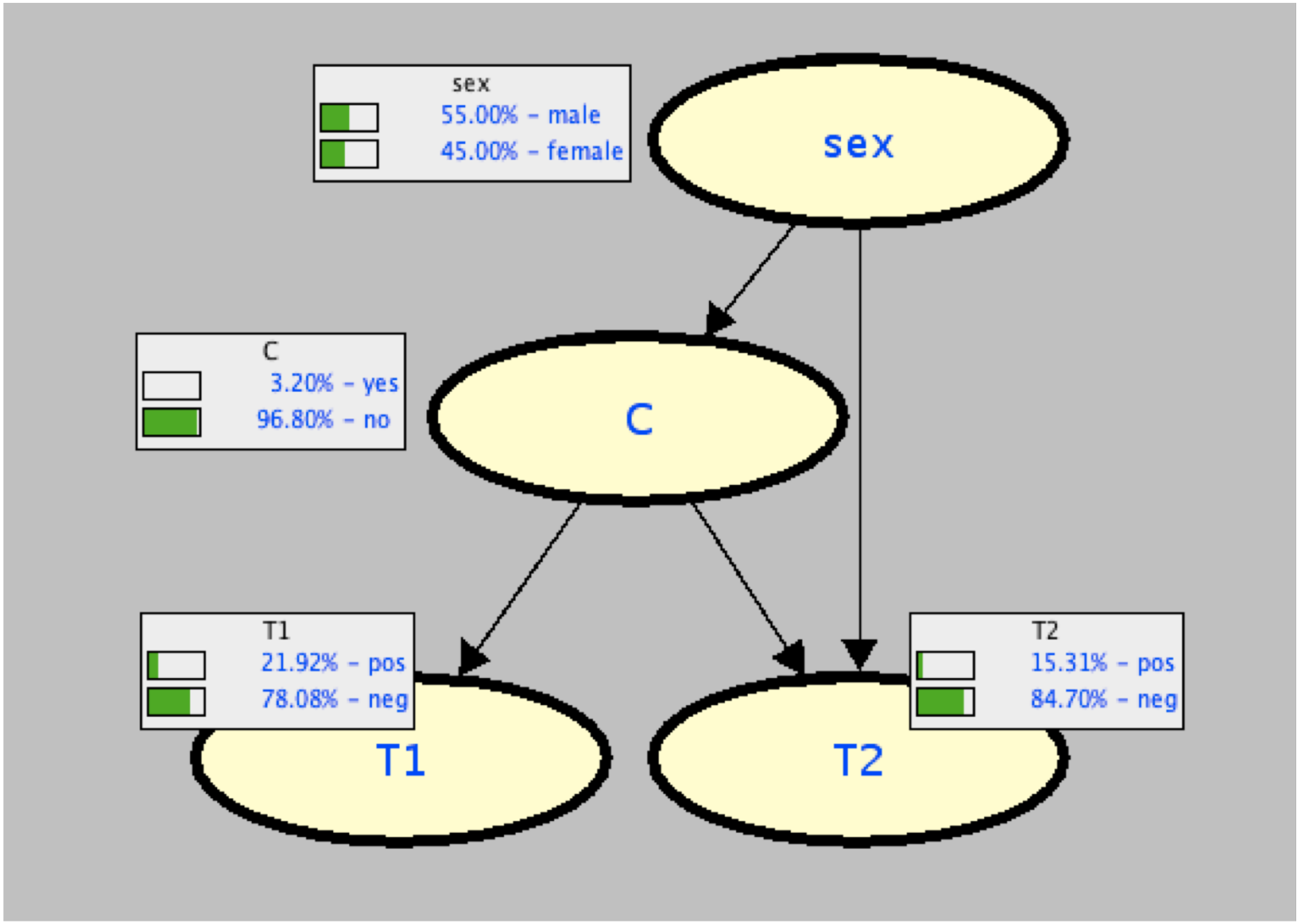
S	θ_s
male	.55
female	.45

S	C	$\theta_{c\|s}$
male	yes	.05
male	no	.95
female	yes	.01
female	no	.99

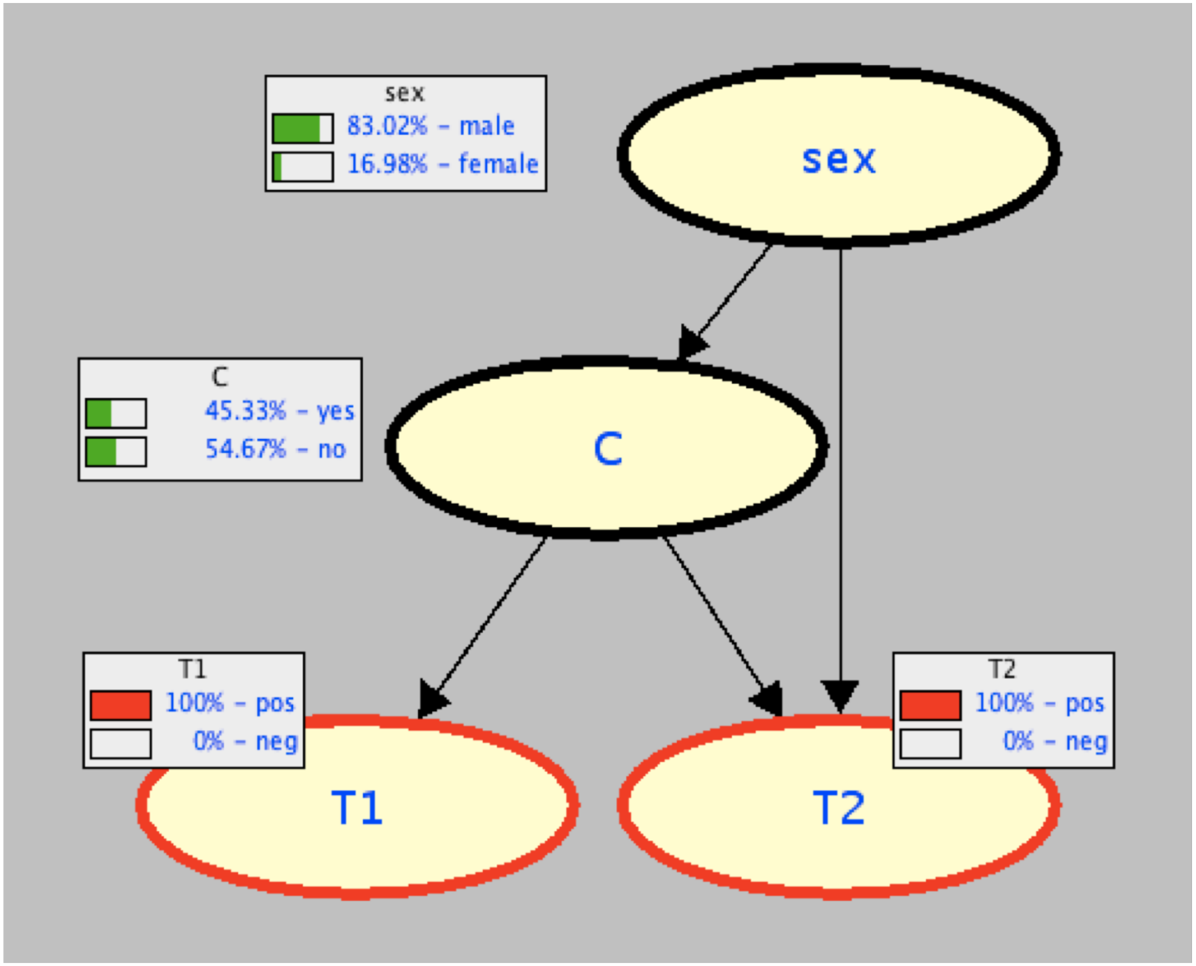
C	T_1	$\theta_{t_1\|c}$	
yes	+ve	.80	
yes	−ve	.20	= fn
no	+ve	.20	= fp
no	−ve	.80	

S	C	T_2	$\theta_{t_2\|c,s}$	
male	yes	+ve	.80	
male	yes	−ve	.20	= fn
male	no	+ve	.20	= fp
male	no	−ve	.80	
female	yes	+ve	.95	
female	yes	−ve	.05	= fn
female	no	+ve	.05	= fp
female	no	−ve	.95	

