On the Role of Logic

in Probabilistic Inference & Machine Learning

Adnan Darwiche
Computer Science Department
UCLA

December, 2017

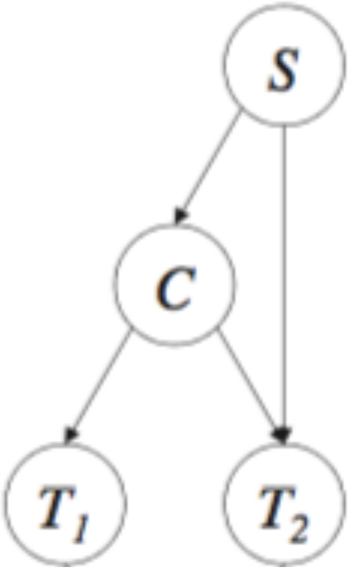
Agenda

- Logic in Probabilistic Inference
  - Four probabilistic queries
  - Beyond NP: PP, NP^{PP} & PP^{PP}
- Logic in Machine Learning
  - Beyond data: Background knowledge
  - Beyond classical datasets

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



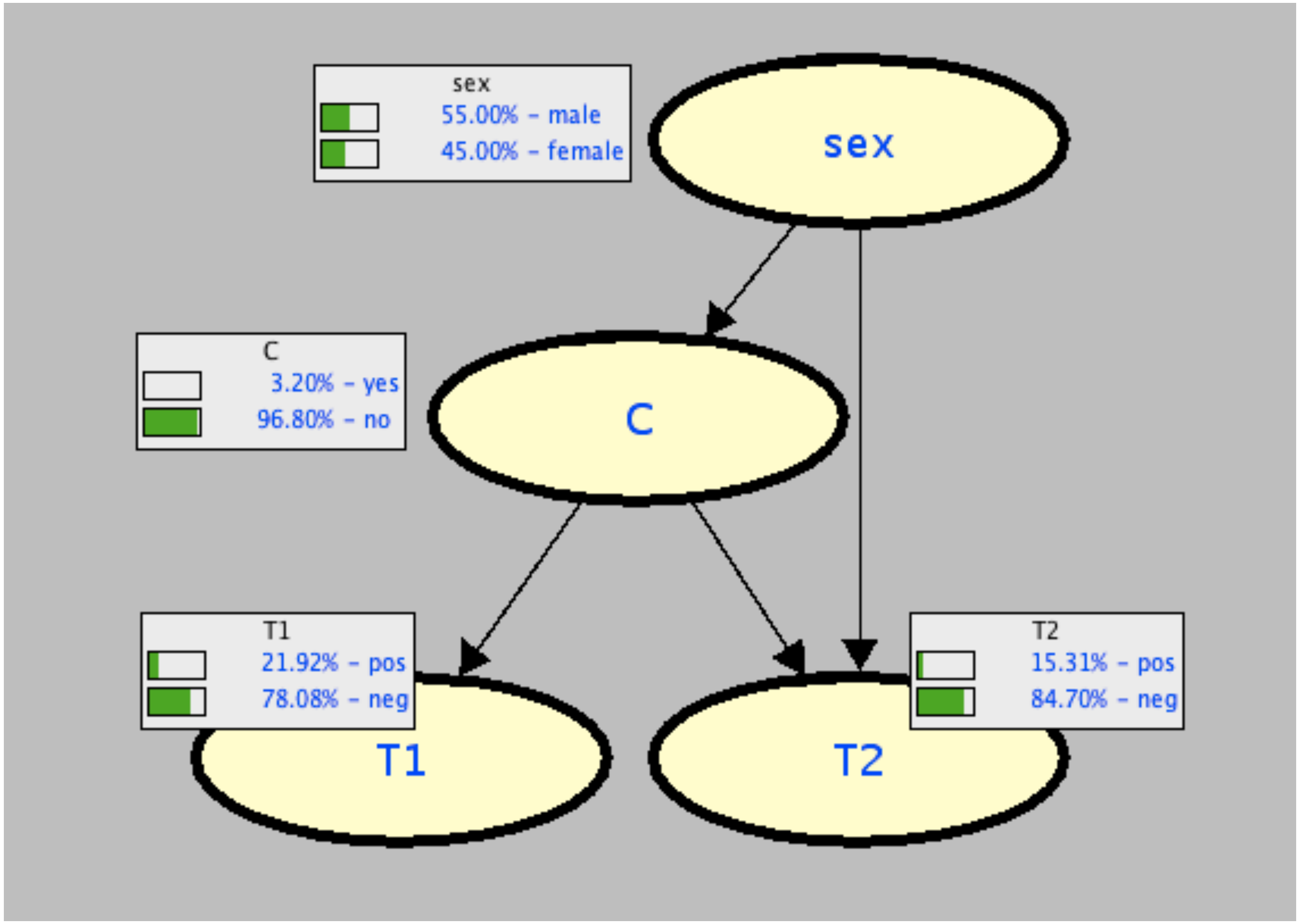
S	θ_s
male	.55
female	.45

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

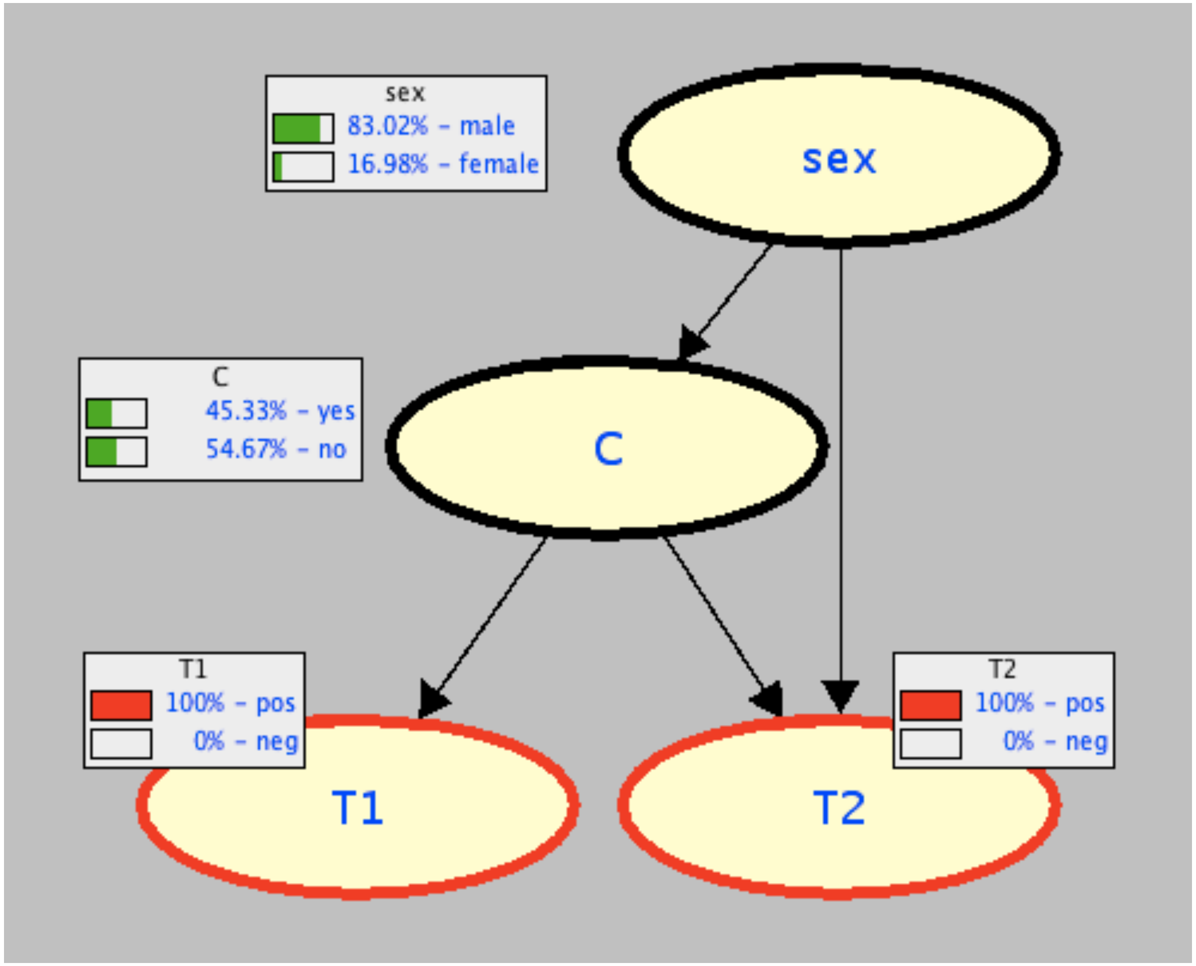
C	T_1	$\theta_{t_1\|c}$	
yes	+ve	.80	
yes	−ve	.20	= fn
no	+ve	.20	= fp
no	−ve	.80	

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	

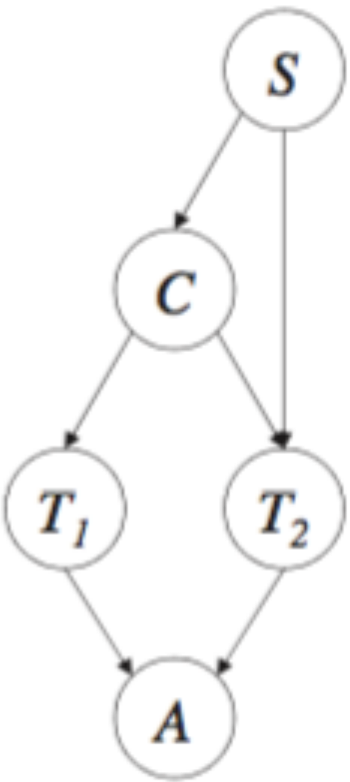
Prior Marginals



Posterior Marginals



Bayesian Network



S	θ_s
male	.55
female	.45

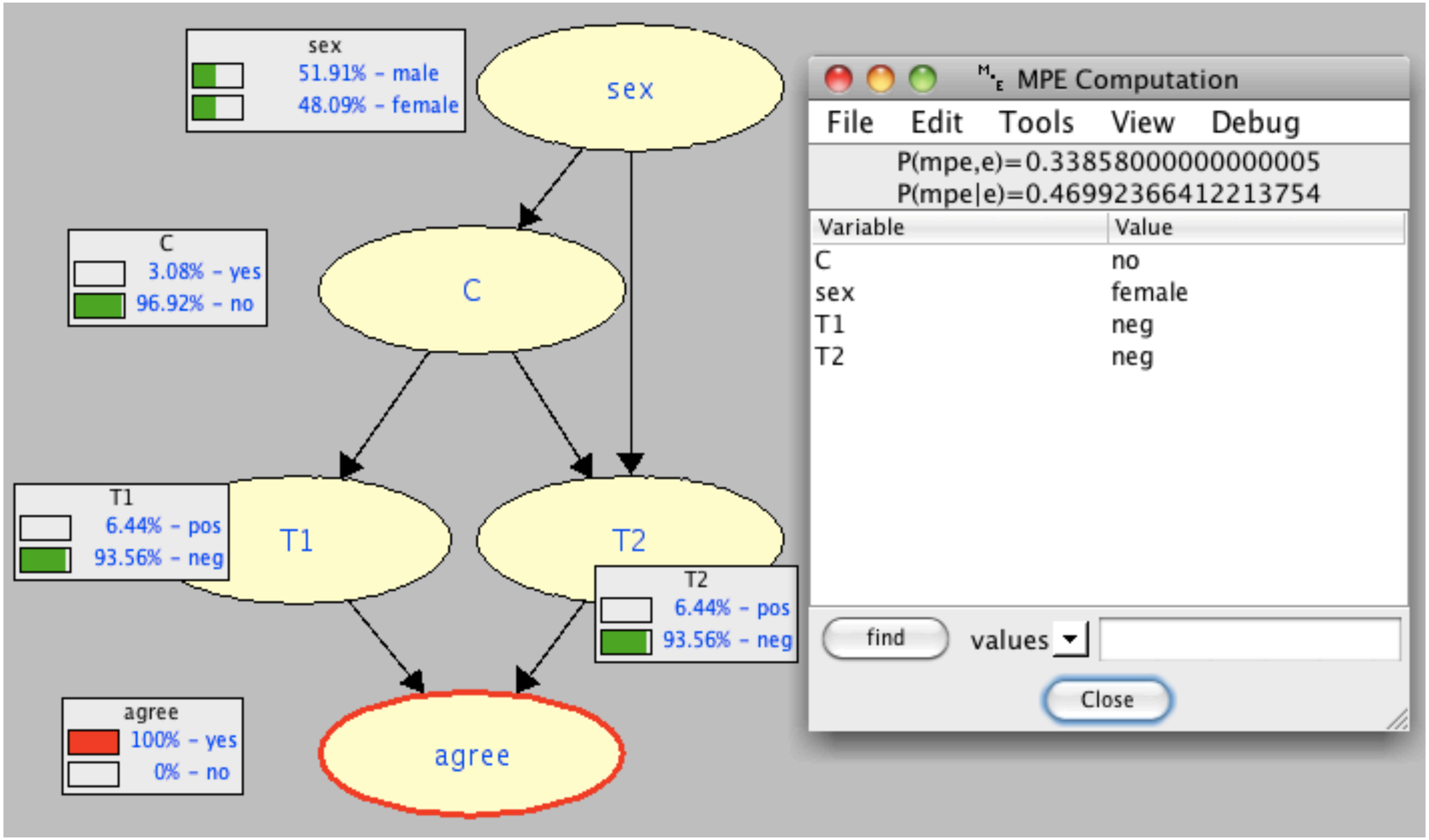
S	C	$\theta_{c\|s}$
male	yes	.05
male	no	.95
female	yes	.01
female	no	.99

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

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

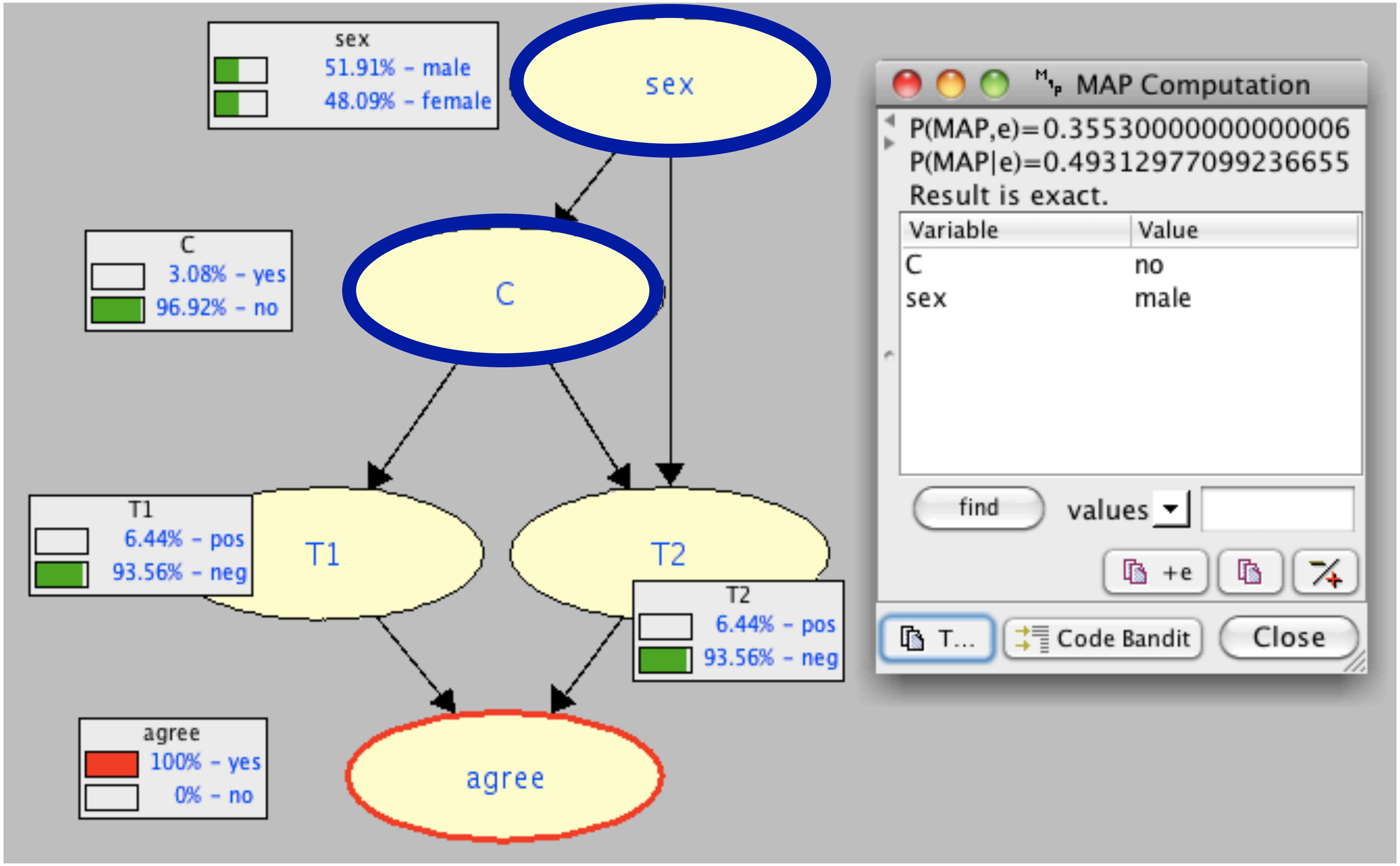
T_1	T_2	A	$\theta_{a\|t_1,t_2}$
+ve	+ve	yes	1
+ve	+ve	no	0
+ve	−ve	yes	0
+ve	−ve	no	1
−ve	+ve	yes	0
−ve	+ve	no	1
−ve	−ve	yes	1
−ve	−ve	no	0

MPE



sex
51.91% – male
48.09% – female
sex
C
3.08% – yes
96.92% – no
C
T1
6.44% – pos
93.56% – neg
T1
T2
T2
6.44% – pos
93.56% – neg
agree
100% – yes
0% – no
agree
MPE Computation
File Edit Tools View Debug
P(mpe,e)=0.33858000000000005
P(mpe|e)=0.46992366412213754
Variable Value
C no
sex female
T1 neg
T2 neg
find
values
Close