Prior Marginals

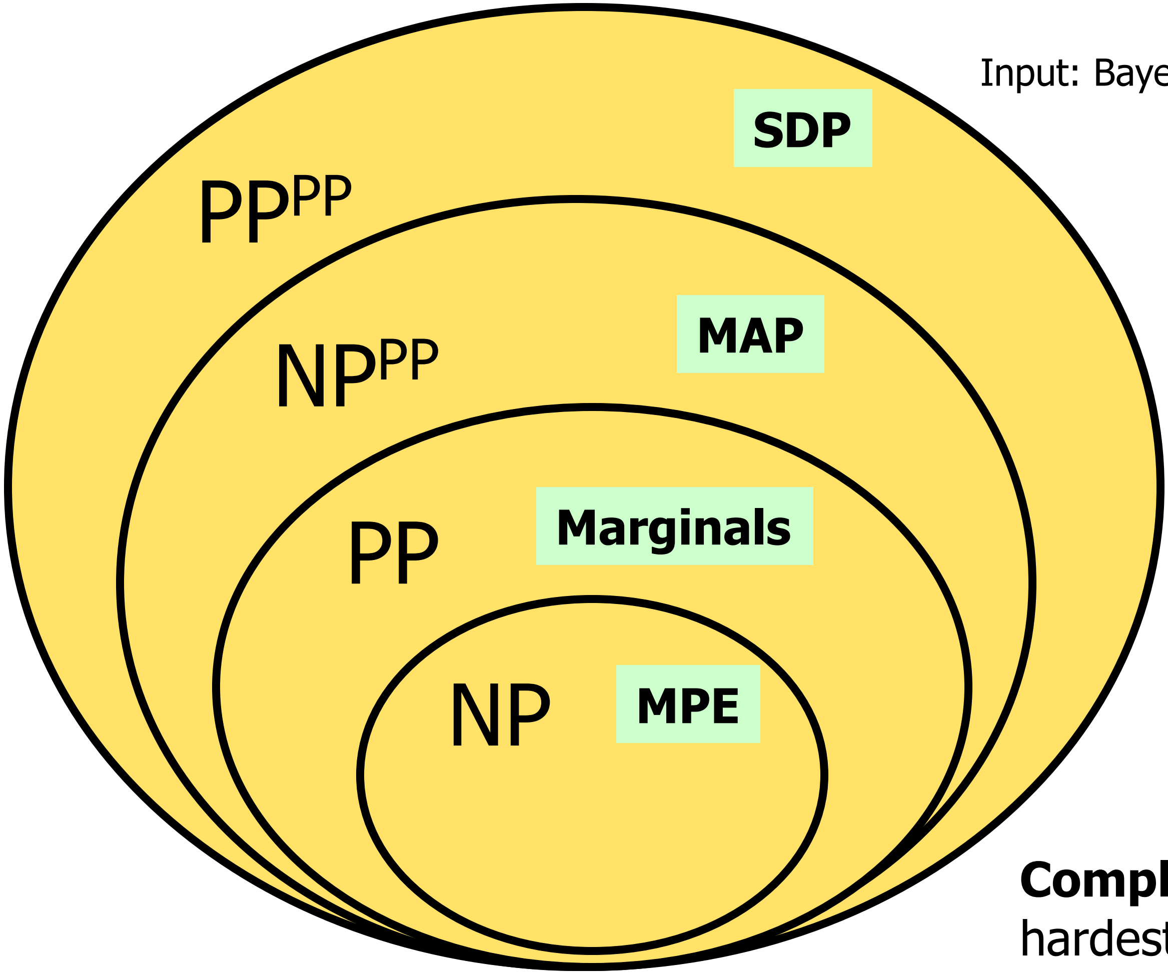


Posterior Marginals



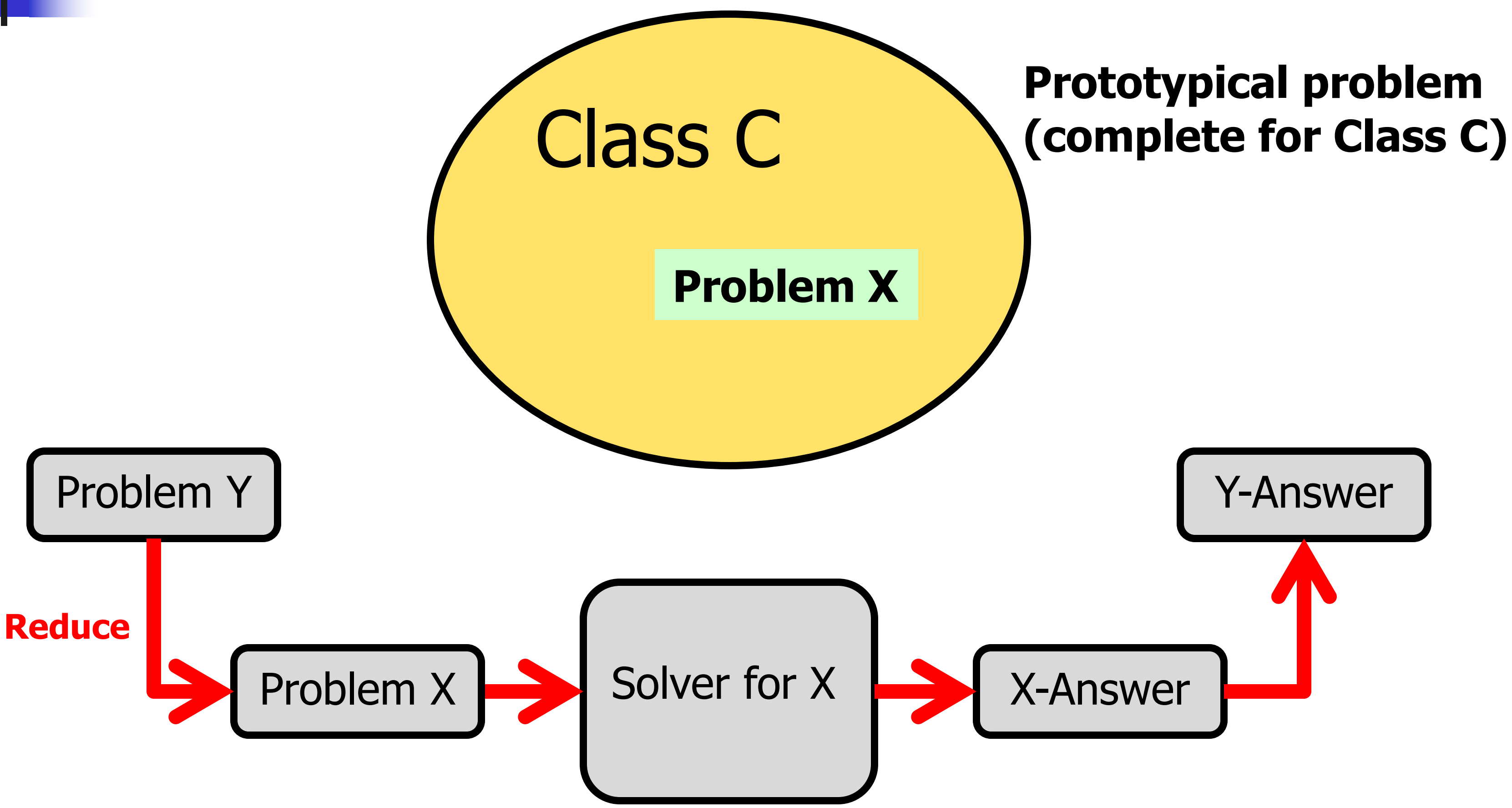
Probabilistic Inference

Input: Bayesian Network

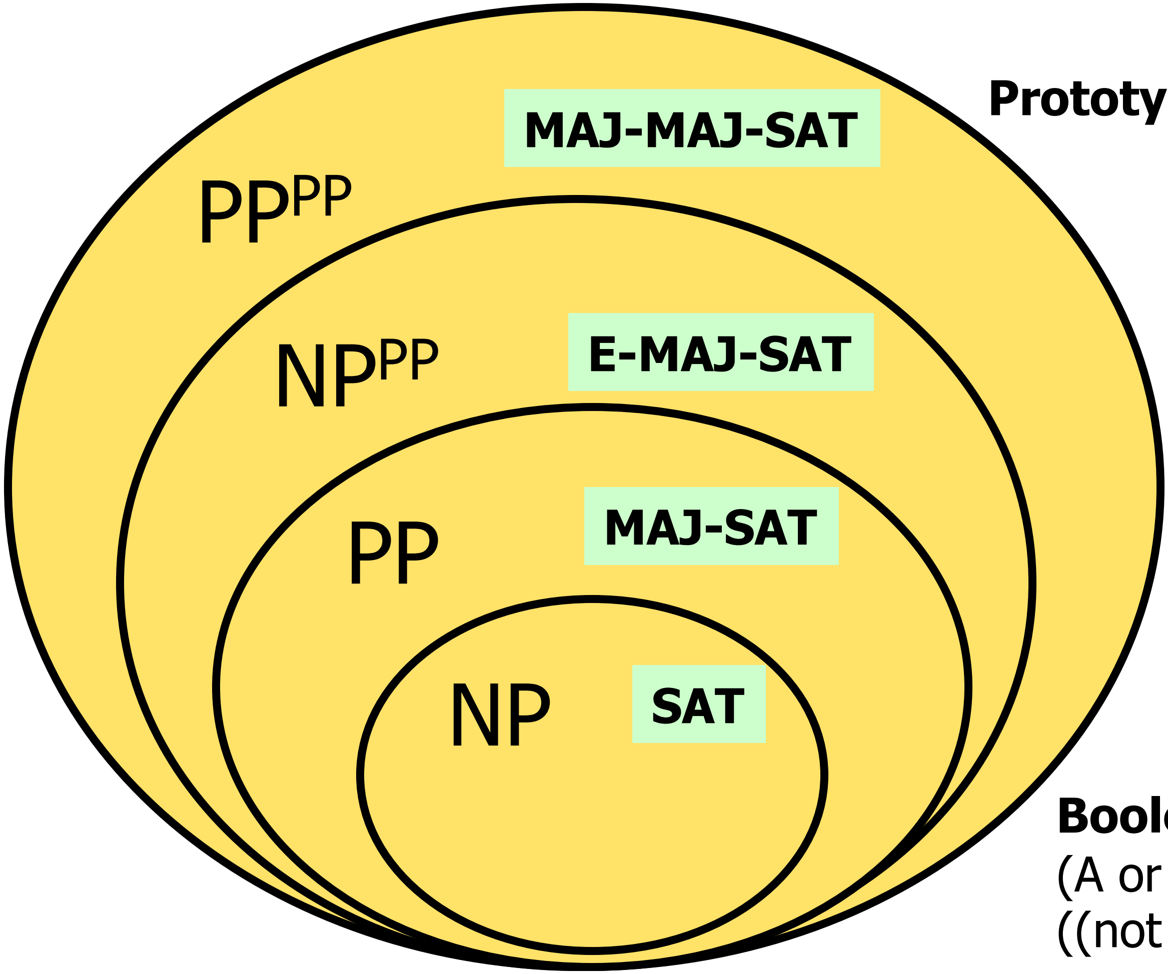


Complete problems:
hardest in their class

Reduction Approaches



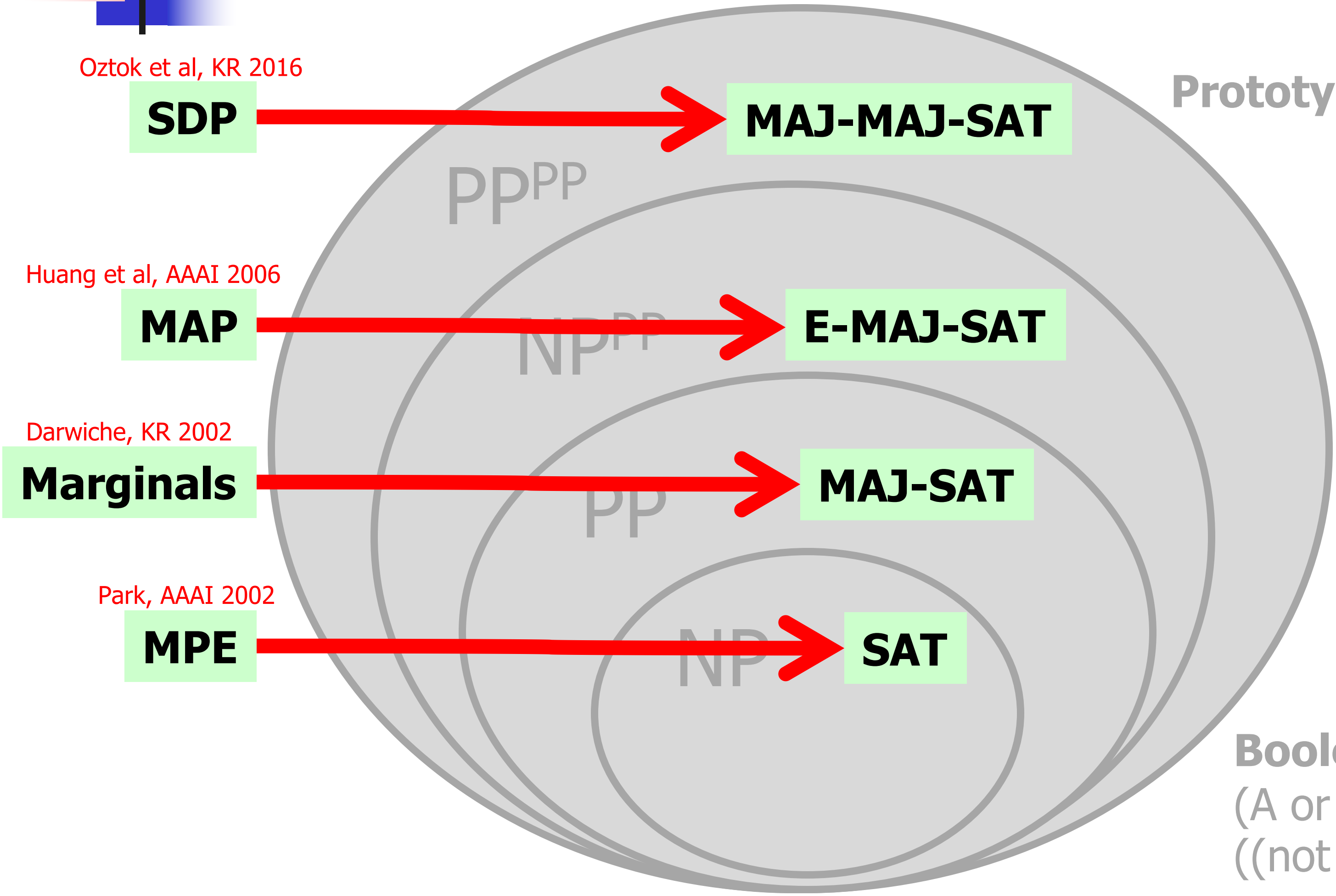
Reduction Approaches



Prototypical problems

Boolean expressions:
(A or (not B) or C),
((not A) or D or (not E)),
….

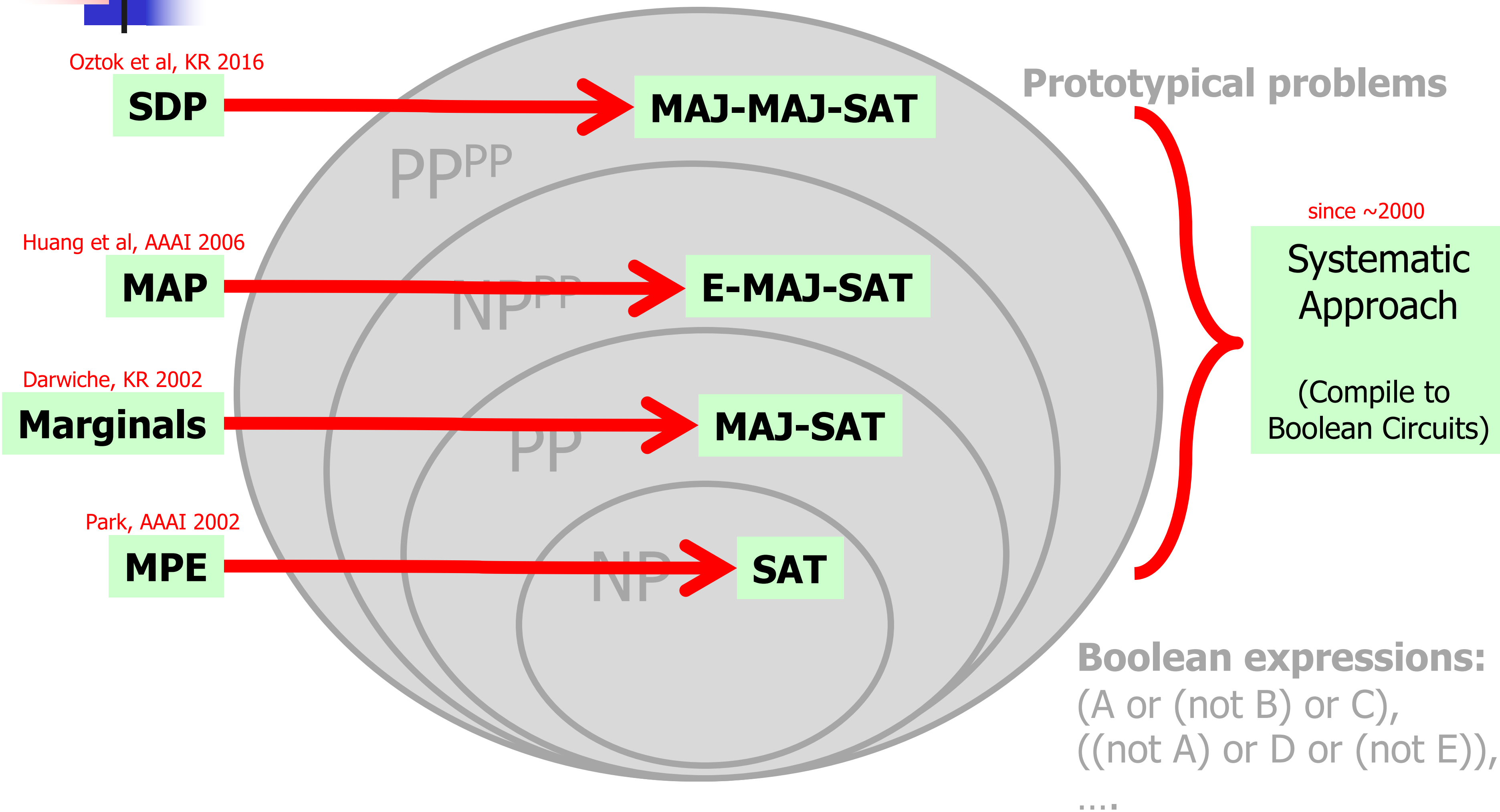
1st Line of Developments



Prototypical problems

Boolean expressions:
(A or (not B) or C),
((not A) or D or (not E)),
….

2^{nd} Line of Developments



SAT: NP-complete

Boolean expression:
(A or B) and (not C)

SAT: Is there a satisfying instantiation?
Yes

A	B	C
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

MAJ-SAT: PP-complete

Boolean expression:
(A or B) and (not C)

MAJ-SAT: Are the majority of instantiations satisfying?
No

A	B	C
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

MAJ-SAT Variant

Model Counting

Boolean expression:
(A or B) and (not C)

#SAT: How many satisfying assignment?
3

A	B	C
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

MAJ-SAT Variant

Weighted Model Counting

Boolean expression:
(A or B) and (not C)

WMC: The added weight of satisfying assignments?
0.14 = 0.04 + 0.10 + 0.00

$$w(A,\neg B,C) = w(A)w(\neg B)w(C)$$

A	B	C	
T	T	T	0.08
T	T	F	0.04
T	F	T	0.10
T	F	F	0.10
F	T	T	0.00
F	T	F	0.00
F	F	T	0.42
F	F	F	0.06

E-MAJ-SAT: NP^{PP}-complete

Boolean expression:

(A or B) and (not C)

Split variables **X**={C}, **Y**={A,B}

E-MAJ-SAT: Is there an **X**-instantiation under which the majority of **Y**-instantiations satisfying?

Yes

A	B	C
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

E-MAJ-SAT: NP^{PP}-complete

Boolean expression:

(A or B) and (not C)

Split variables **X**={C}, **Y**={A,B}

E-MAJ-SAT: Is there an **X**-instantiation under which the majority of **Y**-instantiations satisfying?

Yes

A	B	C
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

MAJ-MAJ-SAT: PP^{PP}-complete

Boolean expression:

(A or B) and (not C)

Split variables **X**={C}, **Y**={A,B}

MAJ-MAJ-SAT: Is there a majority of **X**-instantiation under which the majority of **Y**-instantiations satisfying?

No

A	B	C
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

MAJ-MAJ-SAT: PP^{PP}-complete

Boolean expression:

(A or B) and (not C)