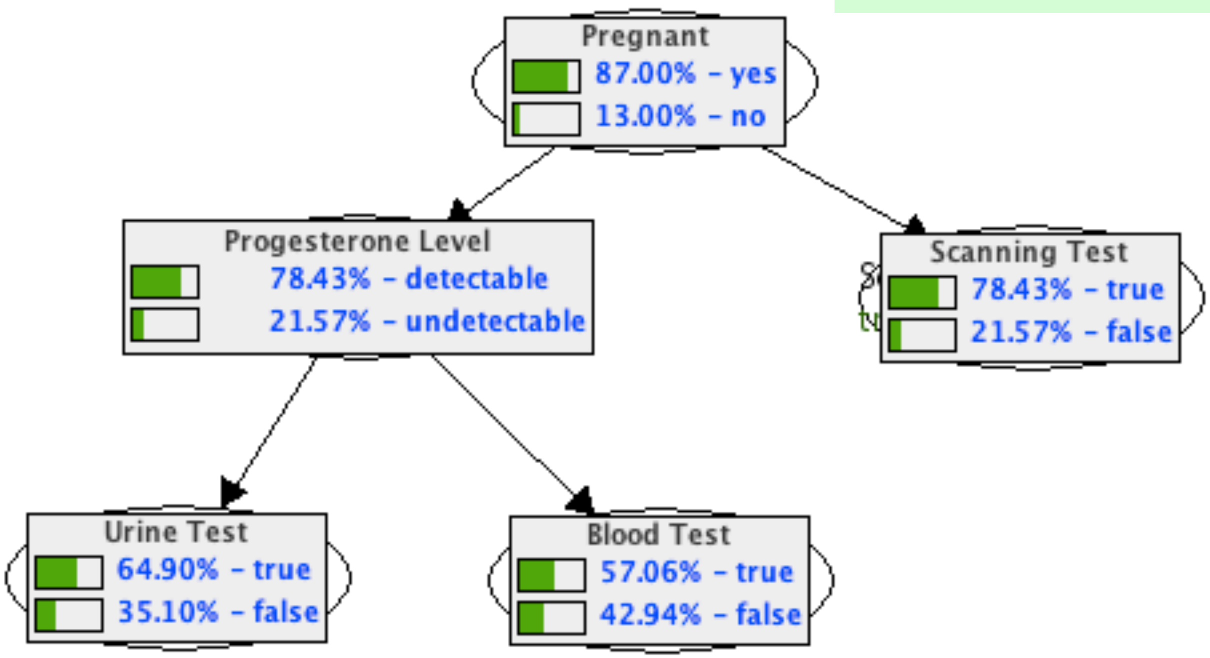
MAP



sex
51.91% - male
48.09% - female
sex
C
3.08% - yes
96.92% - no
C
T1
6.44% - pos
93.56% - neg
T1
T2
T2
6.44% - pos
93.56% - neg
agree
100% - yes
0% - no
agree
MAP Computation
P(MAP,e)=0.35530000000000006
P(MAP|e)=0.49312977099236655
Result is exact.
Variable
Value
C
no
sex
male
find
values
+e
T...
Code Bandit
Close

Same-Decision Probability

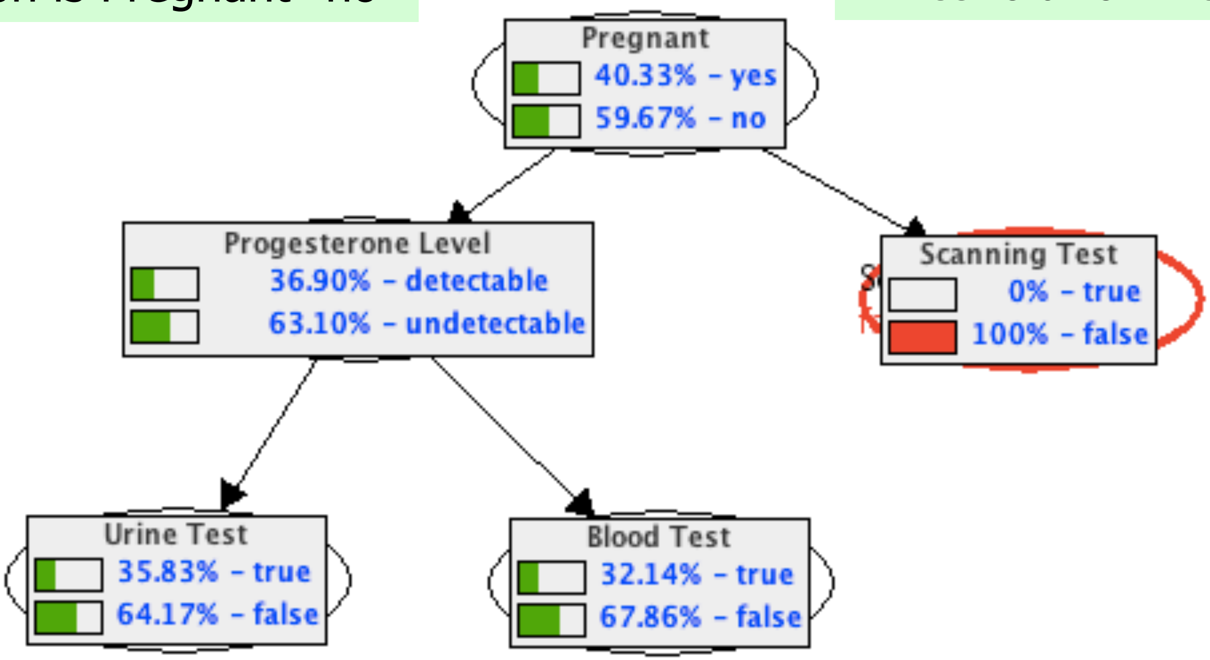
Threshold for Pregnancy is 90%



Same-Decision Probability

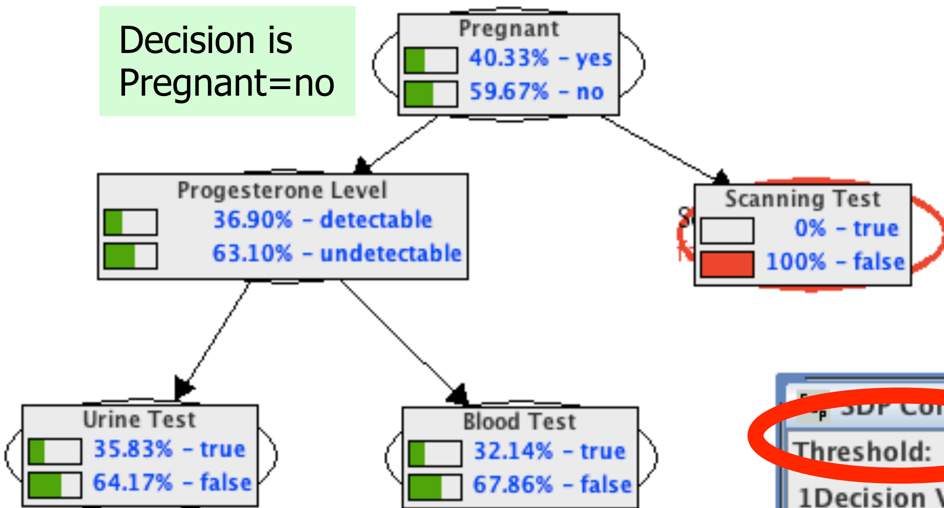
Decision is Pregnant=no

Threshold for Pregnancy is 90%

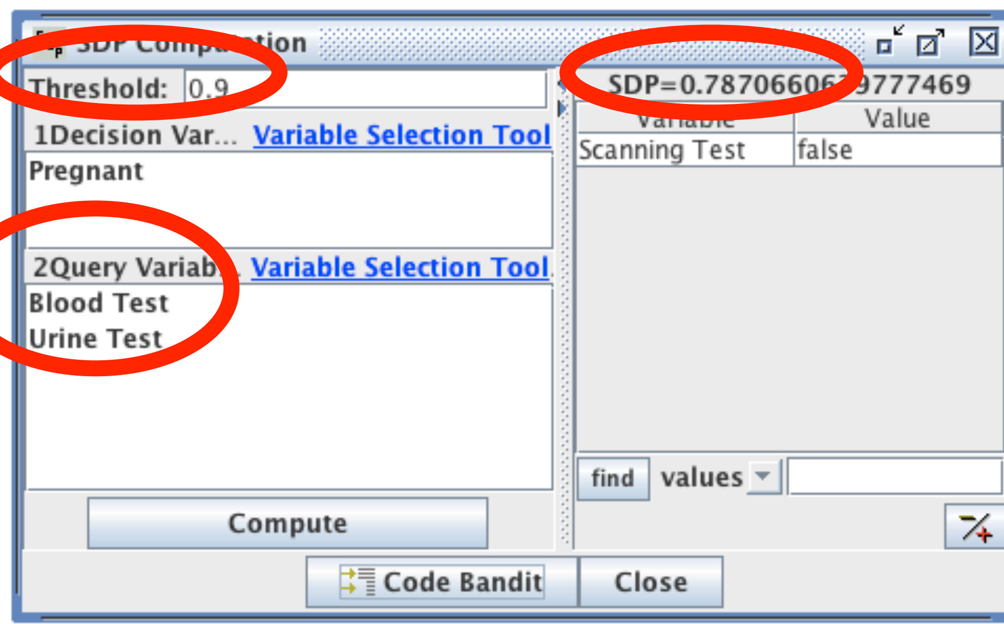


Same-Decision Probability

Decision is Pregnant=no



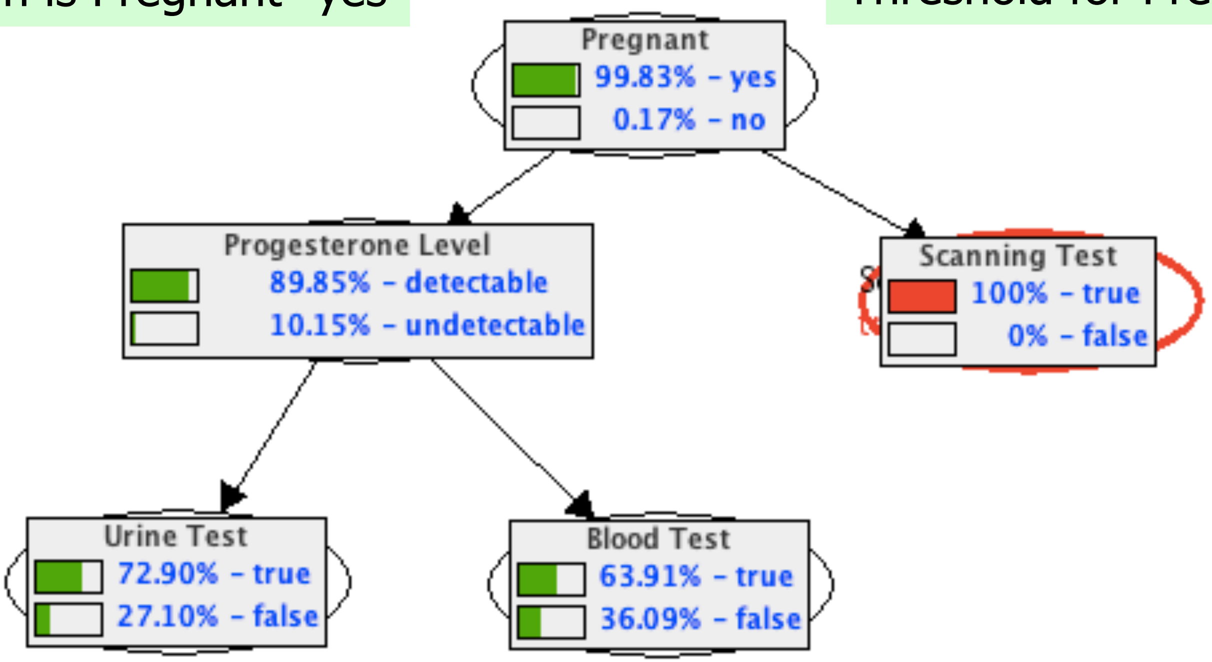
78.7% chance you will still make the same decision after collecting the blood and urine tests.



Same-Decision Probability

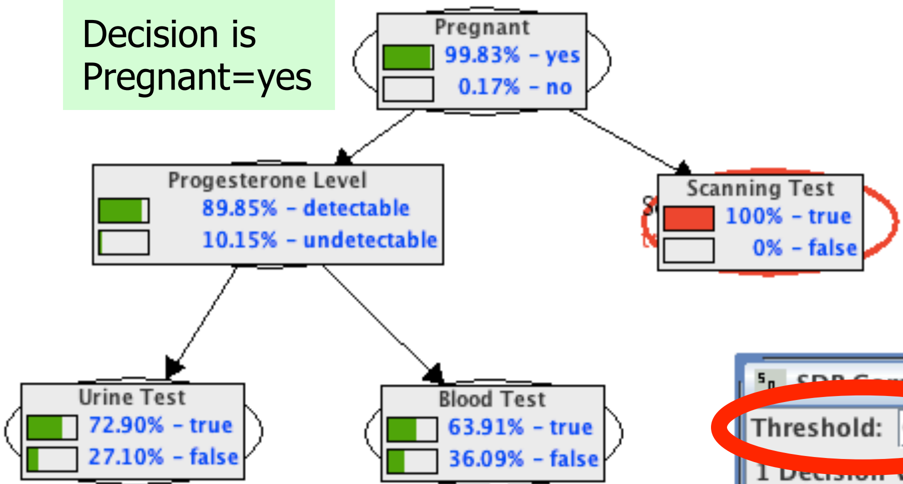
Decision is Pregnant=yes

Threshold for Pregnancy is 90%

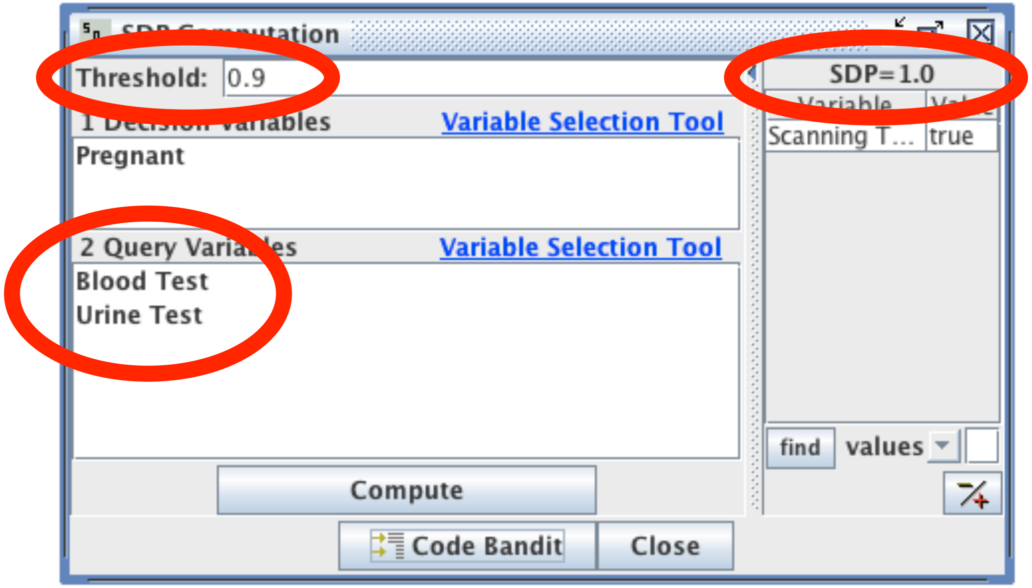


Same-Decision Probability

Decision is Pregnant=yes



100% chance you will still make the same decision after collecting the blood and urine tests.

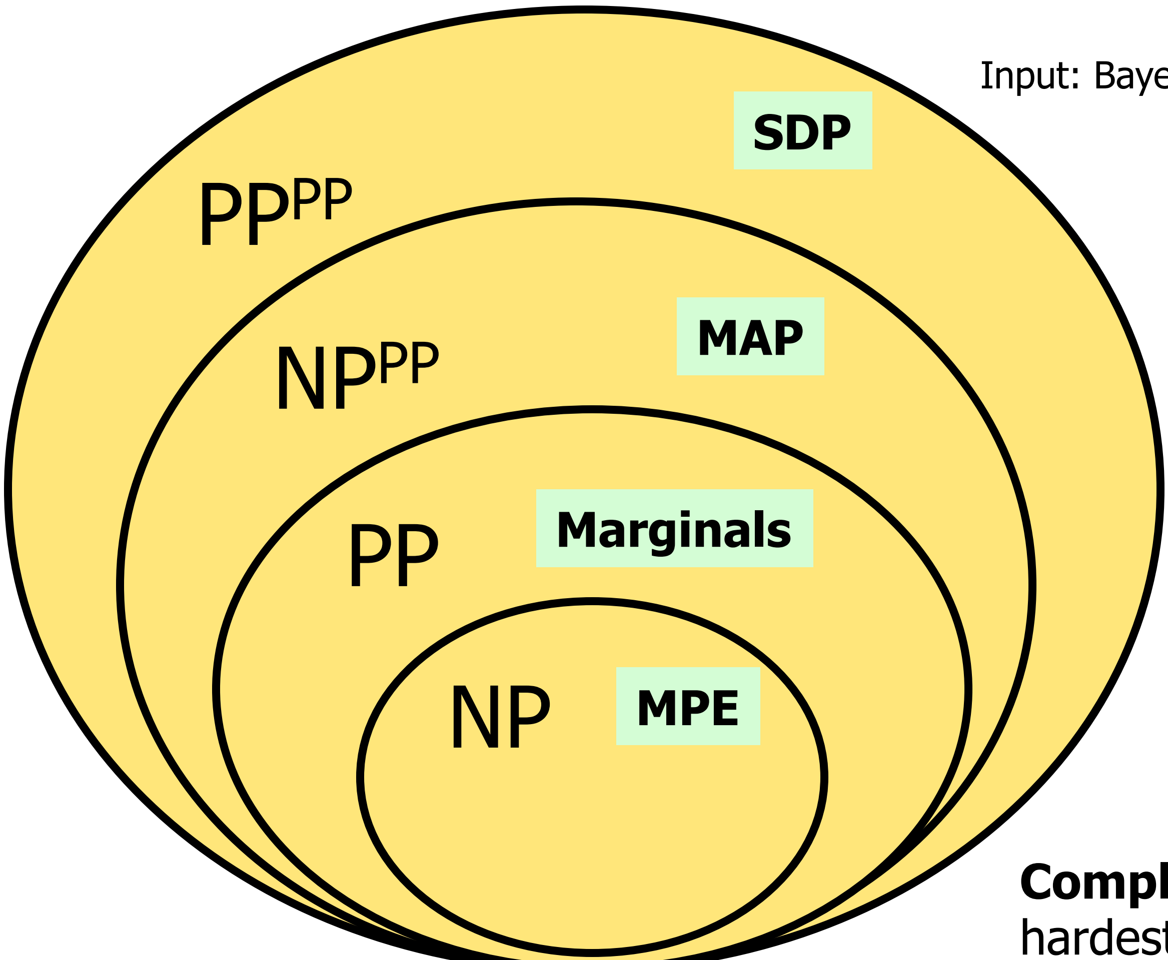


Same-Decision Probability (SDP)

Darwiche & Choi, PGM 2010

probability that a decision will stick after observing some additional variables.