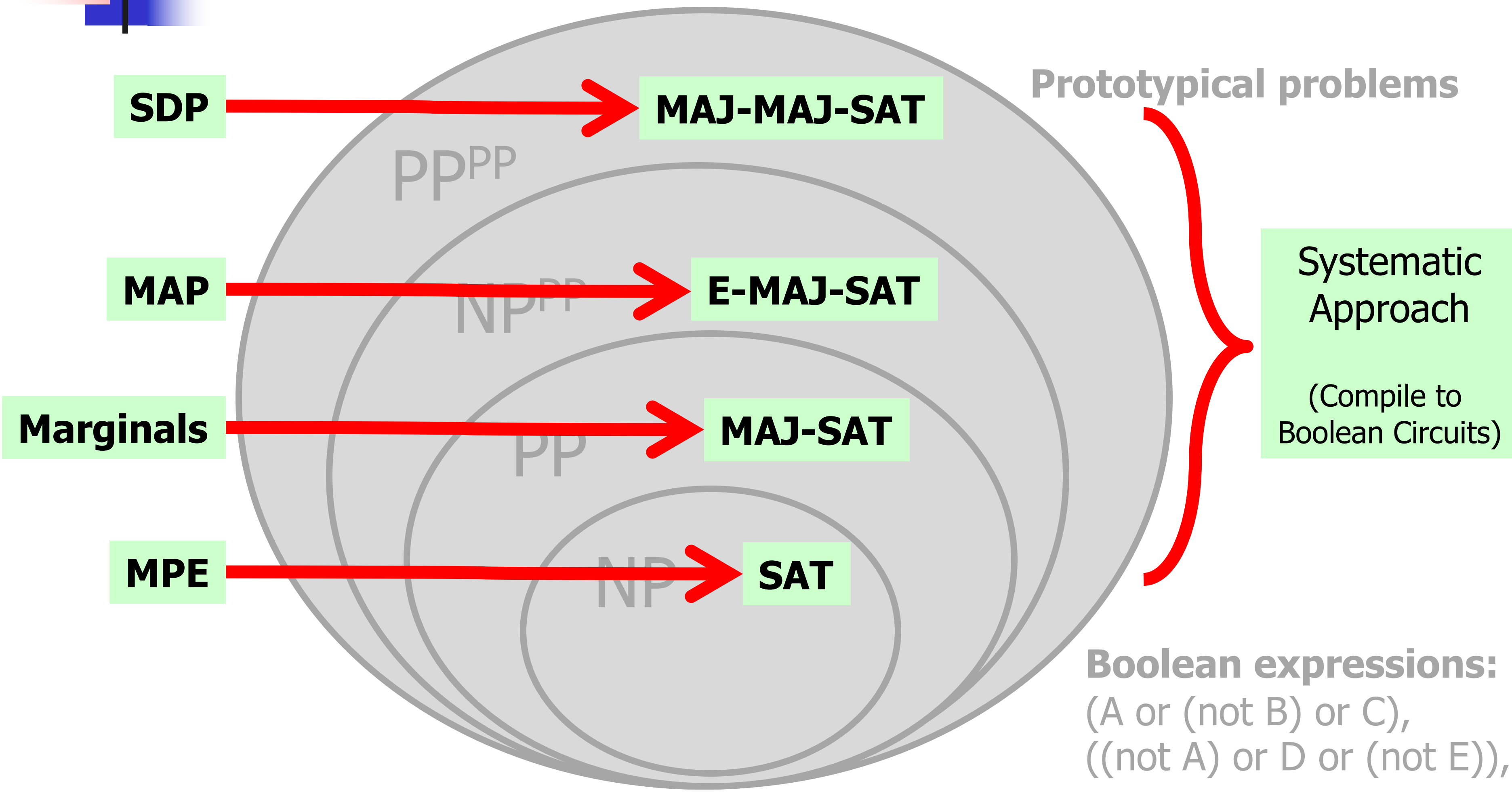
Split variables **X**={C}, **Y**={A,B}

MAJ-MAJ-SAT: Is there a majority of **X**-instantiation under which the majority of **Y**-instantiations satisfying?

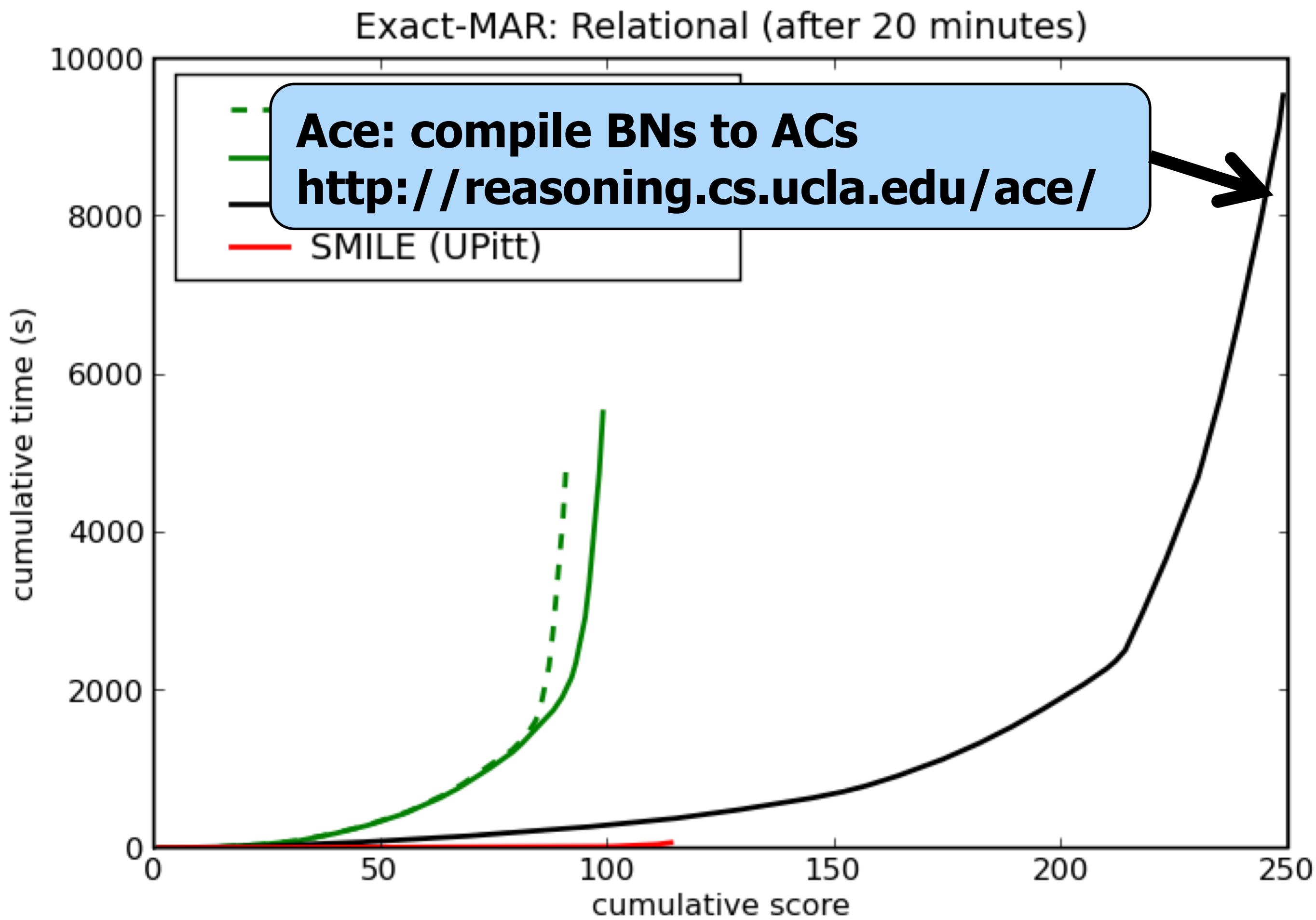
No

A	B	C
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

Reductions



Boolean expressions:
(A or (not B) or C),
((not A) or D or (not E)),
….



- Relational networks (251 networks)
  - Average cluster size is 50

Machine Learning

Learning with Background Knowledge

Logic (L)
Knowledge Representation (K)
Probability (P)
Artificial Intelligence (A)

Data

L	K	P	A	Students
0	0	1	0	6
0	0	1	1	54
0	1	1	1	10
1	0	0	0	5
1	0	1	0	1
1	0	1	1	0
1	1	0	0	17
1	1	1	0	4
1	1	1	1	3

Background Knowledge

Must take at least one of Probability or Logic.
Probability is a prerequisite for AI.
The prerequisites for KR is either AI or Logic.

$$P \vee L \quad A \Rightarrow P \quad K \Rightarrow (P \vee L)$$

Learning with Background Knowledge

unstructured

L	K	P	A
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

- Must take at least one of Probability or Logic.
- Probability is a prerequisite for AI.
- The prerequisites for KR is either AI or Logic.