Probabilistic Inference



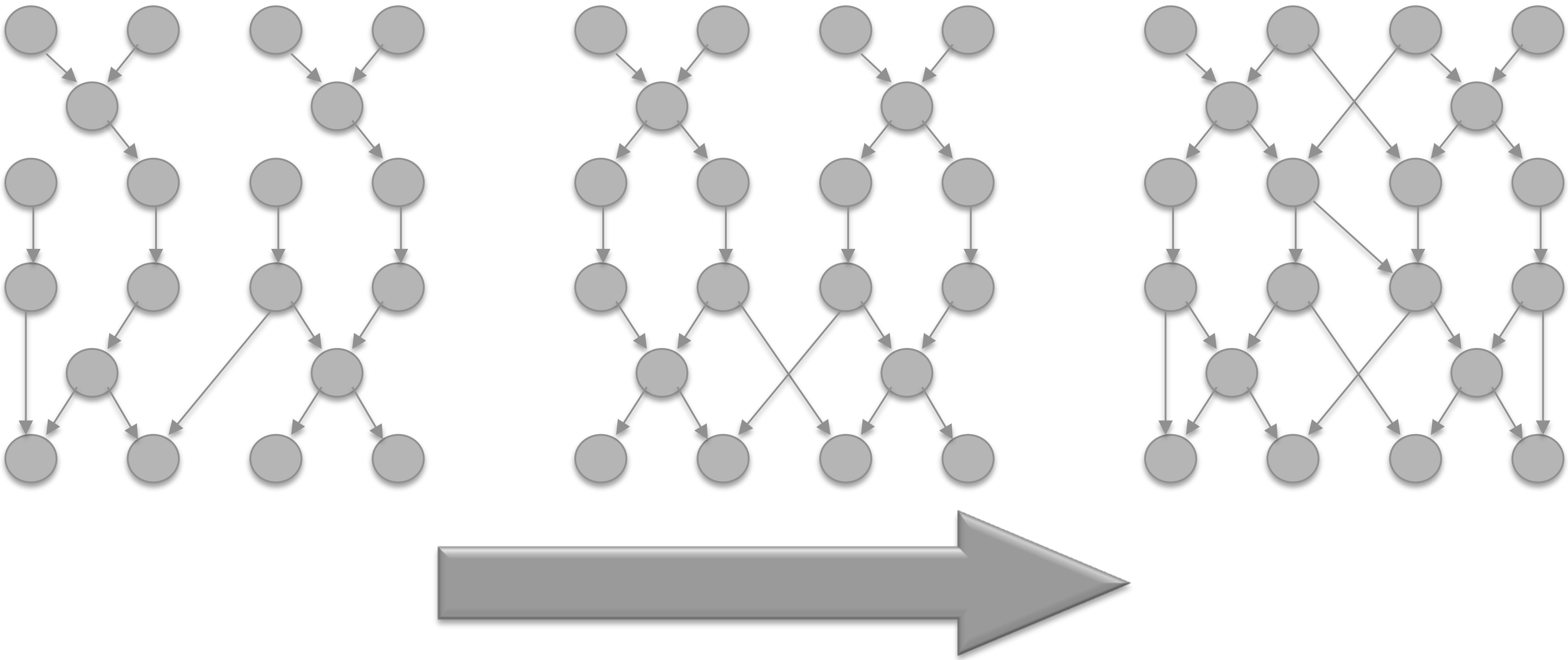
Complete problems:
hardest in their class

Why Not Classical Approaches?

- **Exact Inference:**
  Jointree, Variable Elimination, Conditioning

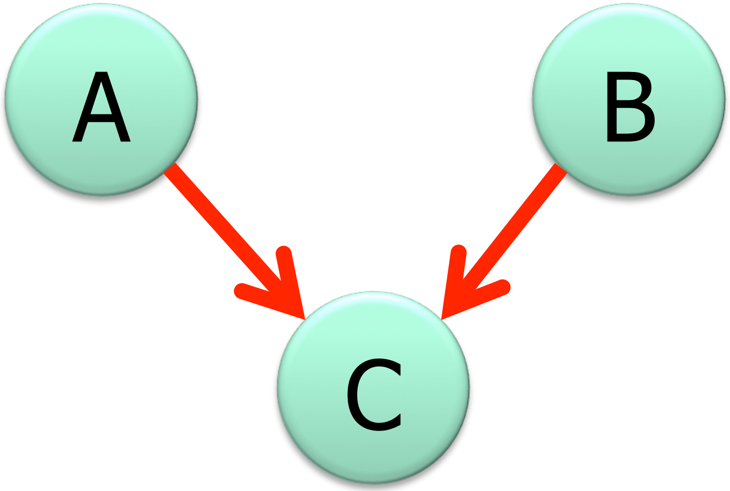
- **Approximate Inference:**
  BP, MCMC, Variational

Barrier: Treewidth

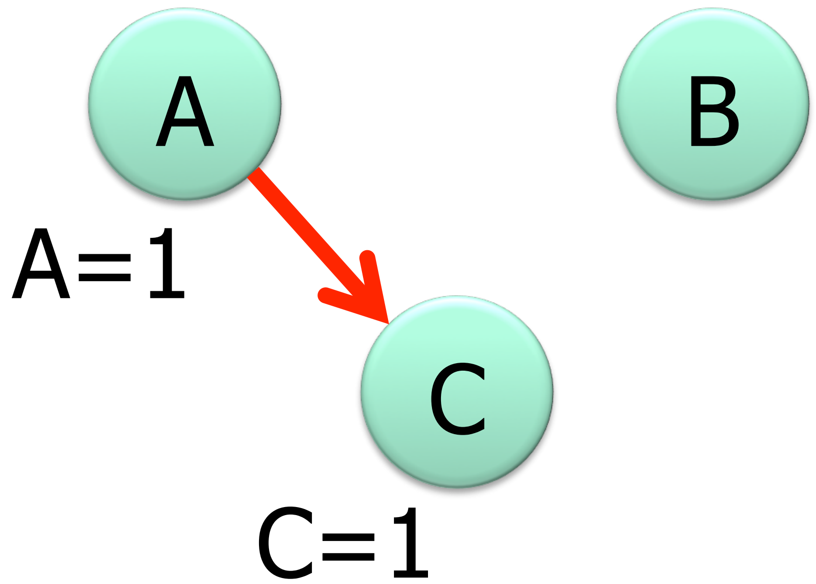


$$O(n \exp\{w\})$$

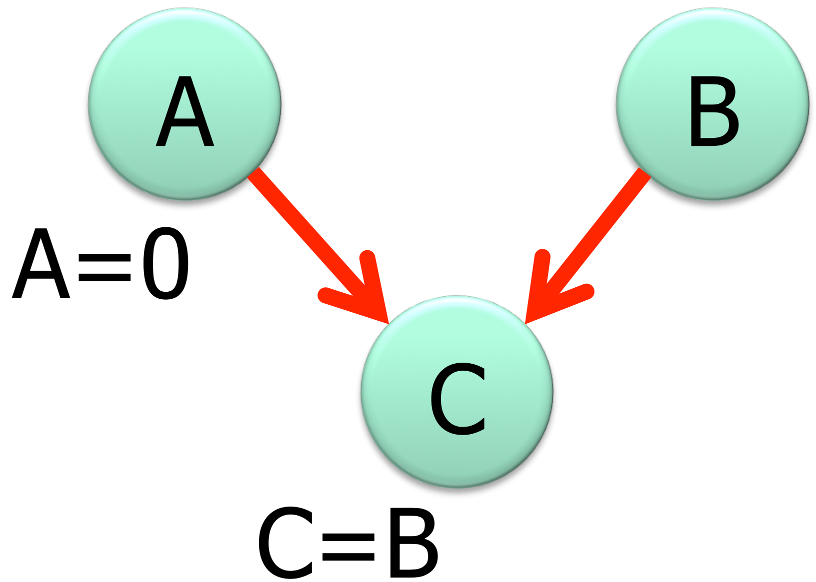
Context-Specific Independence



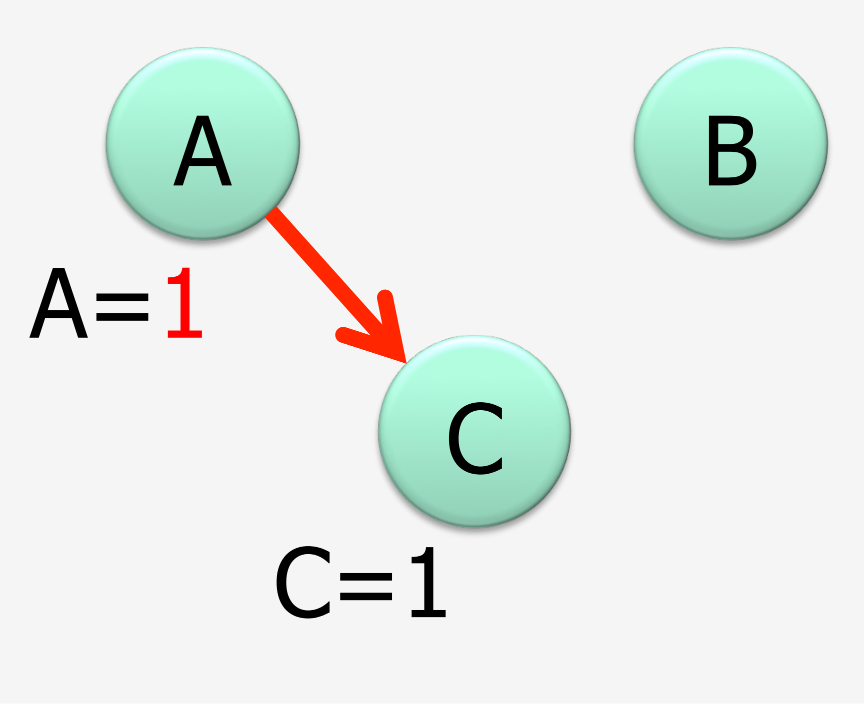
Context-Specific Independence

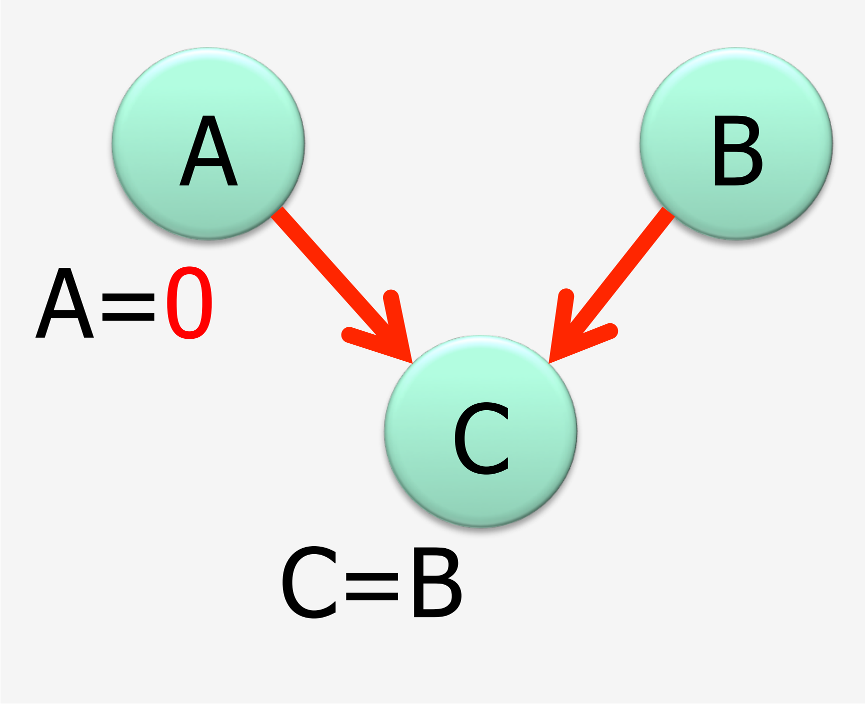


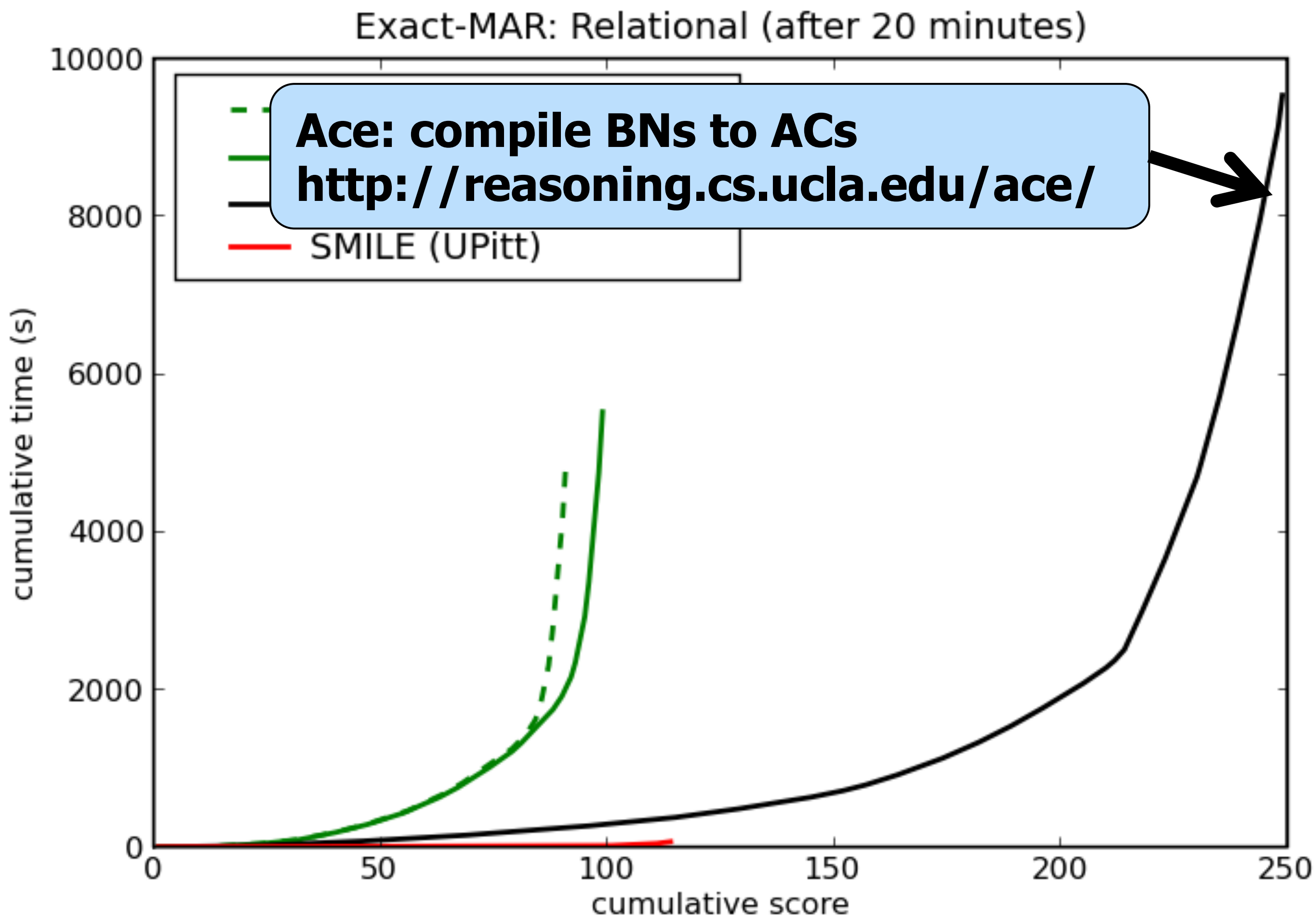
Context-Specific Independence



Context-Specific Independence





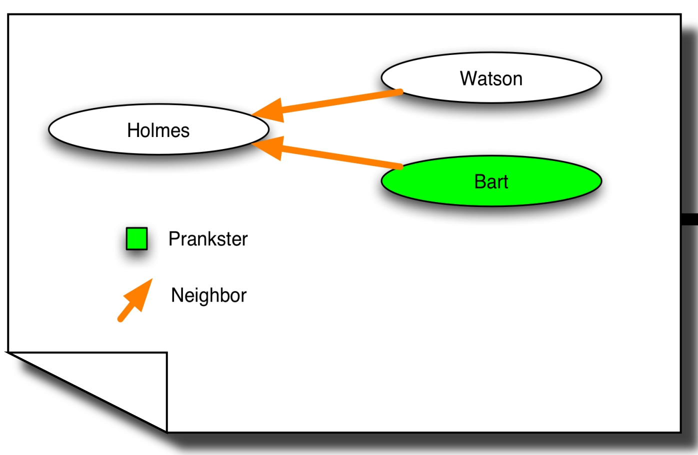


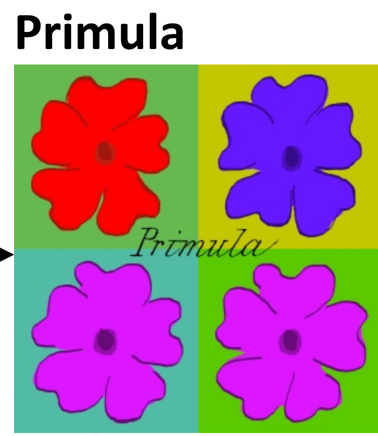
- Relational networks (251 networks)
    - Average cluster size is 50

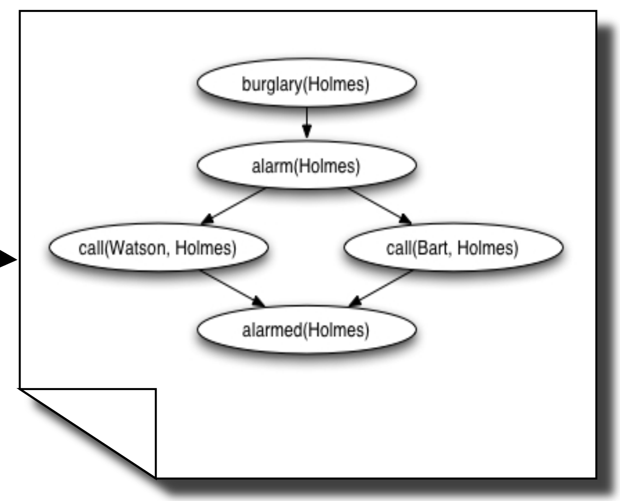
Reducing
Relational Probabilistic Models

Chavira et al, IJAR 2006

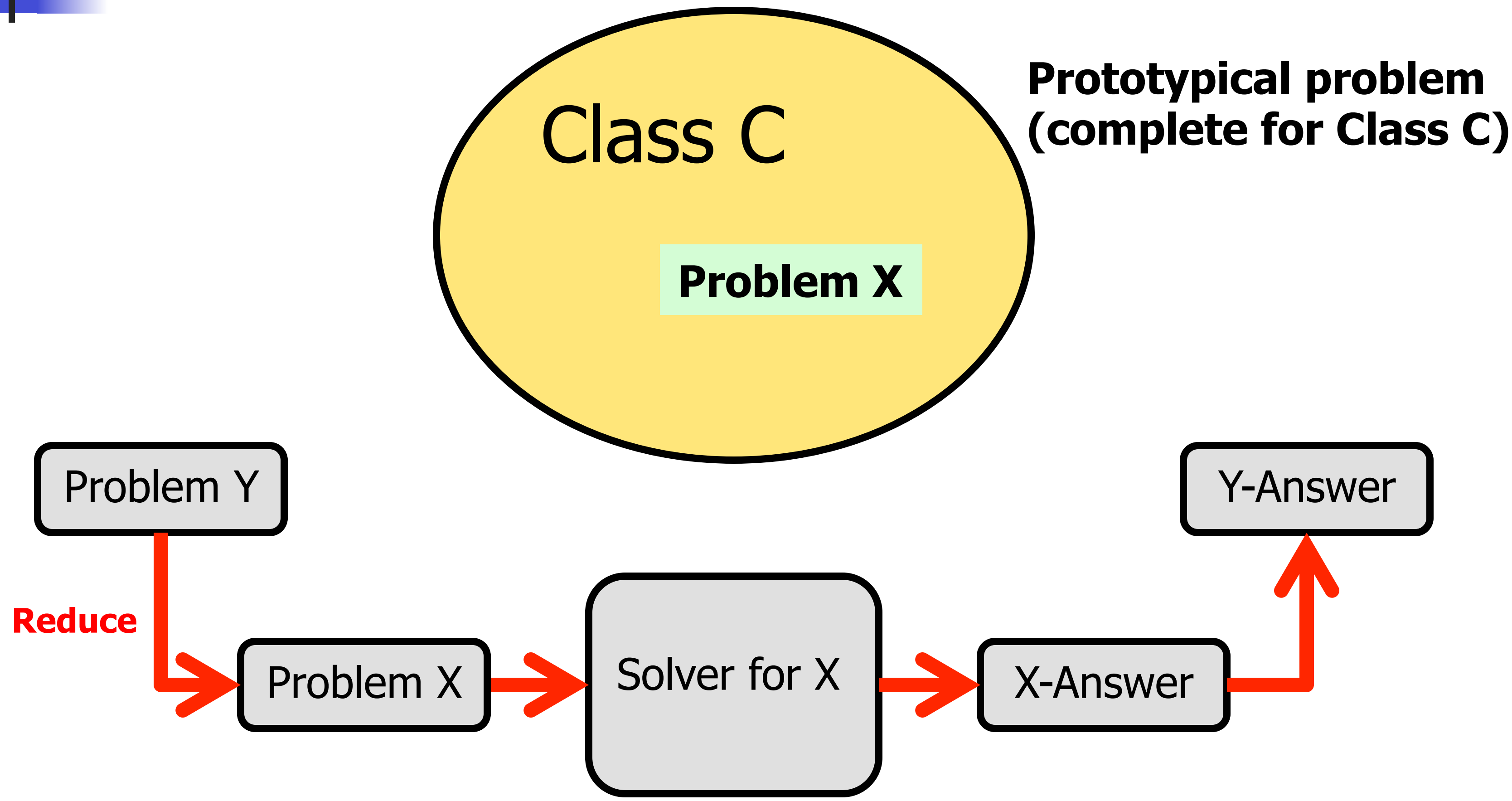
```
burglary(v)=0.005;
alarm(v)=(burglary(v):0.95,0.01);
calls(v,w)=
  (neighbor(v,w):
   (prankster(v)):
    (alarm(w):0.9,0.05),
    (alarm(w):0.9,0)),0);
alarmed(v)=
  n-or{calls(w,v)|w:neighbor(w,v)}
```



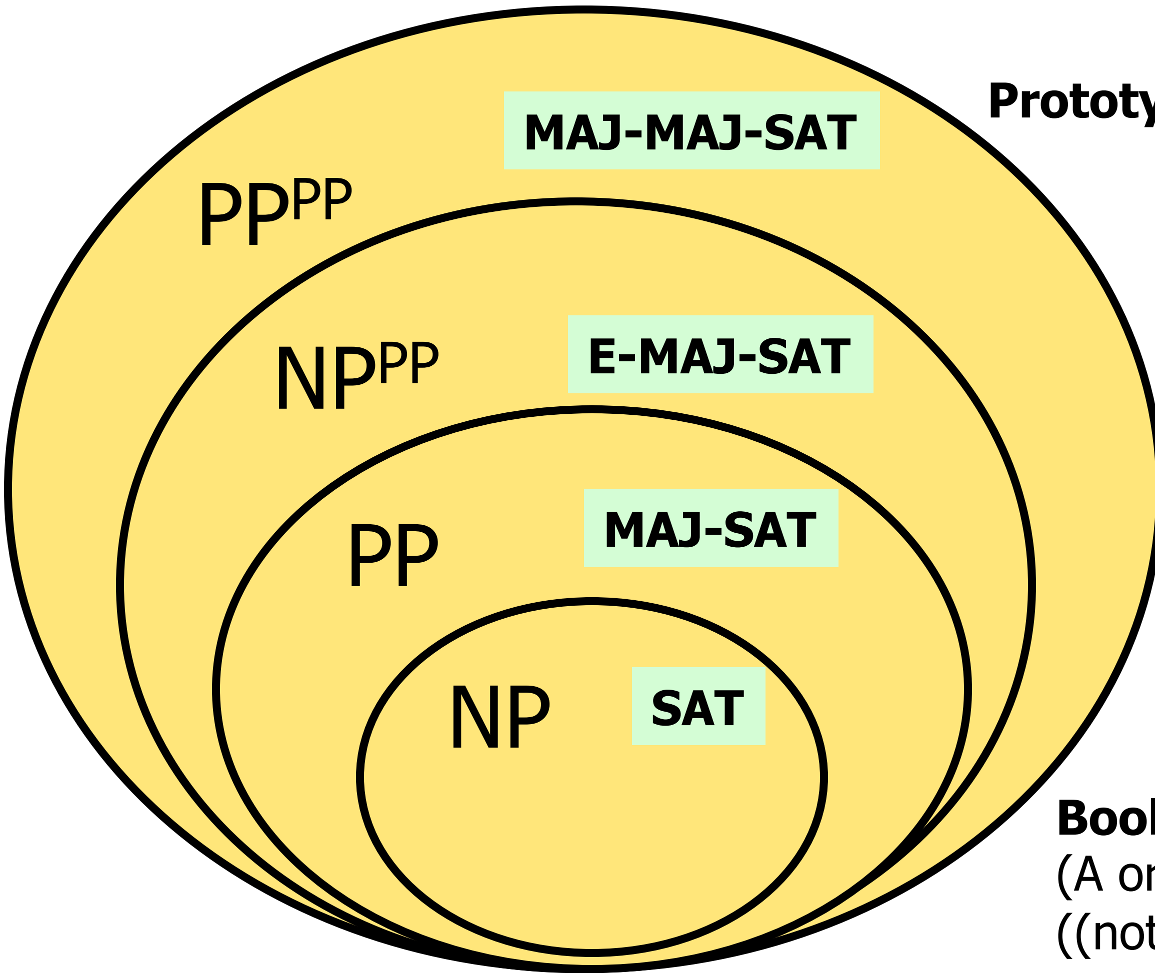




Reduction Approaches



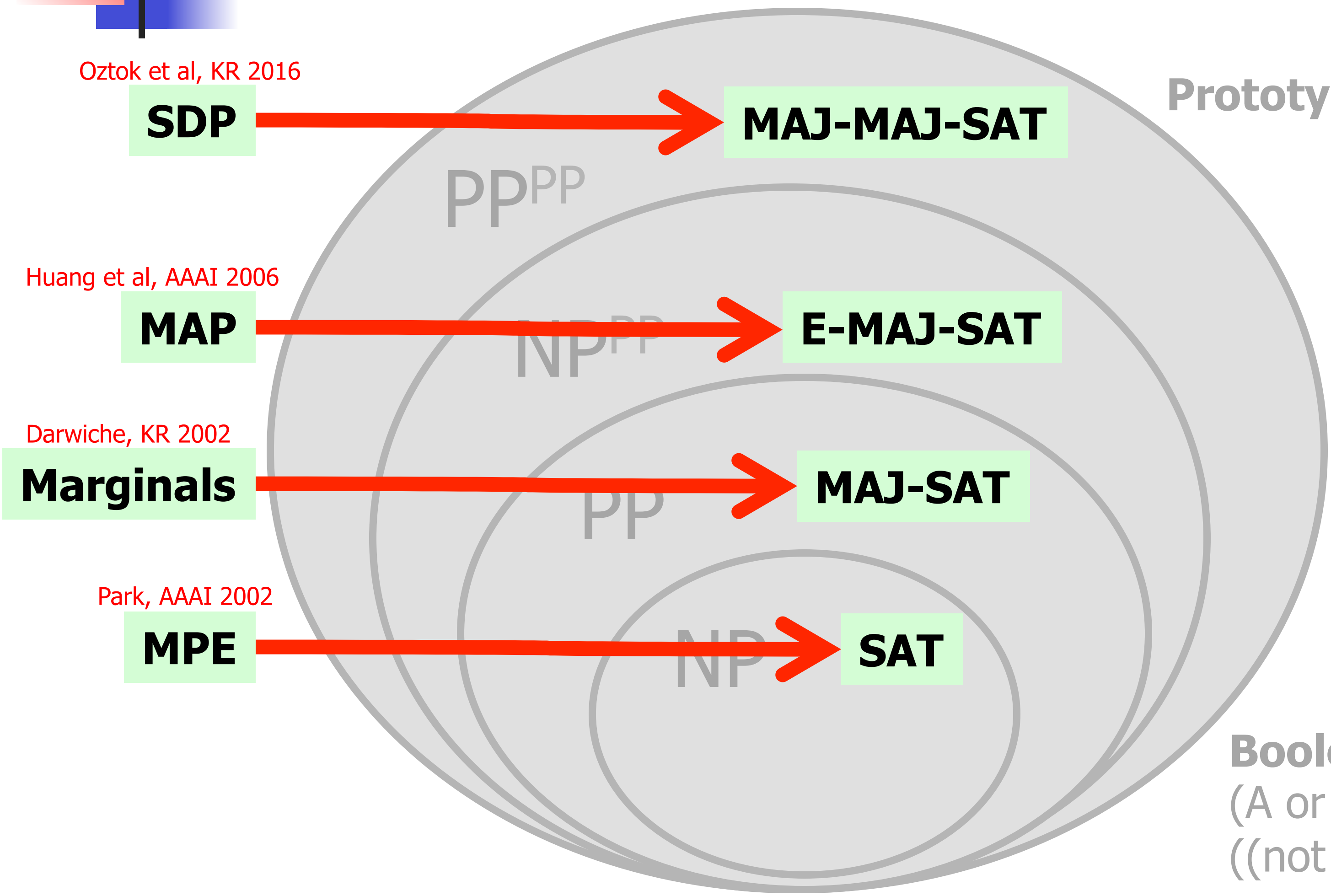
Reduction Approaches



Prototypical problems

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

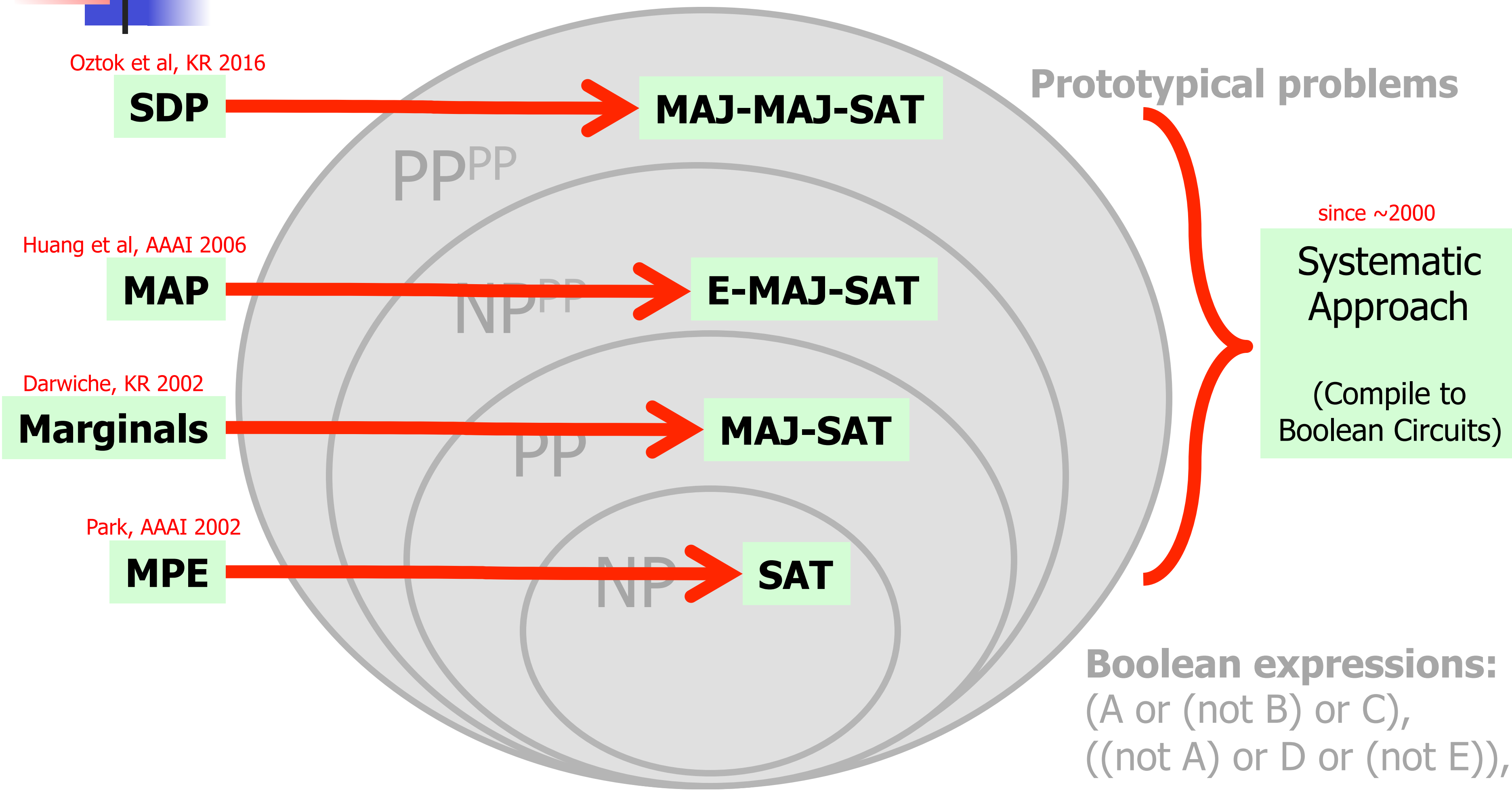
1st Line of Developments



Prototypical problems

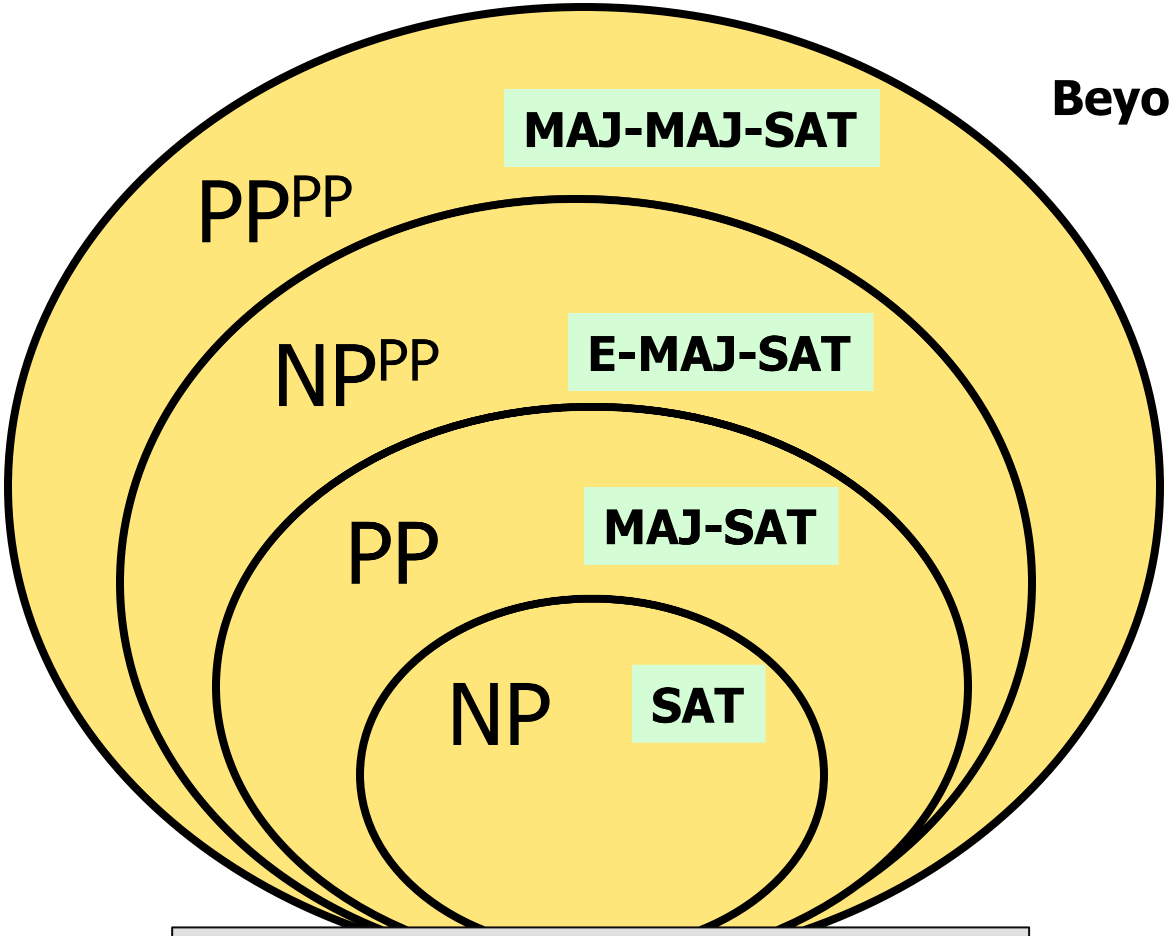
Boolean expressions:
(A or (not B) or C),
((not A) or D or (not E)),
....

2^{nd} Line of Developments



Beyond NP Initiative

BeyondNP.org



1st Workshop on Beyond NP: AAAI'16

Why Reductions?

- **Independent of representations**
  (Bayesian networks, Markov networks, Relational Probabilistic Models, Probabilistic Programs, …)
- **Effective**
  (more later)
- **Systematic and Streamlined**

From the mission statement of BeyondNP

> A new computational paradigm has emerged in computer science over the past few decades, which is exemplified by the use of SAT solvers to tackle problems in the complexity class NP. According to this paradigm, a significant research and engineering investment is made towards developing highly efficient solvers for a prototypical problem (e.g., SAT), that is representative of a broader class of problems (e.g., NP). The cost of this investment is then amortized as these solvers are applied to a broader class of problems via reductions (in contrast to developing dedicated algorithms for each encountered problem). SAT solvers, for example, are now routinely used to solve problems in many domains, including diagnosis, planning, software and hardware verification.