7 out of 16 instantiations are impossible

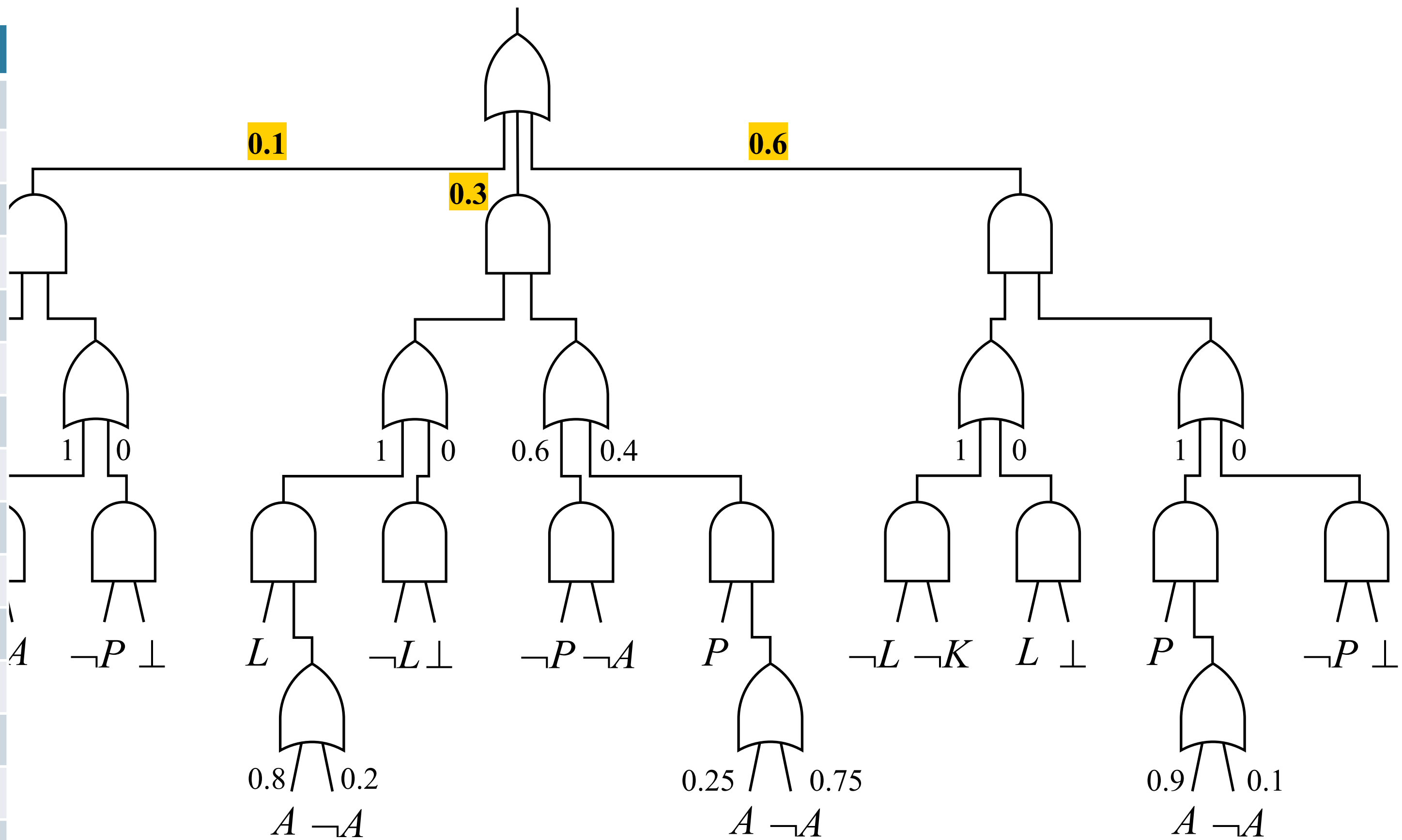
structured

L	K	P	A
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

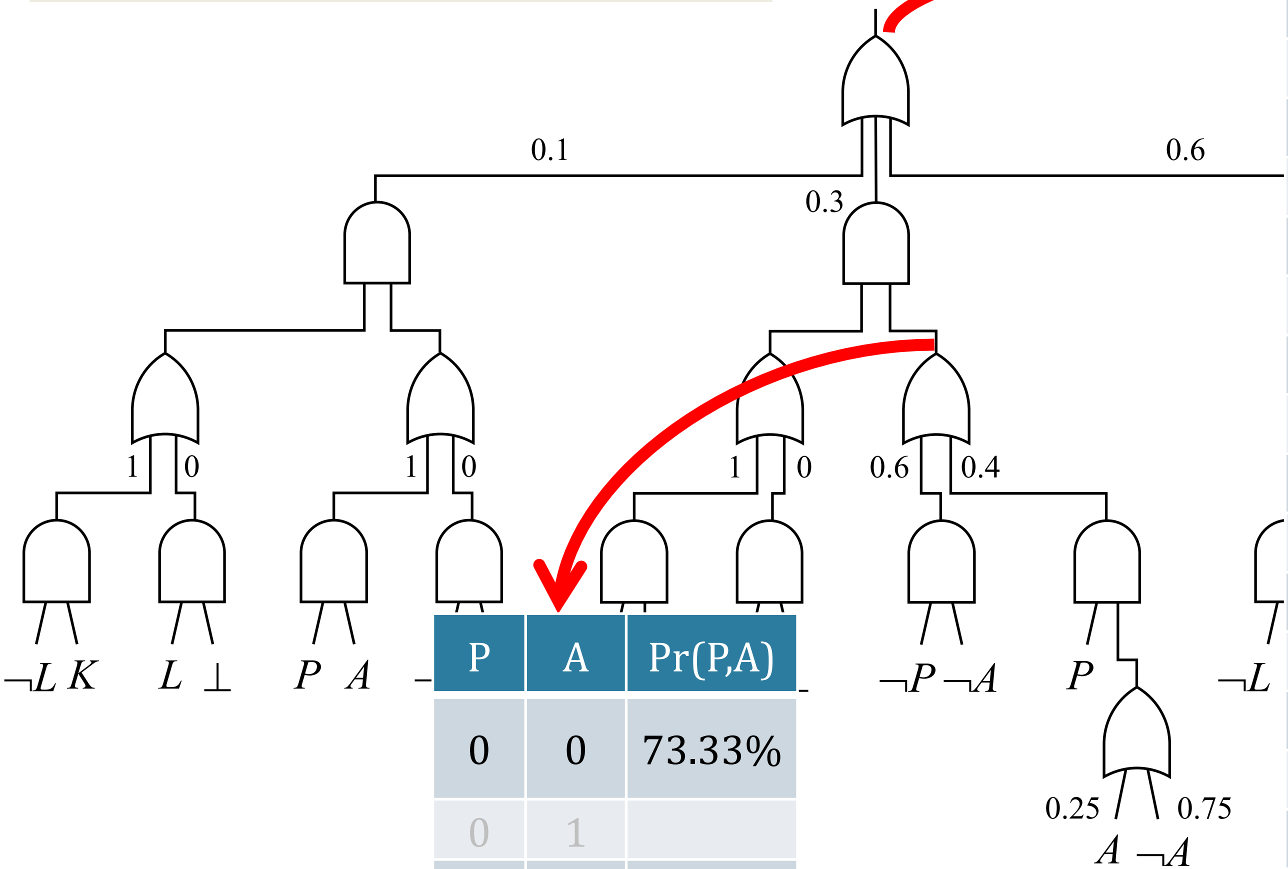
Probabilistic SDD Circuits

Kisa et al, KR 2014

L	K	P	A
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1



induces a normalized distribution over satisfying assignments (structured probability space)



L	K	P	A	P(L,K,P,A)
0	0	0	0	
0	0	0	1	
0	0	1	0	6.0%
0	0	1	1	54.0%
0	1	0	0	
0	1	0	1	
0	1	1	0	
0	1	1	1	10.0%
1	0	0	0	4.4%
1	0	0	1	
1	0	1	0	1.0%
1	0	1	1	0.6%
1	1	0	0	17.6%
1	1	0	1	
1	1	1	0	4.0%
1	1	1	1	2.4%

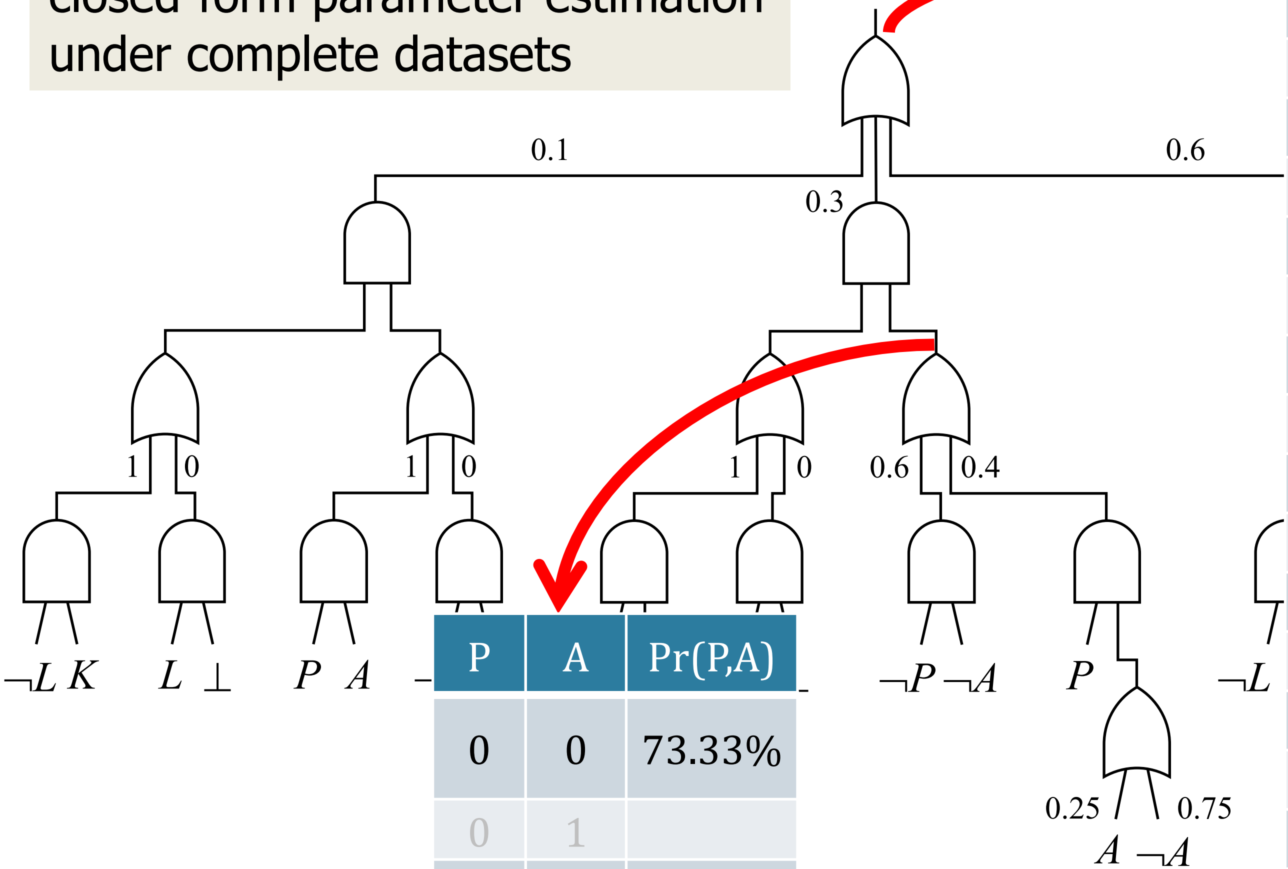
P	A	Pr(P,A)
0	0	73.33%
0	1	
1	0	16.67%
1	1	10.00%

A. Darwiche

complete & canonical representation

tractable for MPE & marginals

closed-form parameter estimation
under complete datasets



L	K	P	A	P(L,K,P,A)
0	0	0	0	
0	0	0	1	
0	0	1	0	6.0%
0	0	1	1	54.0%
0	1	0	0	
0	1	0	1	
0	1	1	0	
0	1	1	1	10.0%
1	0	0	0	4.4%
1	0	0	1	
1	0	1	0	1.0%
1	0	1	1	0.6%
1	1	0	0	17.6%
1	1	0	1	
1	1	1	0	4.0%
1	1	1	1	2.4%

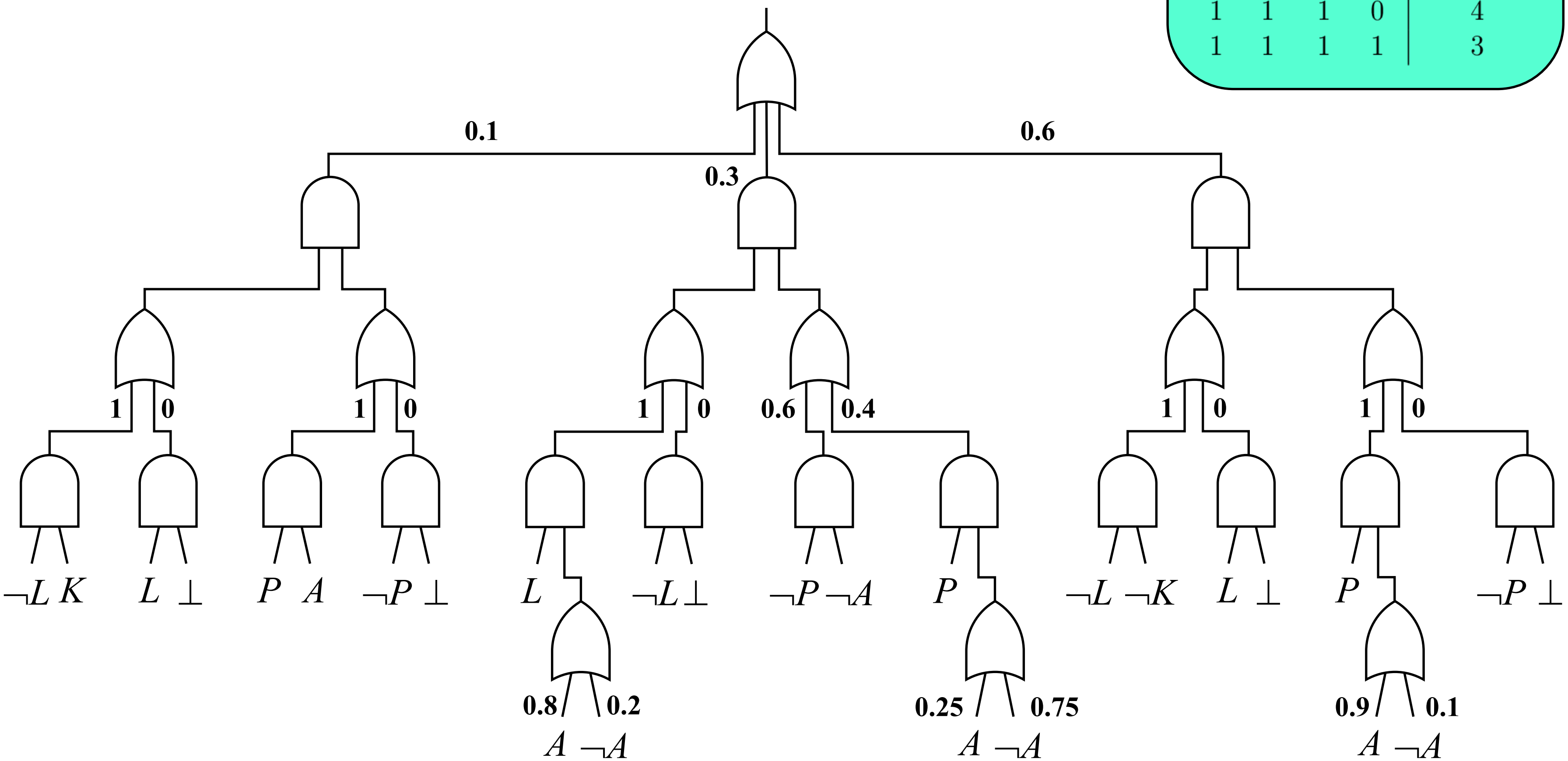
P	A	Pr(P,A)
0	0	73.33%
0	1	
1	0	16.67%
1	1	10.00%