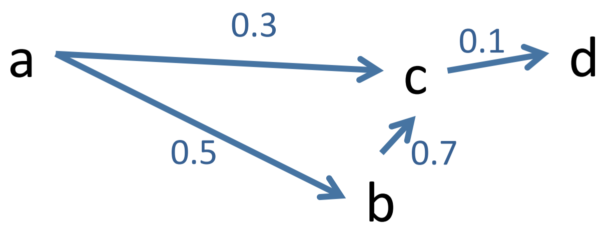
Probabilistic Logic Programming

Fierens et al., TPLP 2015, Vlasselaer et al., IJCAI 2015

Example: Discrete probabilistic reachability program

Logic Program (ProbLog)

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



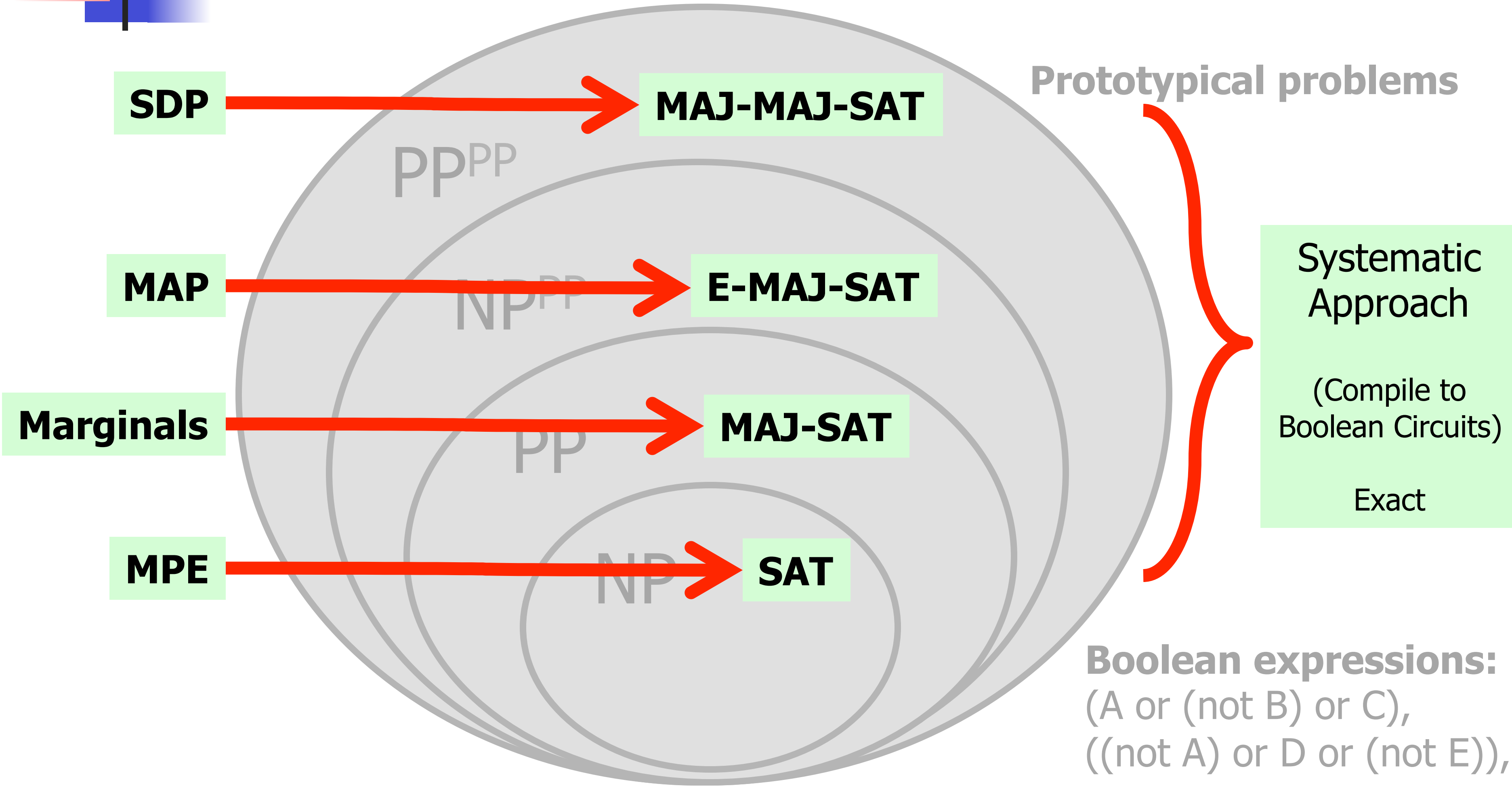
Functional Program (Scala-like)

```
def path(start,end,visited=List())={
    if(start == end)
        return true
    if(visited.contains(start))
        return false
    return start.neighbors.exists{
        path(_,end,(visited+start))
    }
}
nodeA.neighbors = …random vars…
nodeB.neighbors = …random vars…
```

State of the art for **discrete** probabilistic programs

Courtesy G. Van den Broeck

Complete Problems



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

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}

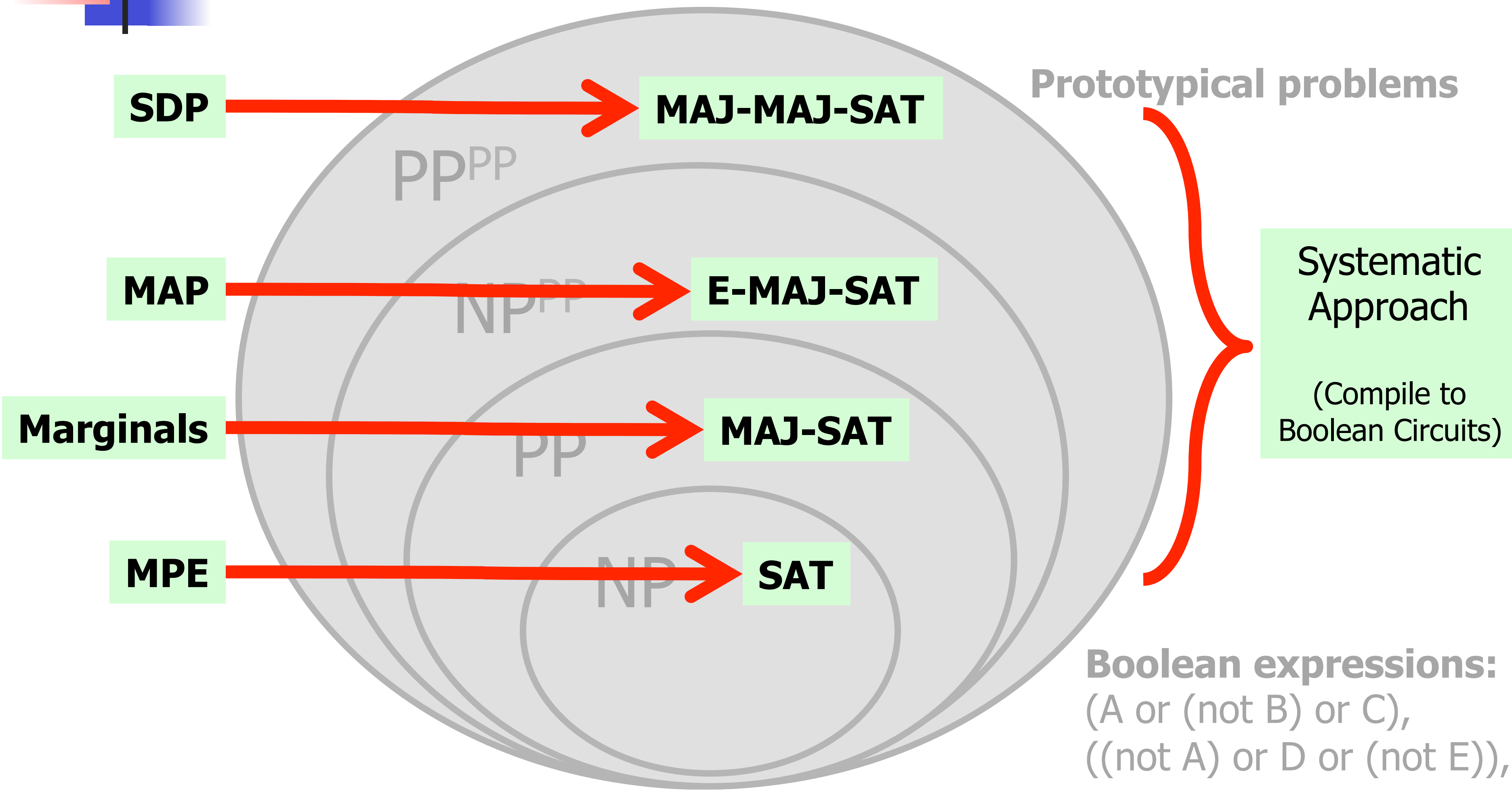
MAJ-MAJ-SAT: Is there a majority of **X**-instantiation under which the majority of **Y**-instantiations satisfying?

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

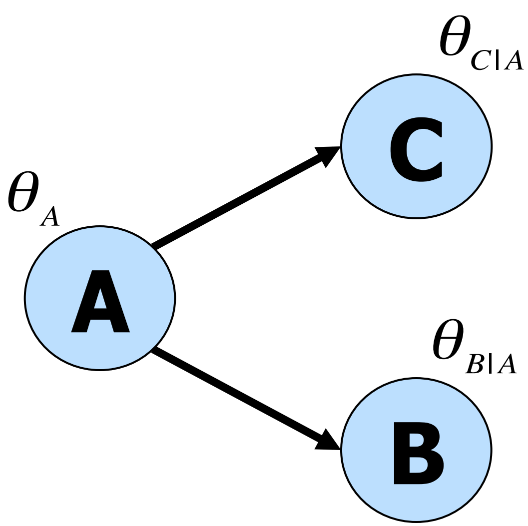
Reductions

Prototypical problems



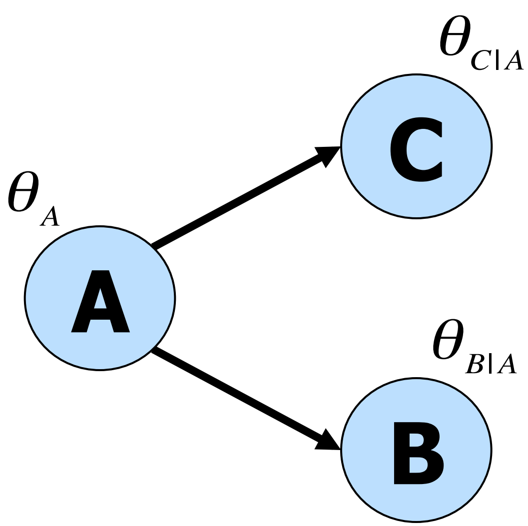
Boolean expressions:
(A or (not B) or C),
((not A) or D or (not E)),
….

Example Reduction



A	B	C	Pr(.)
T	T	T	$\theta_A\theta_{B\|A}\theta_{C\|A}$
T	T	F	$\theta_A\theta_{B\|A}\theta_{\neg C\|A}$
T	F	T	$\theta_A\theta_{\neg B\|A}\theta_{C\|A}$
T	F	F	$\theta_A\theta_{\neg B\|A}\theta_{\neg C\|A}$
F	T	T	$\theta_{\neg A}\theta_{B\|\neg A}\theta_{C\|\neg A}$
F	T	F	$\theta_{\neg A}\theta_{B\|\neg A}\theta_{\neg C\|\neg A}$
F	F	T	$\theta_{\neg A}\theta_{\neg B\|\neg A}\theta_{C\|\neg A}$
F	F	F	$\theta_{\neg A}\theta_{\neg B\|\neg A}\theta_{\neg C\|\neg A}$

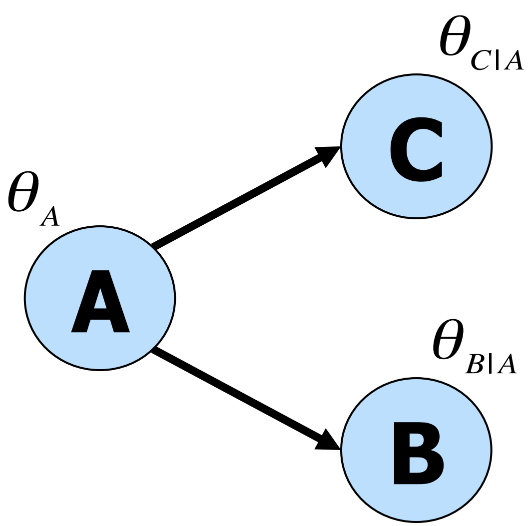
Example Reduction



$\boxed{\Pr(A)}$

A	B	C	Pr(.)
T	T	T	$\theta_A \theta_{B\|A} \theta_{C\|A}$
T	T	F	$\theta_A \theta_{B\|A} \theta_{\neg C\|A}$
T	F	T	$\theta_A \theta_{\neg B\|A} \theta_{C\|A}$
T	F	F	$\theta_A \theta_{\neg B\|A} \theta_{\neg C\|A}$
F	T	T	$\theta_{\neg A} \theta_{B\|\neg A} \theta_{C\|\neg A}$
F	T	F	$\theta_{\neg A} \theta_{B\|\neg A} \theta_{\neg C\|\neg A}$
F	F	T	$\theta_{\neg A} \theta_{\neg B\|\neg A} \theta_{C\|\neg A}$
F	F	F	$\theta_{\neg A} \theta_{\neg B\|\neg A} \theta_{\neg C\|\neg A}$

Example Reduction



A	B	C	Pr(.)
T	T	T	$\theta_A \theta_{B\|A} \theta_{C\|A}$
T	T	F	$\theta_A \theta_{B\|A} \theta_{\neg C\|A}$
T	F	T	$\theta_A \theta_{\neg B\|A} \theta_{C\|A}$
T	F	F	$\theta_A \theta_{\neg B\|A} \theta_{\neg C\|A}$
F	T	T	$\theta_{\neg A} \theta_{B\|\neg A} \theta_{C\|\neg A}$
F	T	F	$\theta_{\neg A} \theta_{B\|\neg A} \theta_{\neg C\|\neg A}$
F	F	T	$\theta_{\neg A} \theta_{\neg B\|\neg A} \theta_{C\|\neg A}$
F	F	F	$\theta_{\neg A} \theta_{\neg B\|\neg A} \theta_{\neg C\|\neg A}$

$\boxed{\Pr(\neg B)}$

A	B	C	θ_A	$\theta_{\neg A}$	$\theta_{B\|A}$	$\theta_{\neg B\|A}$	$\theta_{B\|\neg A}$	$\theta_{\neg B\|\neg A}$	$\theta_{C\|A}$	$\theta_{\neg C\|A}$	$\theta_{C\|\neg A}$	$\theta_{\neg C\|\neg A}$
T	T	T										
T	T	F										
T	F	T										
T	F	F										
F	T	T										
F	T	F										
F	F	T										
F	F	F										

13 variables, 2^{13} instantiations, 8 models

A	B	C	θ_A	$\theta_{\neg A}$	$\theta_{B\|A}$	$\theta_{\neg B\|A}$	$\theta_{B\|\neg A}$	$\theta_{\neg B\|\neg A}$	$\theta_{C\|A}$	$\theta_{\neg C\|A}$	$\theta_{C\|\neg A}$	$\theta_{\neg C\|\neg A}$
T	T	T	T	F	T	F	F	F	T	F	F	F
T	T	F	T	F	T	F	F	F	F	T	F	F
T	F	T	T	F	F	T	F	F	T	F	F	F
T	F	F	T	F	F	T	F	F	F	T	F	F
F	T	T	F	T	F	F	T	F	F	F	T	F
F	T	F	F	T	F	F	T	F	F	F	F	T
F	F	T	F	T	F	F	F	T	F	F	T	F
F	F	F	F	T	F	F	F	T	F	F	F	T

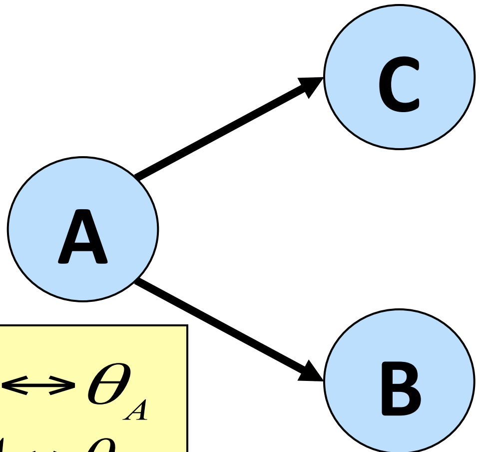
13 variables, 2^{13} instantiations, 8 models

A	B	C	θ_A	$\theta_{\neg A}$	$\theta_{B\|A}$	$\theta_{\neg B\|A}$	$\theta_{B\|\neg A}$	$\theta_{\neg B\|\neg A}$	$\theta_{C\|A}$	$\theta_{\neg C\|A}$	$\theta_{C\|\neg A}$	$\theta_{\neg C\|\neg A}$
T	T	T	T	F	T	F	F	F	T	F	F	F
T	T	F	T	F	T	F	F	F	F	T	F	F
T	F	T	T	F	F	T	F	F	T	F	F	F
T	F	F	T	F	F	T	F	F	F	T	F	F
F	T	T	F	T	F	F	T	F	F	F	T	F
F	T	F	F	T	F	F	T	F	F	F	F	T
F	F	T	F	T	F	F	F	T	F	F	T	F
F	F	F	F	T	F	F	F	T	F	F	F	T

13 variables, 2^{13} instantiations, 8 models

$$w(A) = w(\neg A) = 1, \ldots$$

$$w(\theta_{B|A}) = .6, w(\neg \theta_{B|A}) = 1, \ldots$$



$$A \leftrightarrow \theta_A$$
$$\neg A \leftrightarrow \theta_{\neg A}$$

$$A \wedge C \leftrightarrow \theta_{C|A}$$
$$A \wedge \neg C \leftrightarrow \theta_{\neg C|A}$$
$$\vdots$$

$$A \wedge B \leftrightarrow \theta_{B|A}$$
$$A \wedge \neg B \leftrightarrow \theta_{\neg B|A}$$
$$\vdots$$

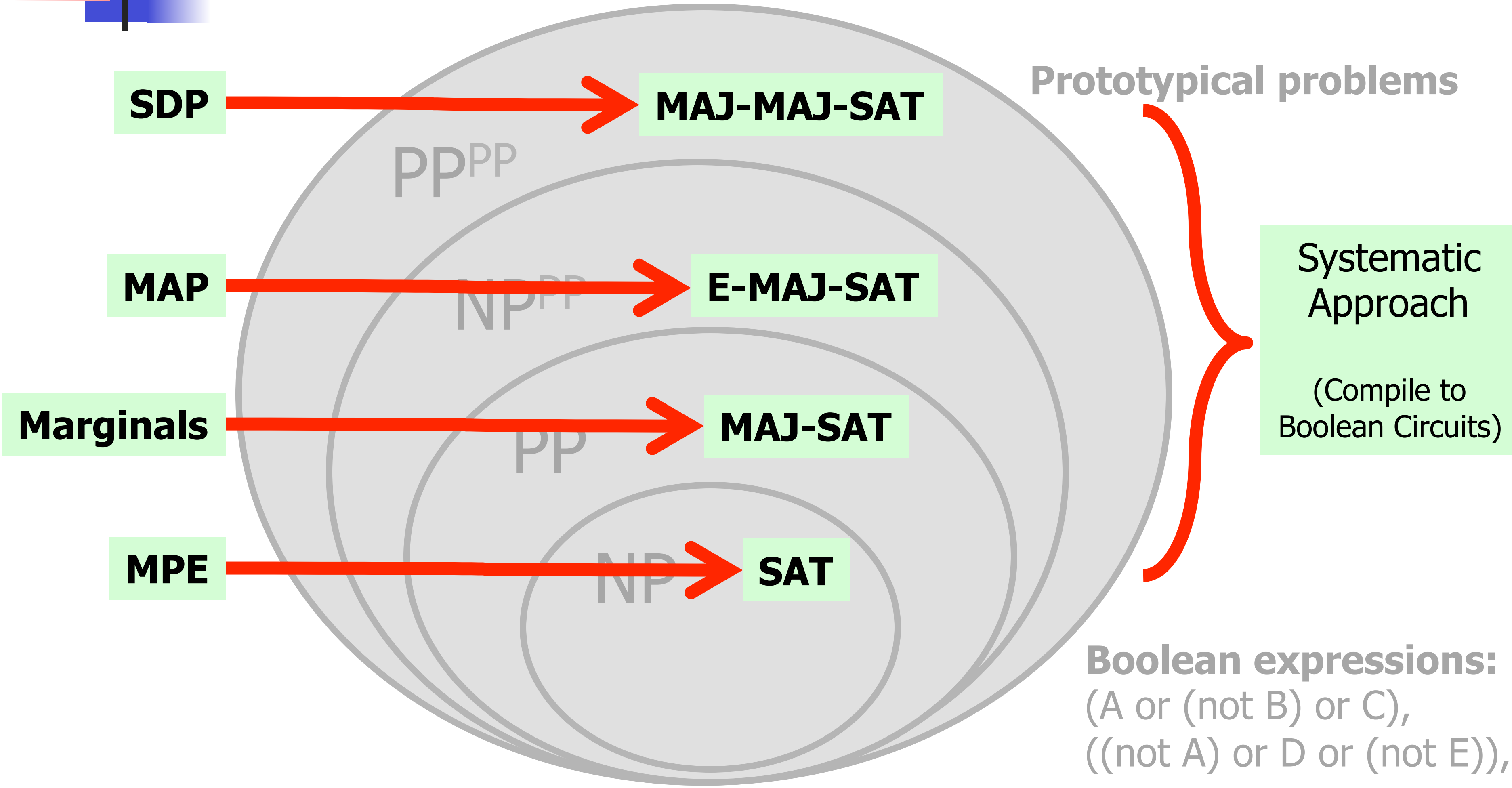
A	B	C	Pr(.)
T	T	T	$\theta_A \theta_{B|A} \theta_{C|A}$
T	T	F	$\theta_A \theta_{B|A} \theta_{\neg C|A}$
T	F	T	$\theta_A \theta_{\neg B|A} \theta_{C|A}$
T	F	F	$\theta_A \theta_{\neg B|A} \theta_{\neg C|A}$
F	T	T	$\theta_{\neg A} \theta_{B|\neg A} \theta_{C|\neg A}$
F	T	F	$\theta_{\neg A} \theta_{B|\neg A} \theta_{\neg C|\neg A}$
F	F	T	$\theta_{\neg A} \theta_{\neg B|\neg A} \theta_{C|\neg A}$
F	F	F	$\theta_{\neg A} \theta_{\neg B|\neg A} \theta_{\neg C|\neg A}$

$$A, B, \neg C, \theta_A, \theta_{B|A}, \theta_{\neg C|A}, \neg \theta_{\neg A}, \ldots, \neg \theta_{\neg C|\neg A}$$

1 1 1 .3 .6 .8 1 1

A. Darwiche

Systematic Approach



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

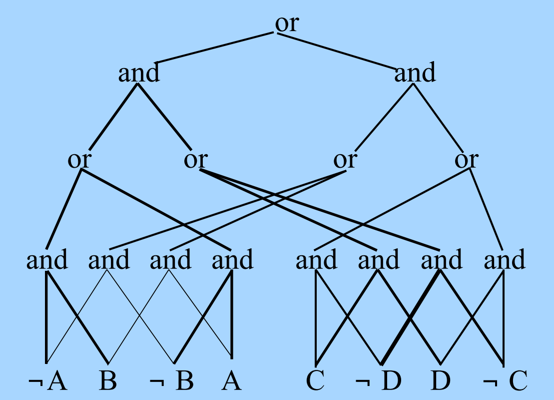
Knowledge Compilation

encoding

```
(A and (not B))
or(C and (not D))
or ((not C) and D)
...
```

Compiler

NNF Circuit

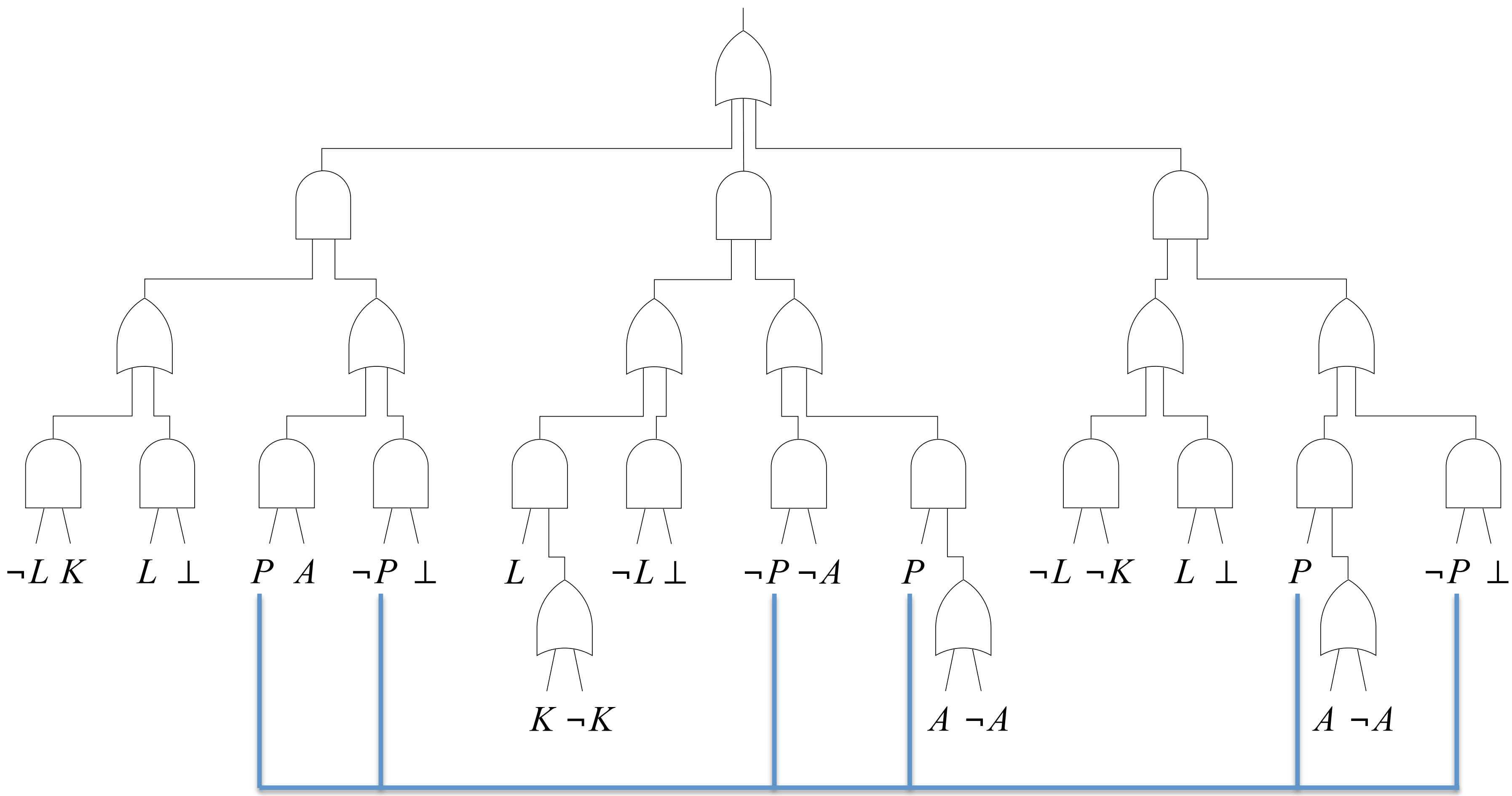


MAJ-MAJ-SAT
E-MAJ-SAT
MAJ-SAT
SAT

Answer in Linear Time

NNF Circuits

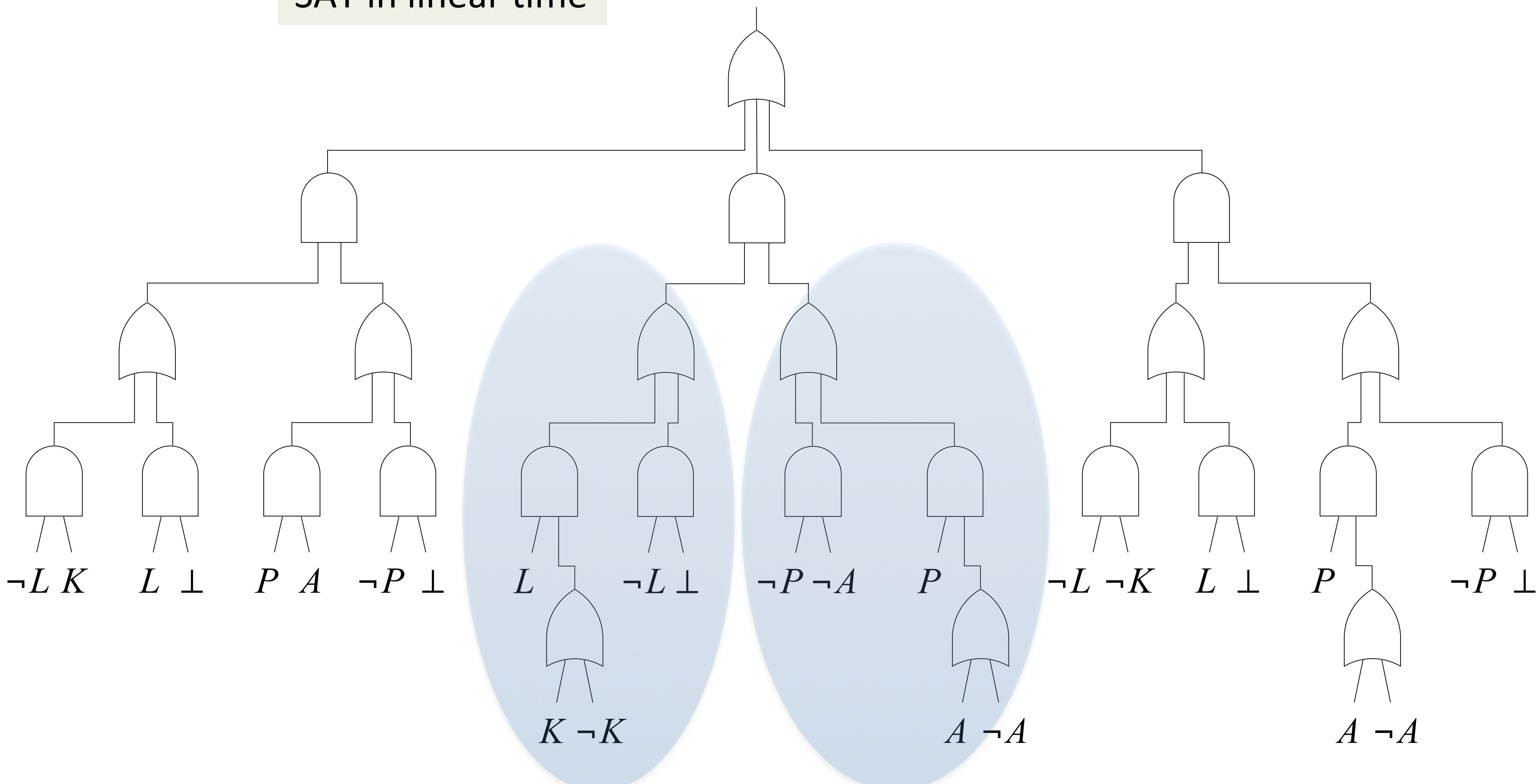
$P \vee L$
$A \Rightarrow P$
$K \Rightarrow (P \vee L)$



Decomposability (DNNF)

Darwiche, JACM 2001

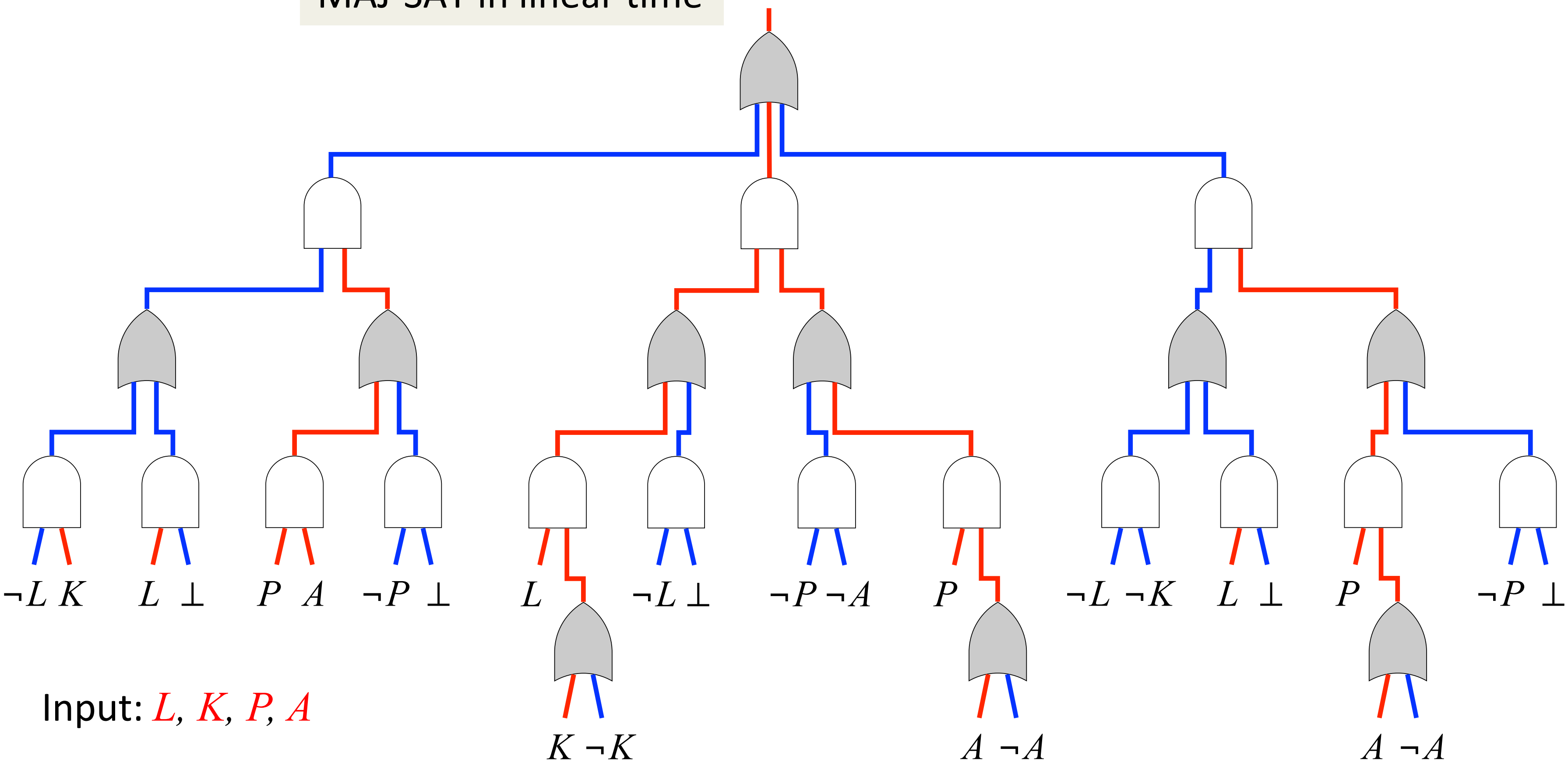
SAT in linear time



Determinism (d-DNNF)