Learned PSDD

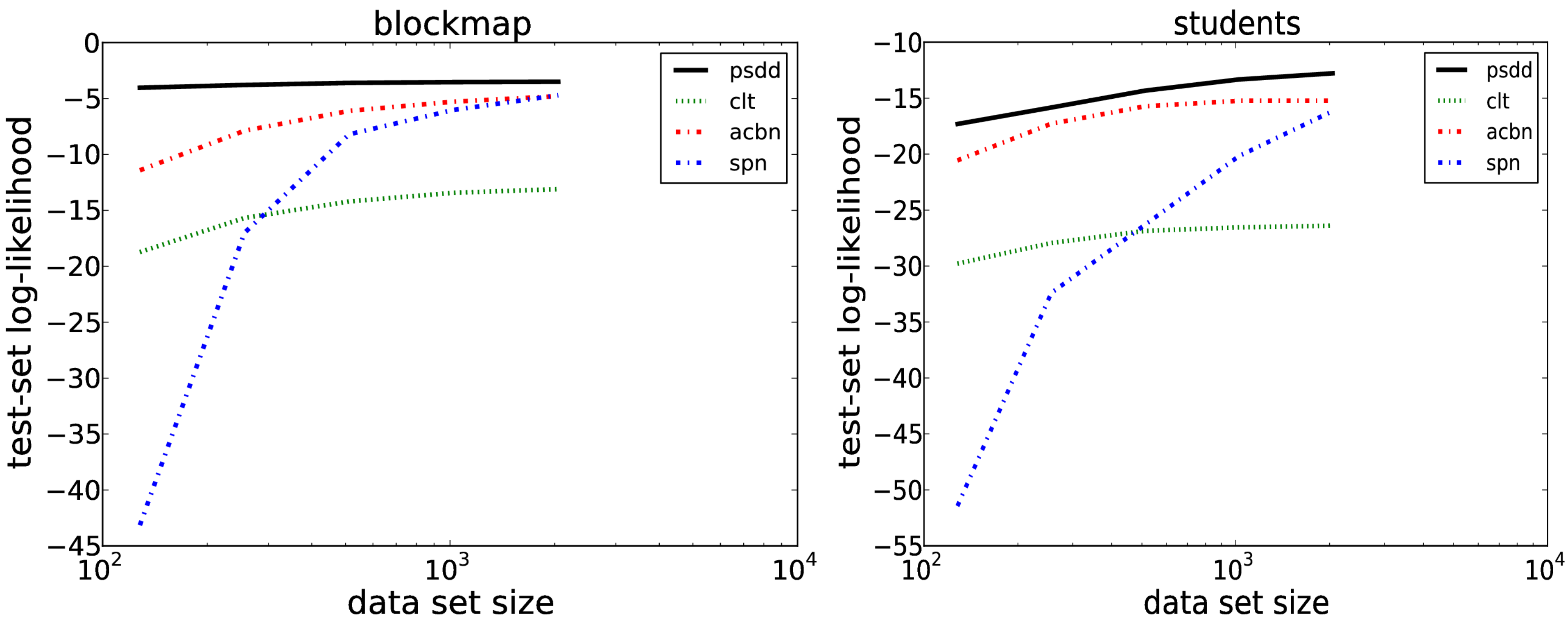
$P \lor L$
$A \Rightarrow P$
$K \Rightarrow (P \lor L)$

closed-form parameter estimation
under complete datasets

L	K	P	A	Students
0	0	1	0	6
0	0	1	1	54
0	1	1	1	10
1	0	0	0	5
1	0	1	0	1
1	0	1	1	0
1	1	0	0	17
1	1	1	0	4
1	1	1	1	3

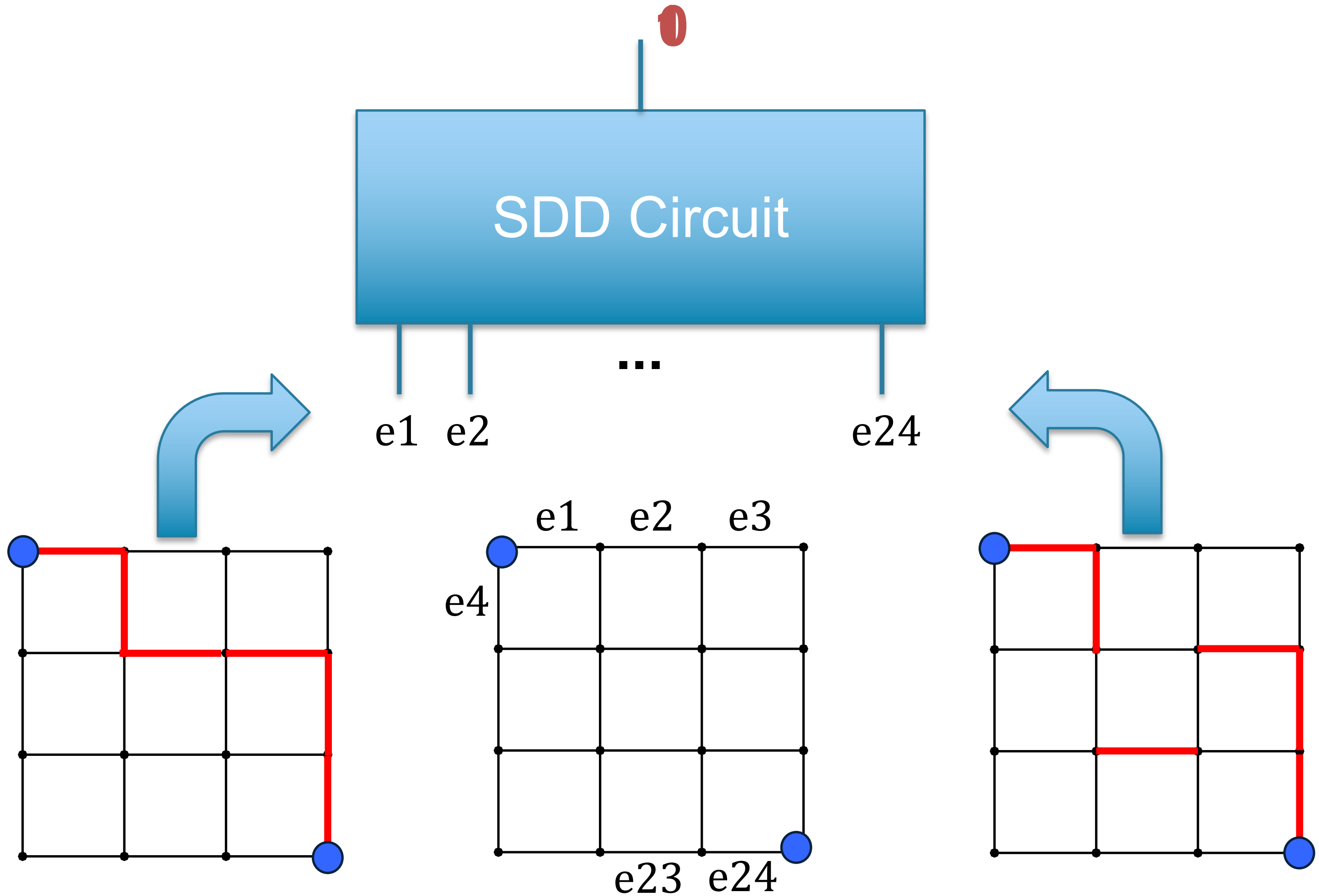


Ignoring Background Knowledge



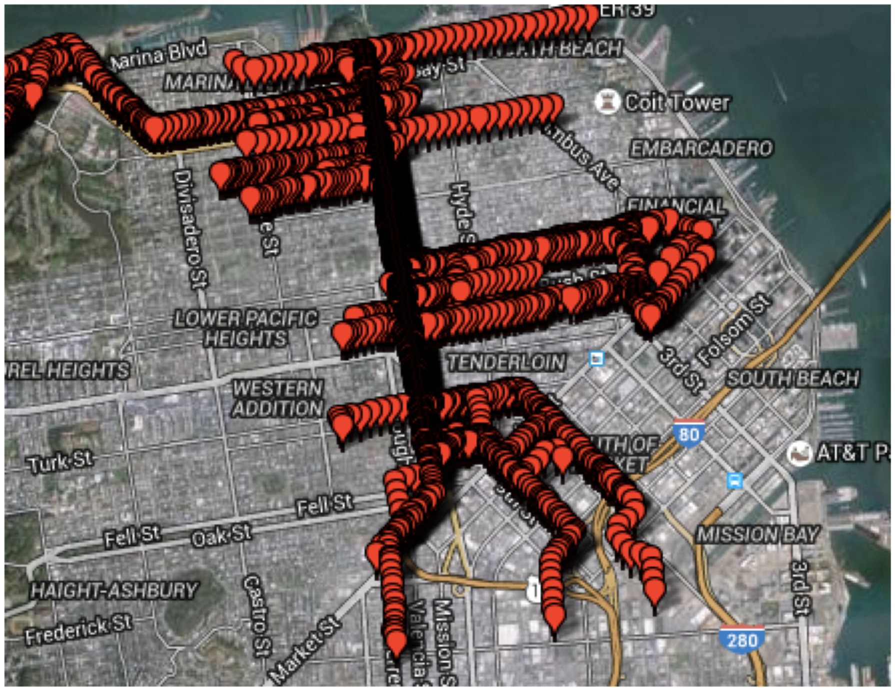
Combinatorial Objects: Routes

Choi et al, AAAI 2016



Combinatorial Objects: Routes

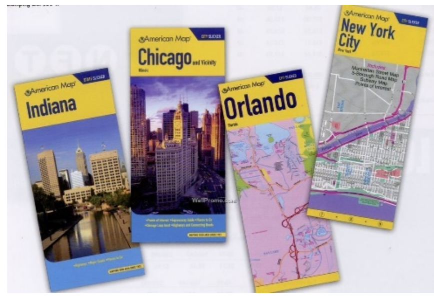
Choi et al, NIPS 2017



- Uber GPS data in SF
- Project GPS coordinates onto a grid/graph, then learn distributions over routes
- Applications:
  - Detect anomalies
  - Given a partial route, predict its most likely completions