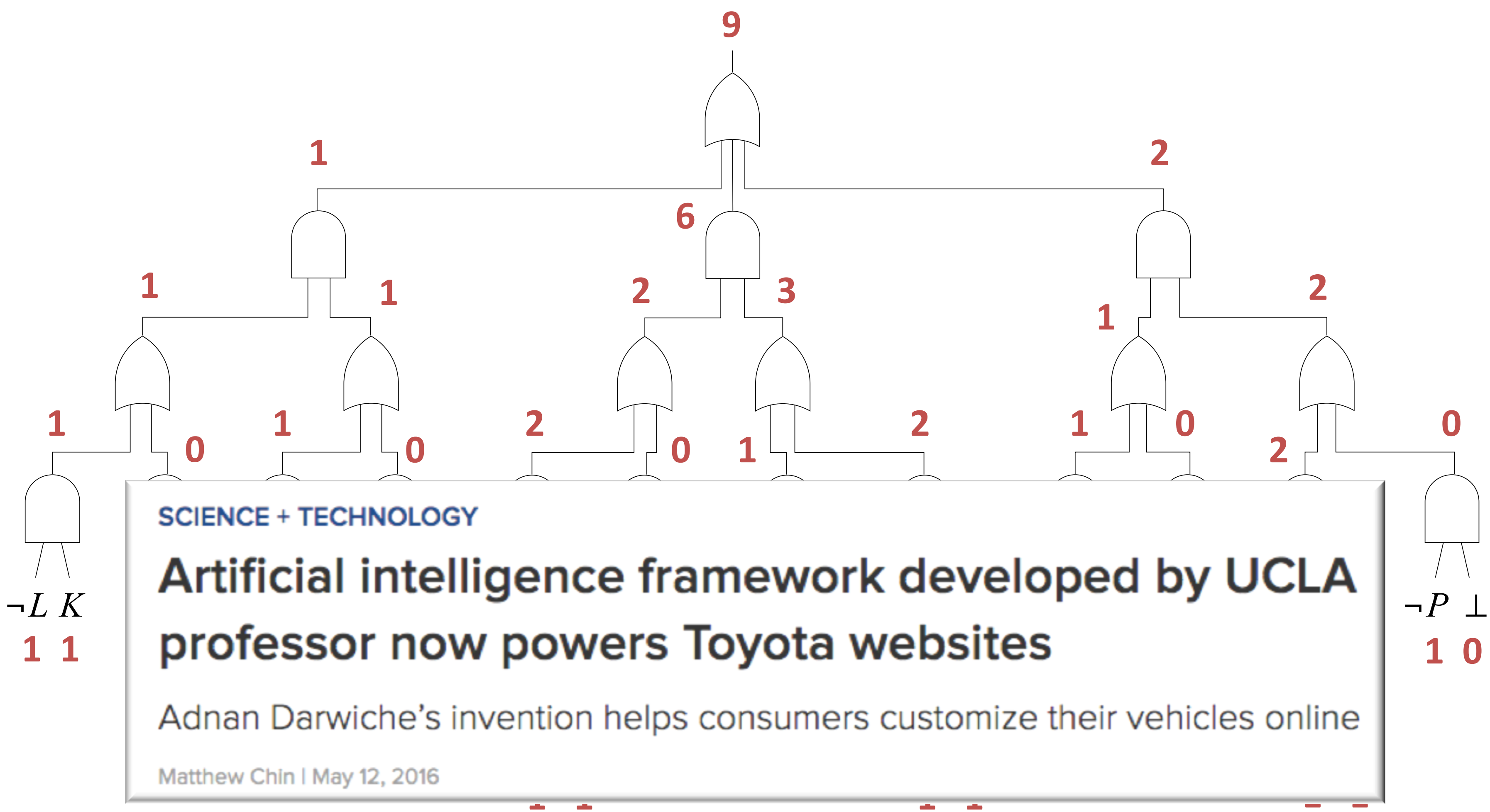
Darwiche, JANCL 2000

MAJ-SAT in linear time



Input: *L*, *K*, *P*, *A*

Model Counting



SCIENCE + TECHNOLOGY

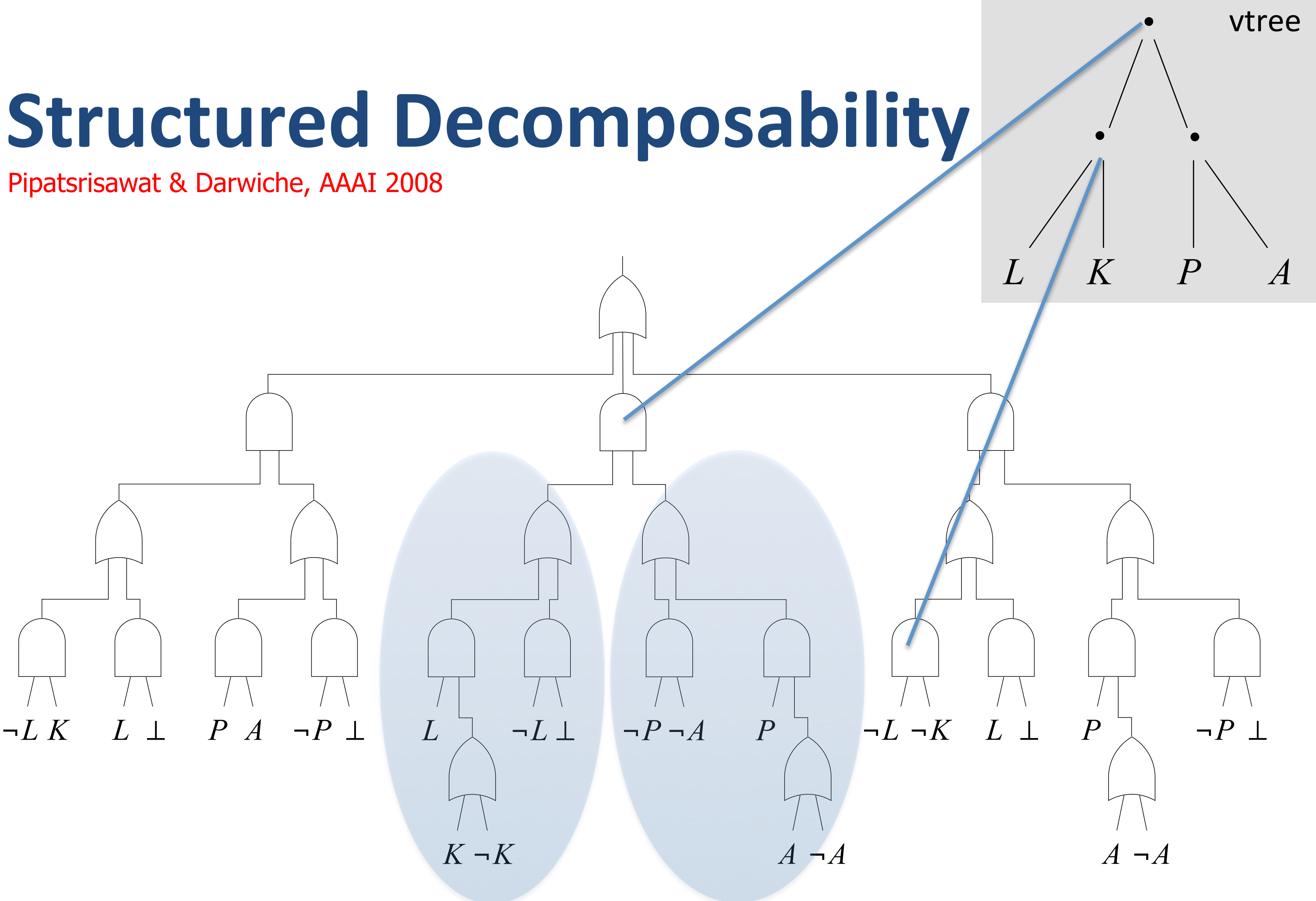
Artificial intelligence framework developed by UCLA professor now powers Toyota websites

Adnan Darwiche's invention helps consumers customize their vehicles online

Matthew Chin | May 12, 2016

Structured Decomposability

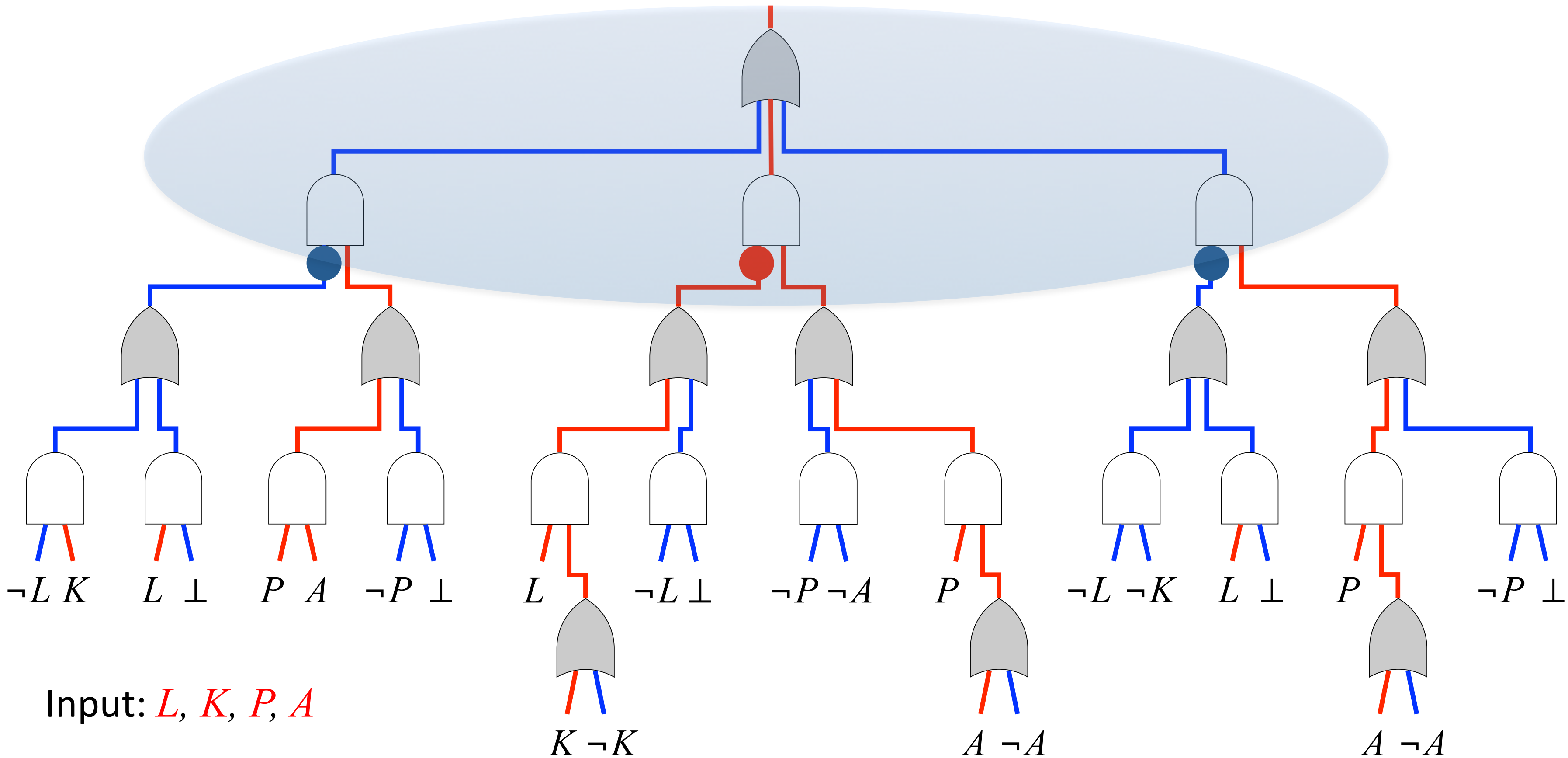
Pipatsrisawat & Darwiche, AAAI 2008



Partitioned Determinism (SDD Circuits)

Darwiche, IJCAI 2011

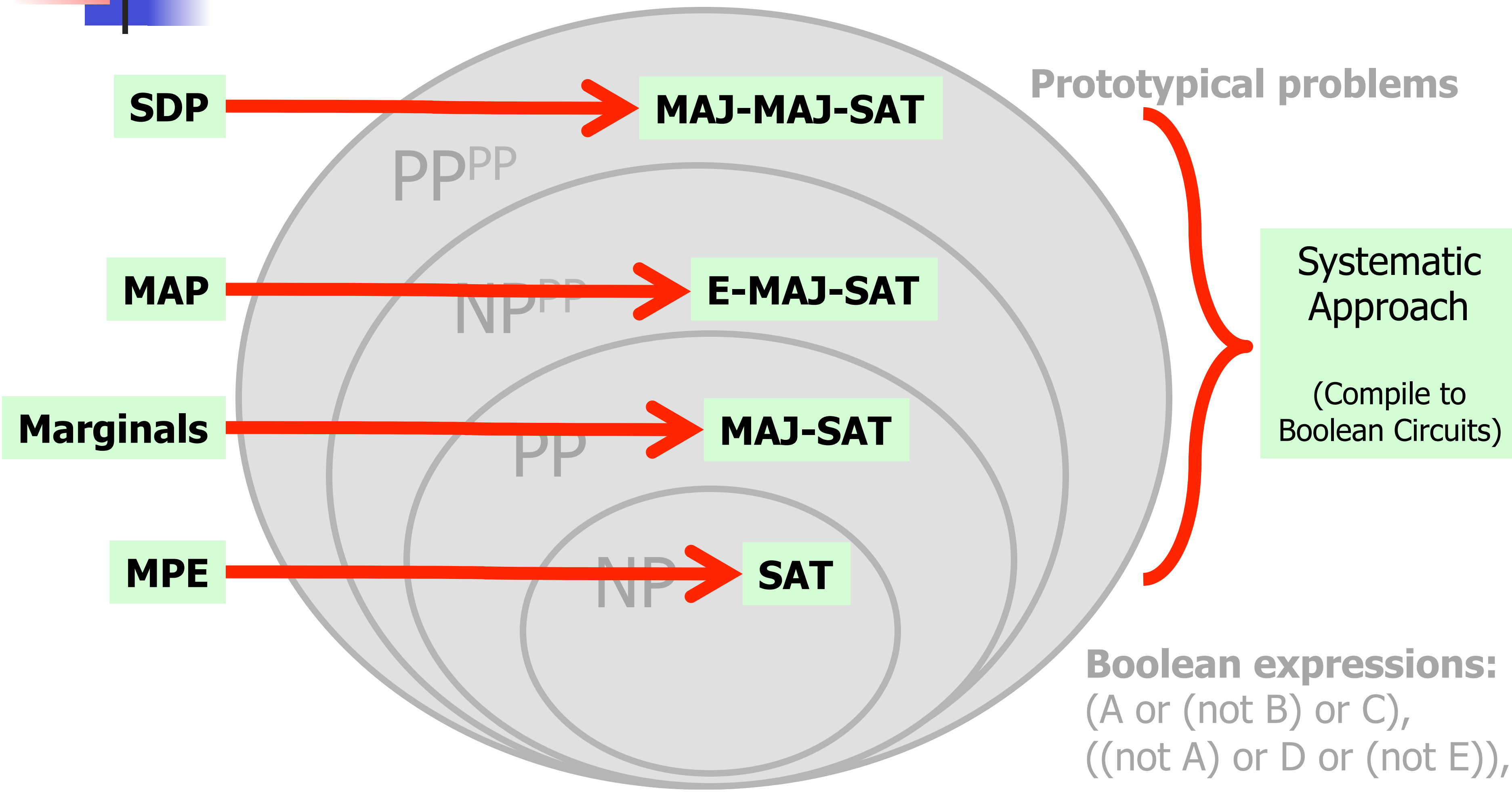
SDD compiler: source out soon



MAJ-MAJ-SAT in linear time using appropriate vtree

Oztok & Darwiche, KR 2016

Beyond NP



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

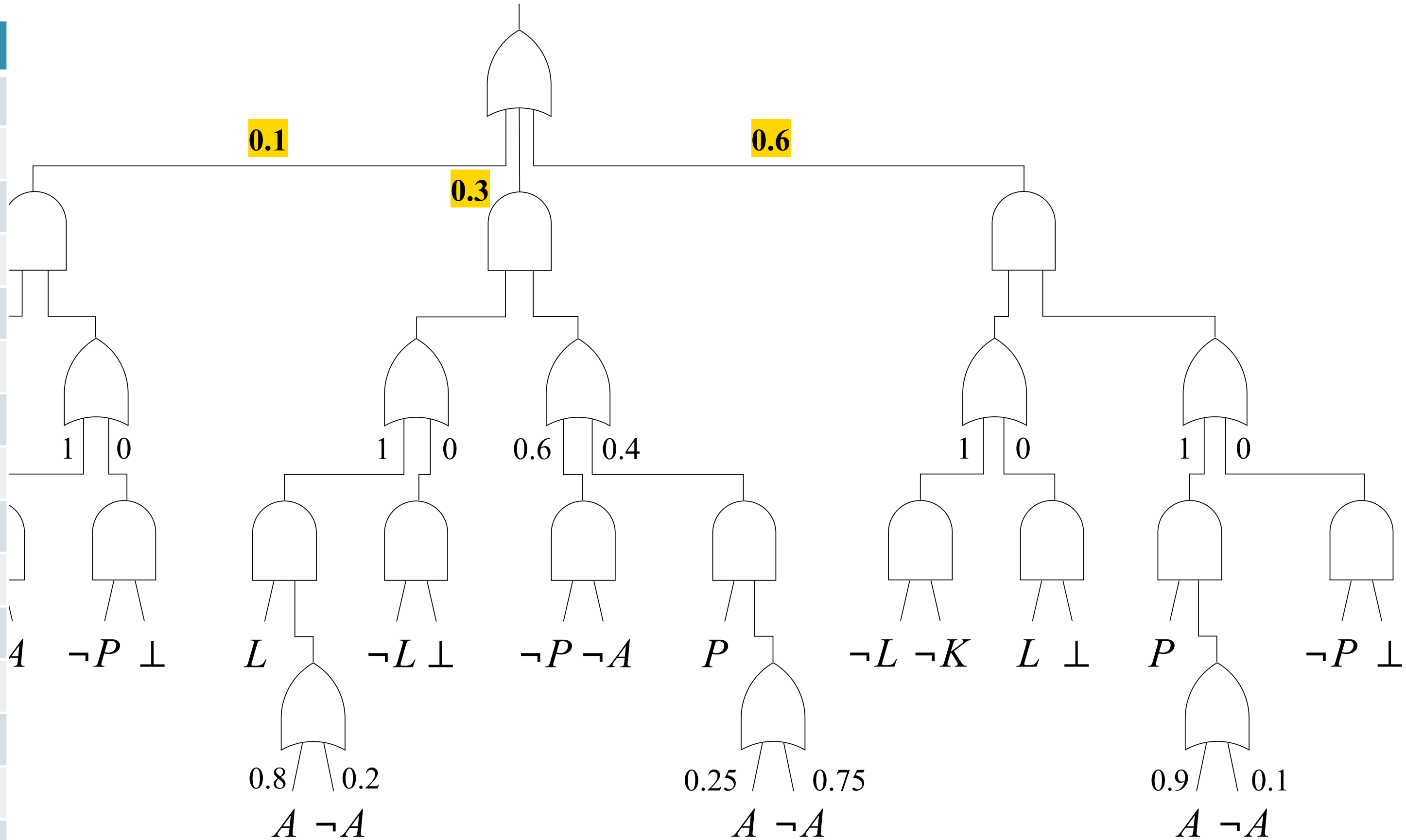
Agenda

- Logic in Probabilistic Inference
  - Four probabilistic queries
  - Beyond NP: PP, NP^{PP} & PP^{PP}
- Logic in Machine Learning
  - Beyond data: Background knowledge
  - Beyond classical datasets

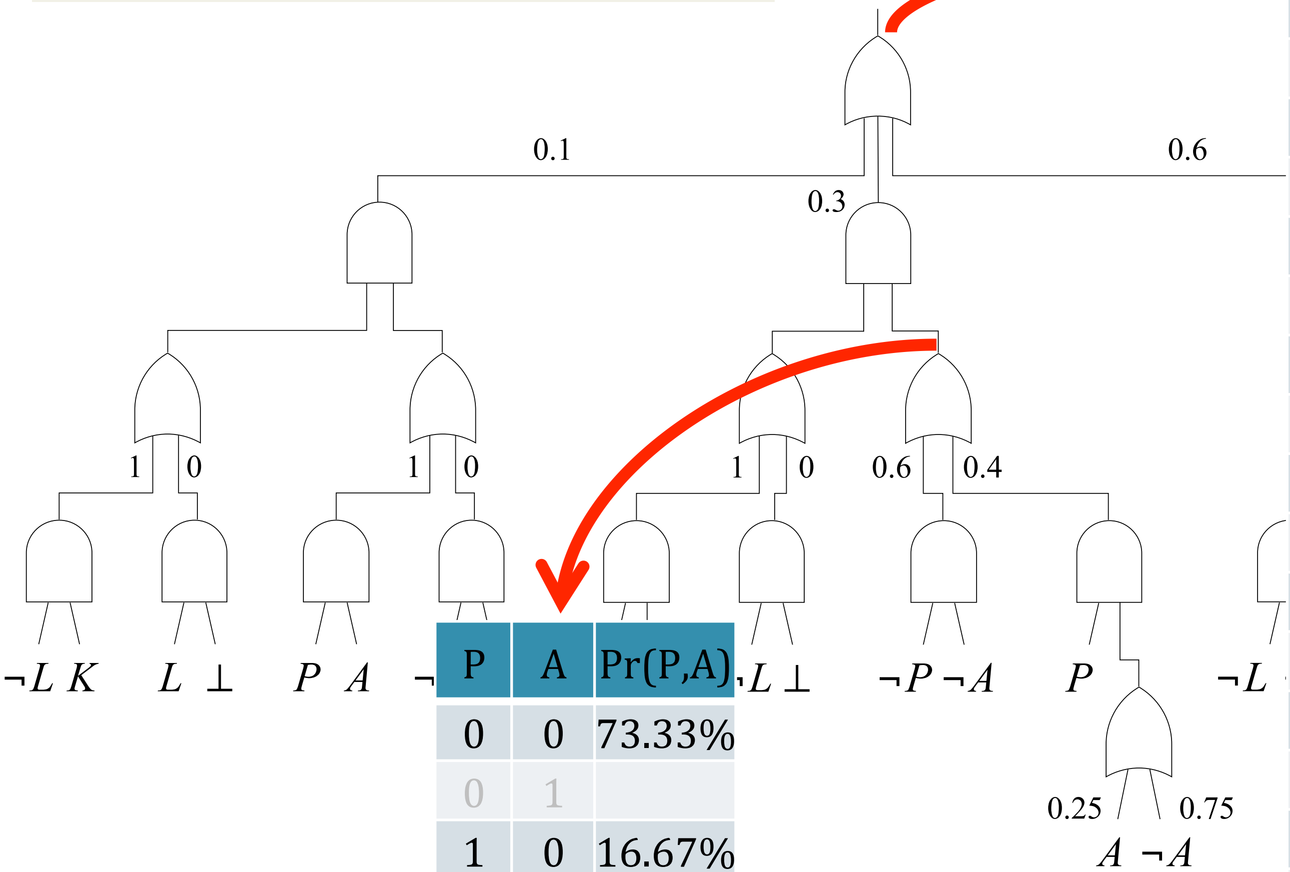
Probabilistic SDD Circuits

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)