Combining Knowledge & Data

Input: Knowledge (a map)



Input: Data (GPS routes)

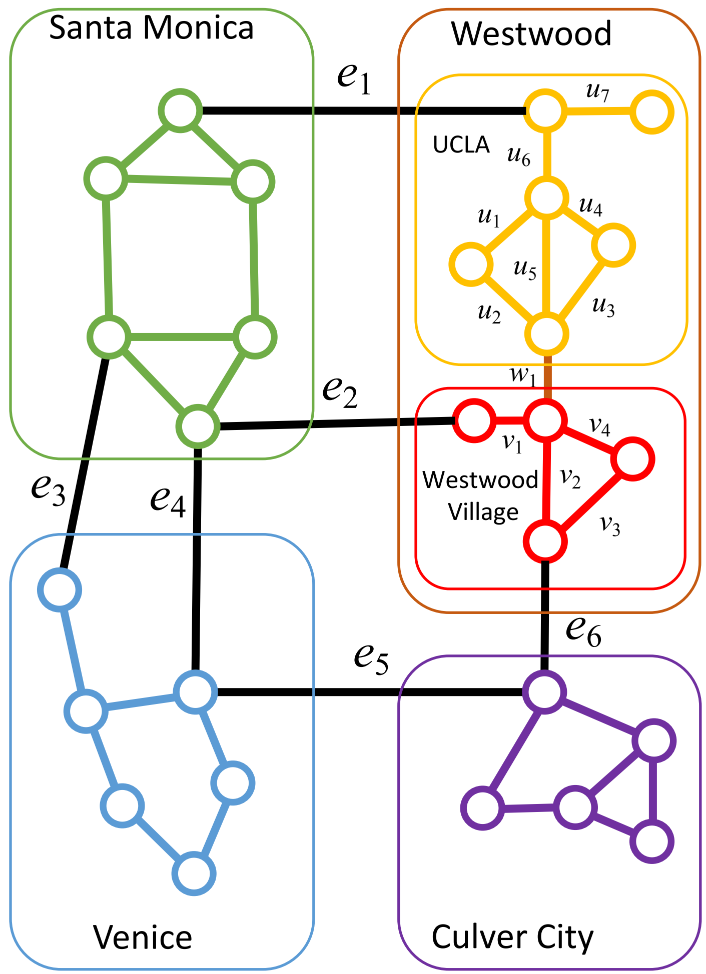
Output: Probabilistic Model over Routes

Estimate traffic
Predict routes
Predict the impact
of an intervention

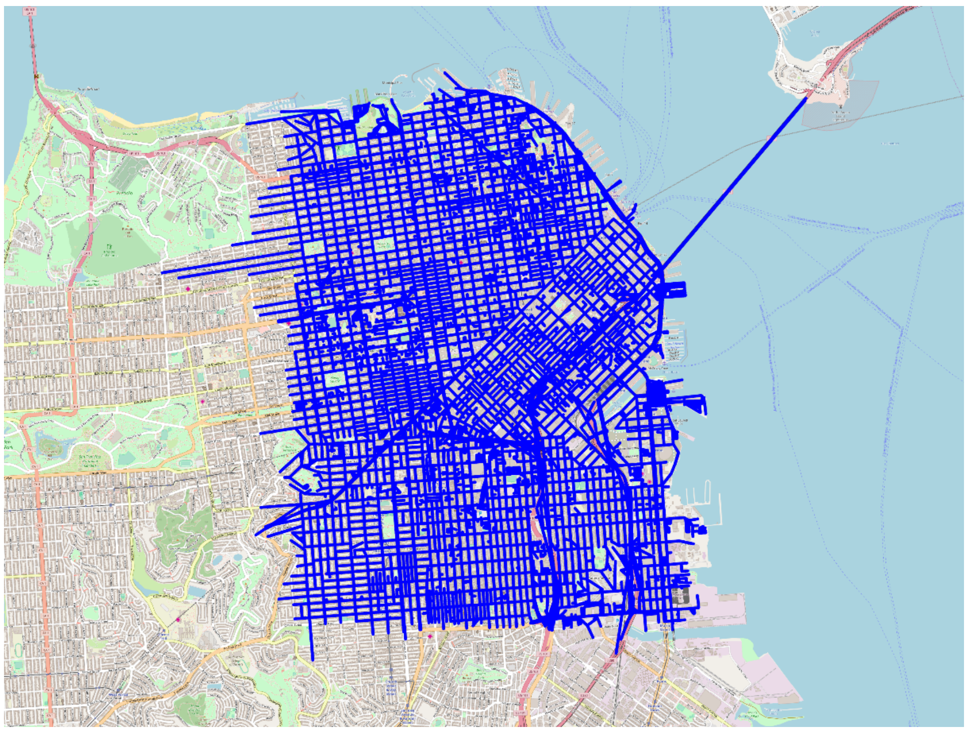
Hierarchical Maps

Choi et al, NIPS 2017

Shen et al, AAAI 2018

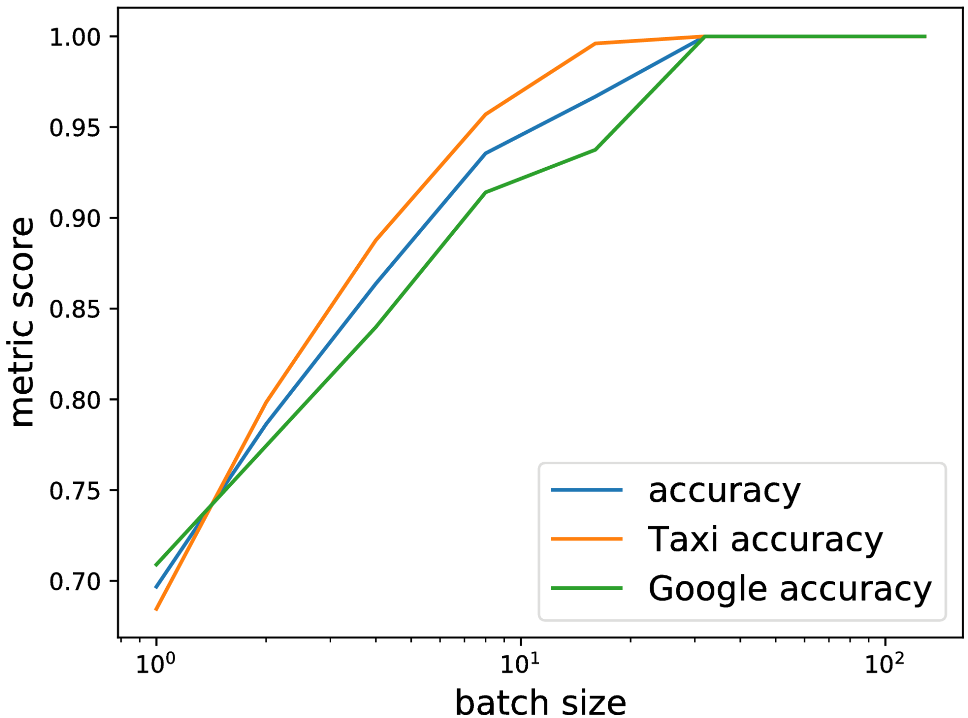


Combinatorial Objects: Routes



- Region of SF with 10,500 edges
- SBN has 1.7M parameters
- **PSDD has 8.9M parameters** (found by PSDD multiply)

Route Classification



- Region of SF with 5,374 edges
- SBN has 611K parameters

- **Taxi dataset:** GPS routes from taxi rides in SF
- **Google dataset:** for each route in taxi dataset: request directions for source, destination, time-of-day and day-of-week
- 172,265 routes (in region)

Combinatorial Objects: Rankings

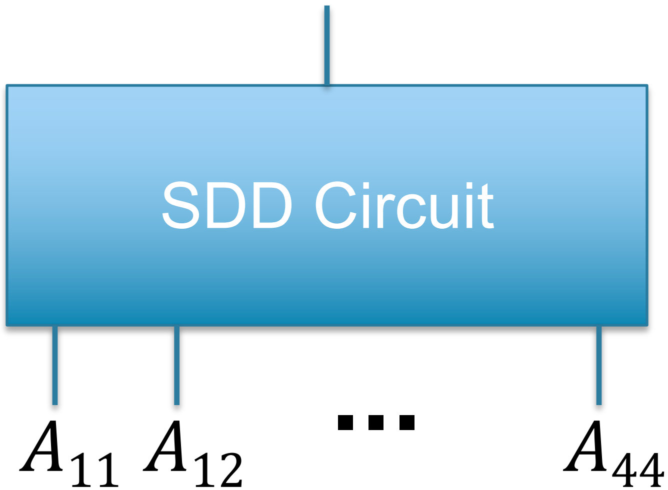
Choi et al, IJCAI 2015

rank	sushi
1	fatty tuna
2	sea urchin
3	salmon roe
4	shrimp
5	tuna
6	squid
7	tuna roll
8	see eel
9	egg
10	cucumber roll

rank	sushi
1	shrimp
2	sea urchin
3	salmon roe
4	fatty tuna
5	tuna
6	squid
7	tuna roll
8	see eel
9	egg
10	cucumber roll

Combinatorial Objects: Rankings

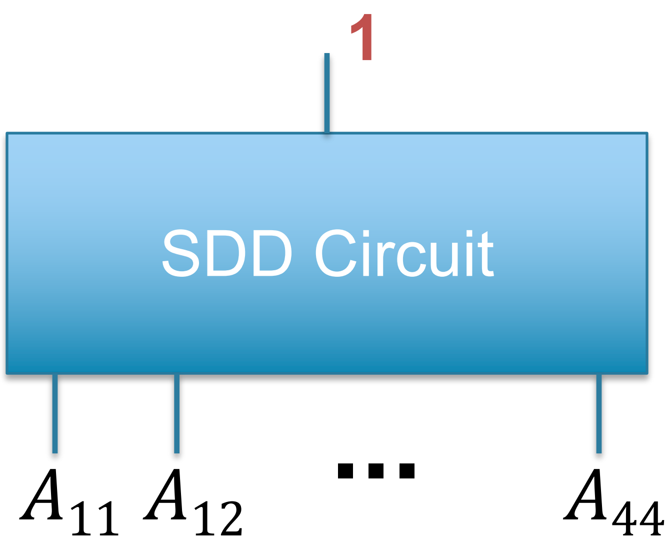
Choi et al, IJCAI 2015



	pos 1	pos 2	pos 3	pos 4
item 1	A_{11}	A_{12}	A_{13}	A_{14}
item 2	A_{21}	A_{22}	A_{23}	A_{24}
item 3	A_{31}	A_{32}	A_{33}	A_{34}
item 4	A_{41}	A_{42}	A_{43}	A_{44}

Combinatorial Objects: Rankings

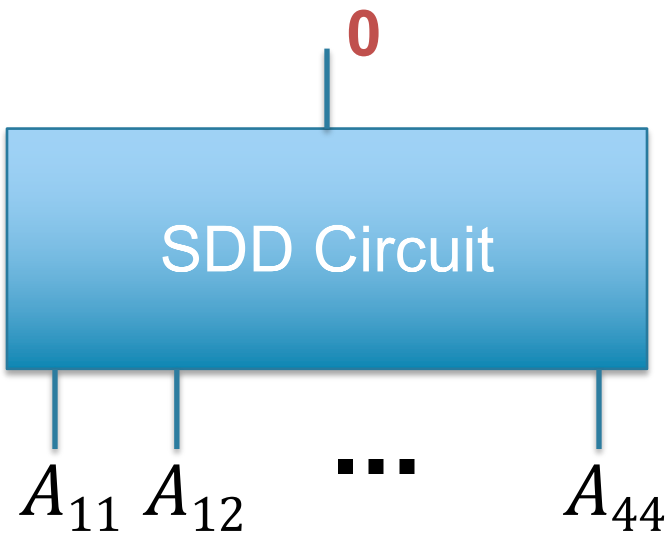
Choi et al, IJCAI 2015



	pos 1	pos 2	pos 3	pos 4
item 1	A_{11}	A_{12}	A_{13}	A_{14}
item 2	A_{21}	A_{22}	A_{23}	A_{24}
item 3	A_{31}	A_{32}	A_{33}	A_{34}
item 4	A_{41}	A_{42}	A_{43}	A_{44}

Combinatorial Objects: Rankings

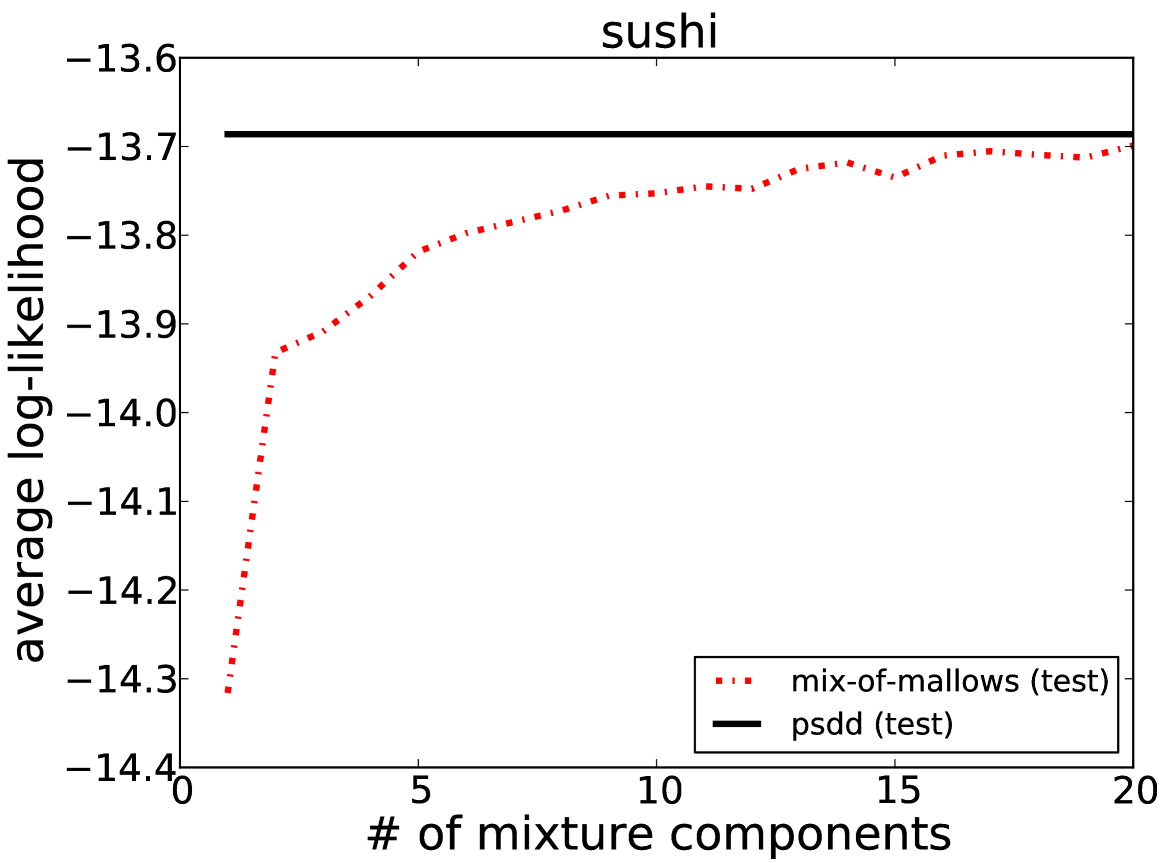
Choi et al, IJCAI 2015



	pos 1	pos 2	pos 3	pos 4
item 1	A_{11}	A_{12}	A_{13}	A_{14}
item 2	A_{21}	A_{22}	A_{23}	A_{24}
item 3	A_{31}	A_{32}	A_{33}	A_{34}
item 4	A_{41}	A_{42}	A_{43}	A_{44}

Learning Distributions over Total Rankings

- training set (3,500)
  testing set (1,500)
- Mixture-of-Mallows
  - # of components from 1 to 20
  - EM with 10 random seeds
  - implementation of Lu & Boutilier



Classical Datasets

a classical
complete dataset

id	X	Y	Z
1	x_1	y_2	z_1
2	x_2	y_1	z_2
3	x_2	y_1	z_2
4	x_1	y_1	z_1
5	x_1	y_2	z_2

a classical
incomplete dataset

id	X	Y	Z
1	x_1	y_2	?
2	x_2	y_1	?
3	?	?	z_2
4	?	y_1	z_1
5	x_1	y_2	z_2

a new type of
incomplete dataset

id	X	Y	Z
1	$X \equiv Z$		
2	x_2 and (y_2 or z_2)		
3	$x_2 \Rightarrow y_1$		
4	$X \oplus Y \oplus Z \equiv 1$		
5	x_1 and y_2 and z_2		

Missed in the
ML literature

Classical Datasets

id	1st sushi	2nd sushi	3rd sushi	…
1	fatty tuna	sea urchin	salmon roe	…
2	fatty tuna	tuna	shrimp	…
3	tuna	tuna roll	sea eel	…
4	fatty tuna	salmon roe	tuna	…
5	egg	squid	shrimp	…

id	1st sushi	2nd sushi	3rd sushi	…
1	fatty tuna	sea urchin	?	…
2	fatty tuna	?	?	…
3	tuna	tuna roll	?	…
4	fatty tuna	salmon roe	?	…
5	egg	?	?	…

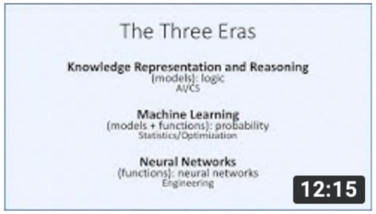
Structured Datasets

id	1^{st} sushi	2^{nd} sushi	3^{rd} sushi	…
1	fatty tuna	sea urchin	salmon roe	…
2	fatty tuna	tuna	shrimp	…
3	tuna	tuna roll	sea eel	…
4	fatty tuna	salmon roe	tuna	…
5	egg	squid	shrimp	…

id	1^{st} sushi	2^{nd} sushi	3^{rd} sushi	…
1	(fatty tuna > sea urchin) and (tuna > sea eel)			…
2	(fatty tuna is 1^{st}) and (salmon roe > egg)			…
3	tuna > squid			…
4	egg is last			…
5	egg > squid > shrimp			…

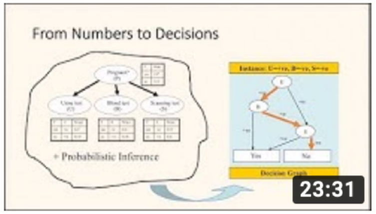
CACM Oct. 2018 - Human-Level Intelligence or Animal...

Association for Computing Ma...
2.7K views • 4 months ago



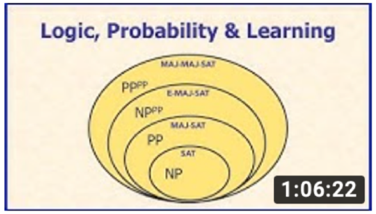
Adnan Darwiche – On AI Education

UCLA Automated Reasoning G...
1.4K views • 5 months ago



Adnan Darwiche – Explaining and Verifying AI Systems

UCLA Automated Reasoning G...
464 views • 4 months ago



Adnan Darwiche – On the Role of Logic in Probabilisti...

UCLA Automated Reasoning G...
2K views • 1 year ago

Adnan Darwiche – On Model-Based versus Model-Blind...

UCR School of Public Policy
3.3K views • 1 year ago

Learning and Reasoning with Bayesian Networks ▶ PLAY ALL

Lectures by Adnan Darwiche for his UCLA course on Bayesian Networks.

Subscribe!

1a. Course Overview with a Historical Perspective on AI

UCLA Automated Reasoning Group • 2.2K views • 1 year ago

Adnan Darwiche's UCLA course: Learning and Reasoning with Bayesian Networks.

Thank You