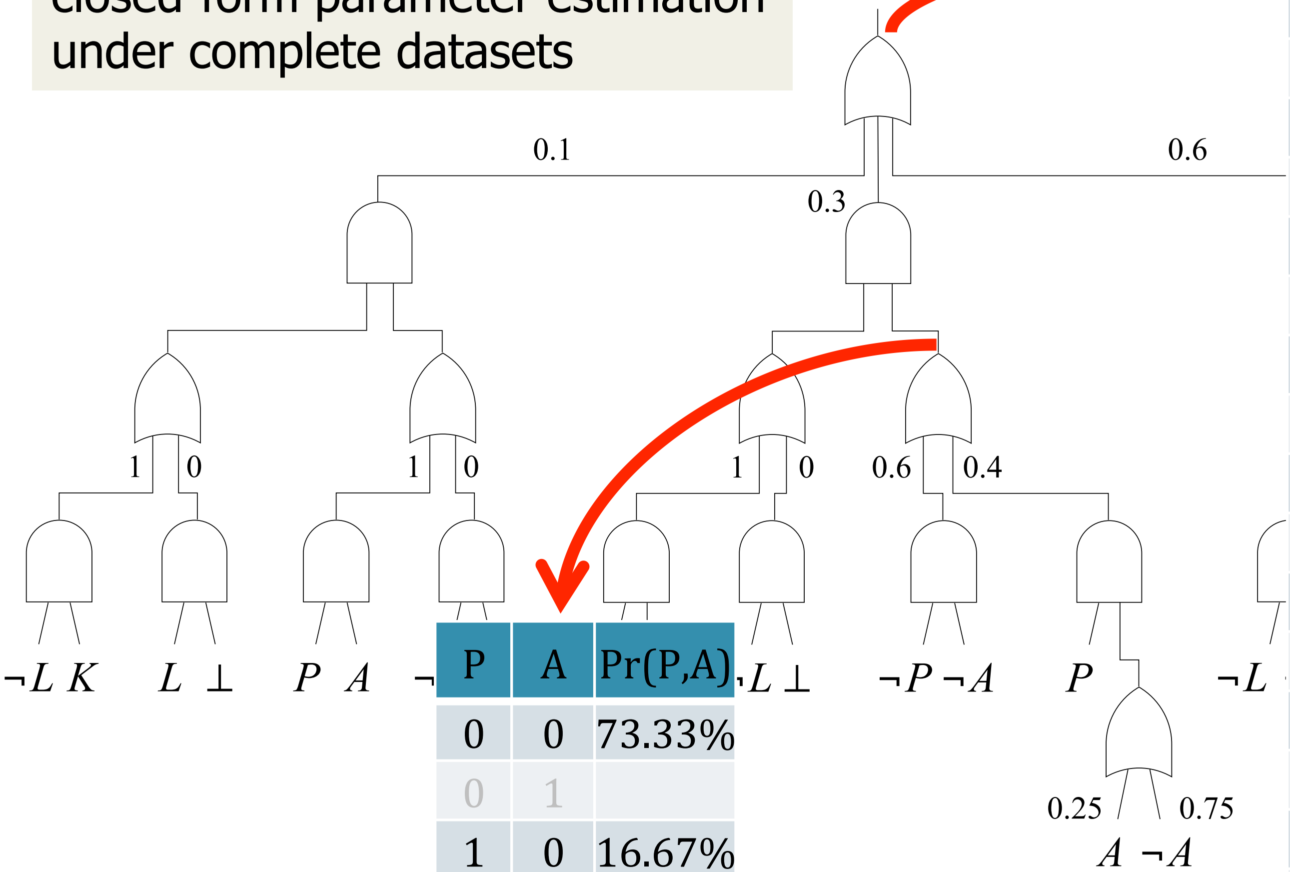
P	A	Pr(P,A)
0	0	73.33%
0	1	
1	0	16.67%
1	1	10.00%

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%

complete & canonical representation

tractable for MPE & marginals

closed-form parameter estimation
under complete datasets



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

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%

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

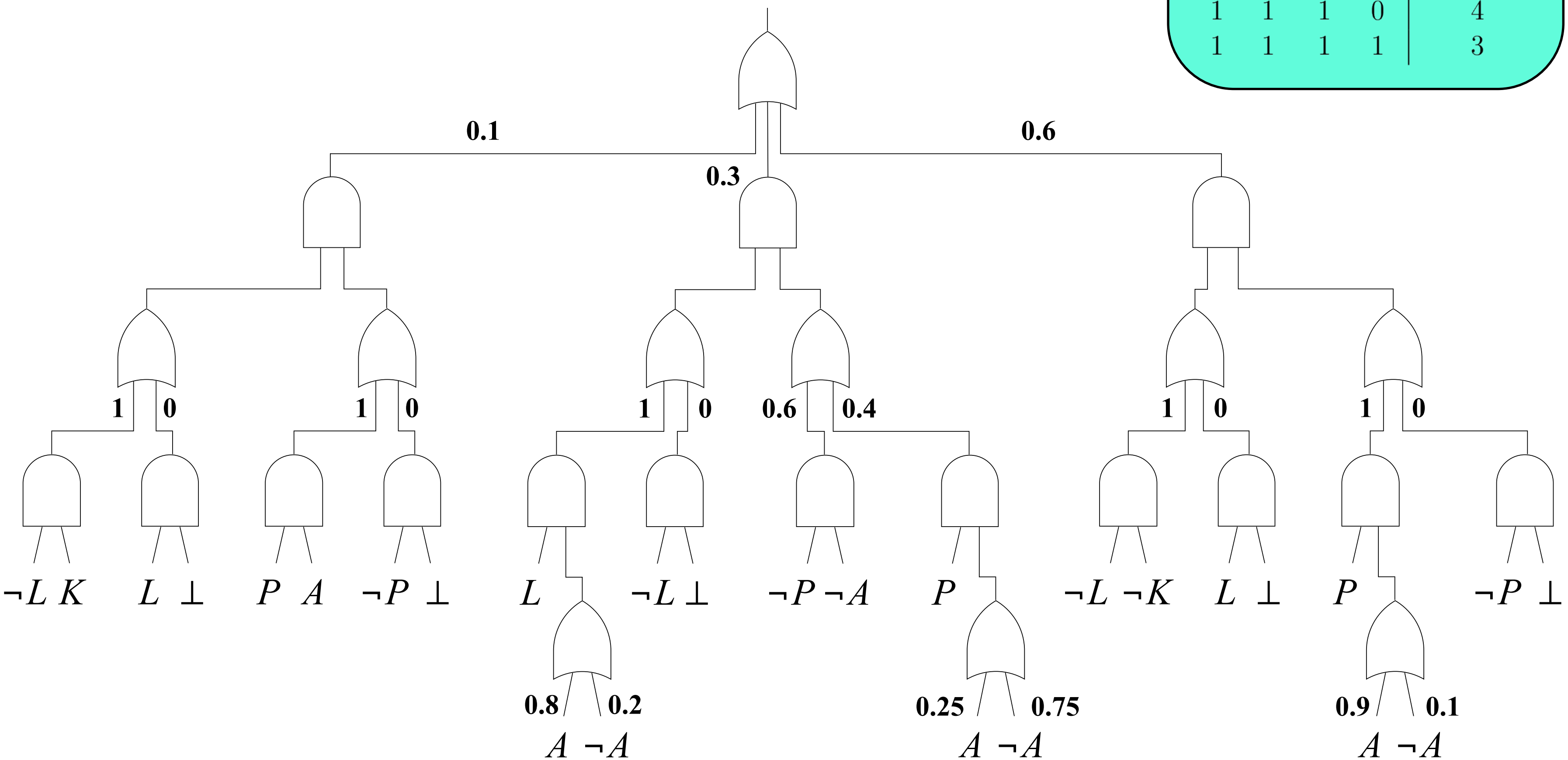
L	K	P	A
0	0	0	0
0	0	0	1
0	0	1	0
0	0	1	1
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	0
1	1	0	1
1	1	1	0
1	1	1	1

Learned PSDD

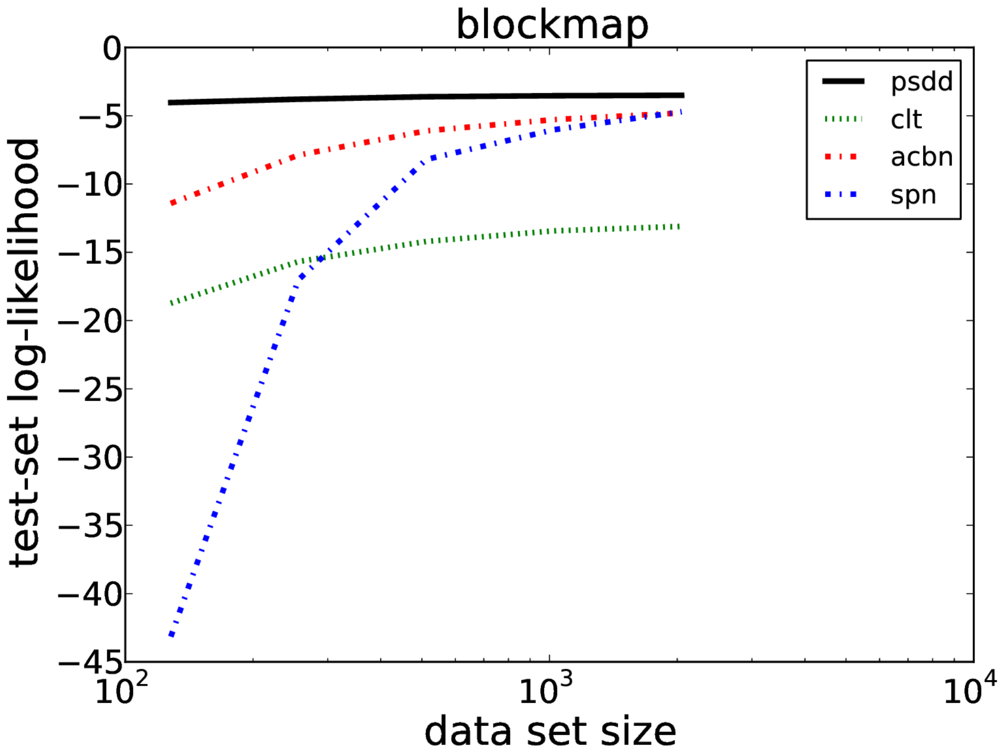
$$P \vee L$$
$$A \Rightarrow P$$
$$K \Rightarrow (P \vee L)$$

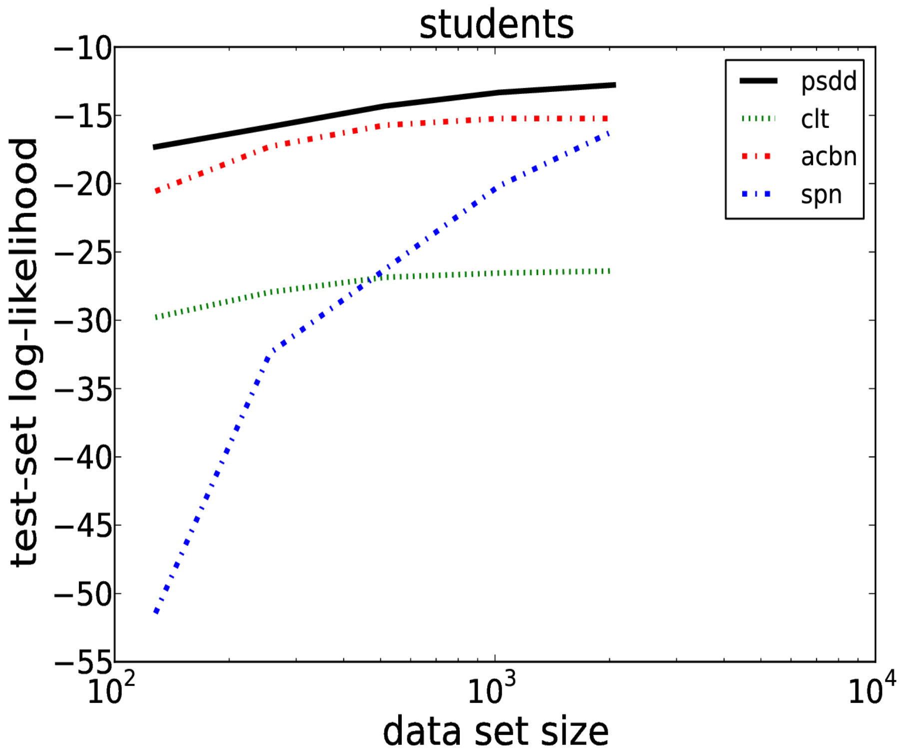
closed-form parameter estimation under complete datasets

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



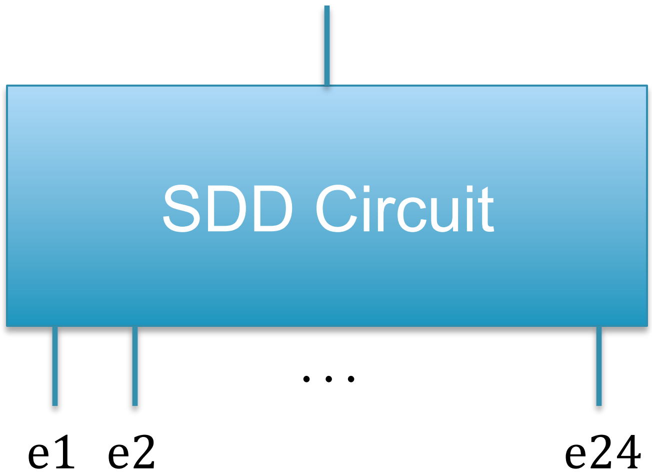
Ignoring Background Knowledge

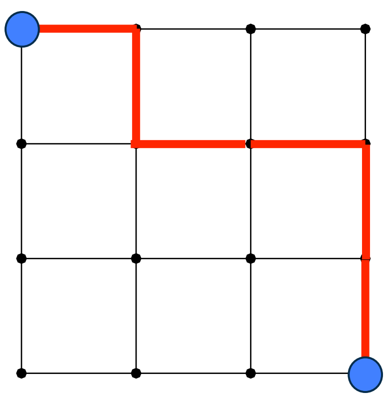


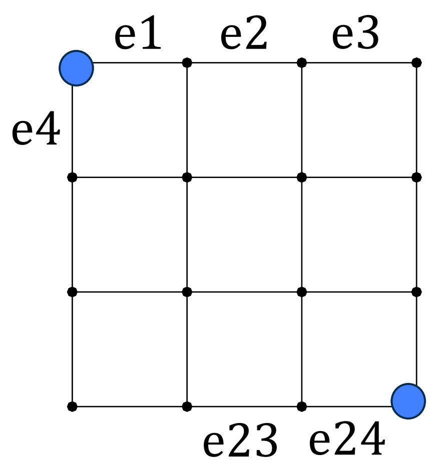


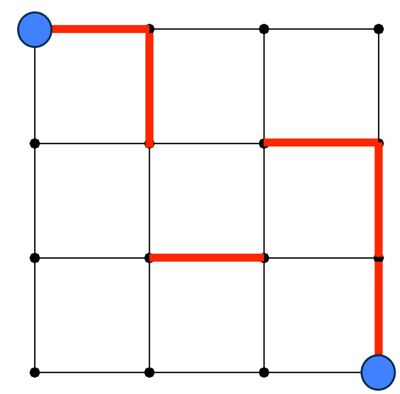
Combinatorial Objects: Routes

Choi et al, AAAI 2016



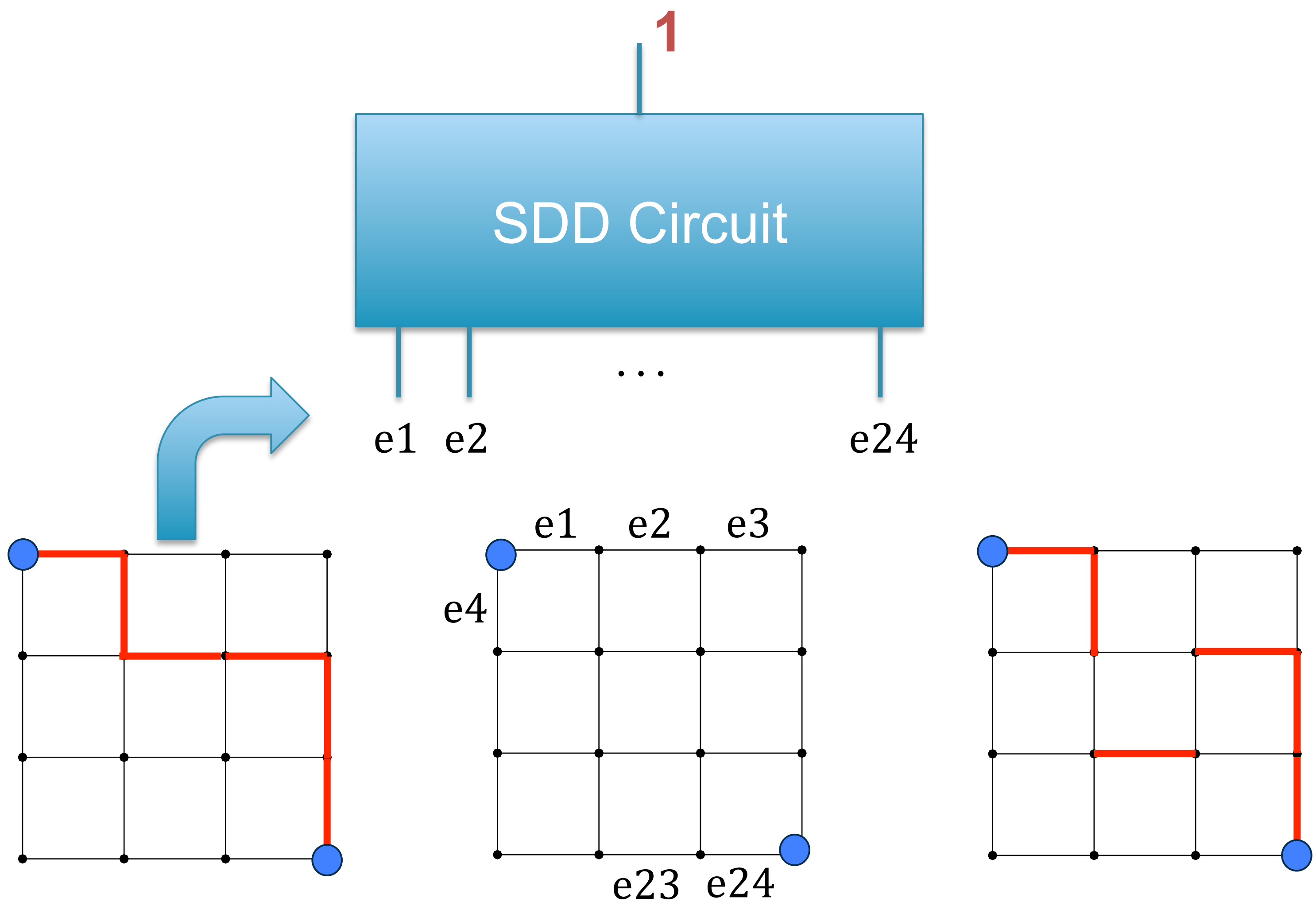






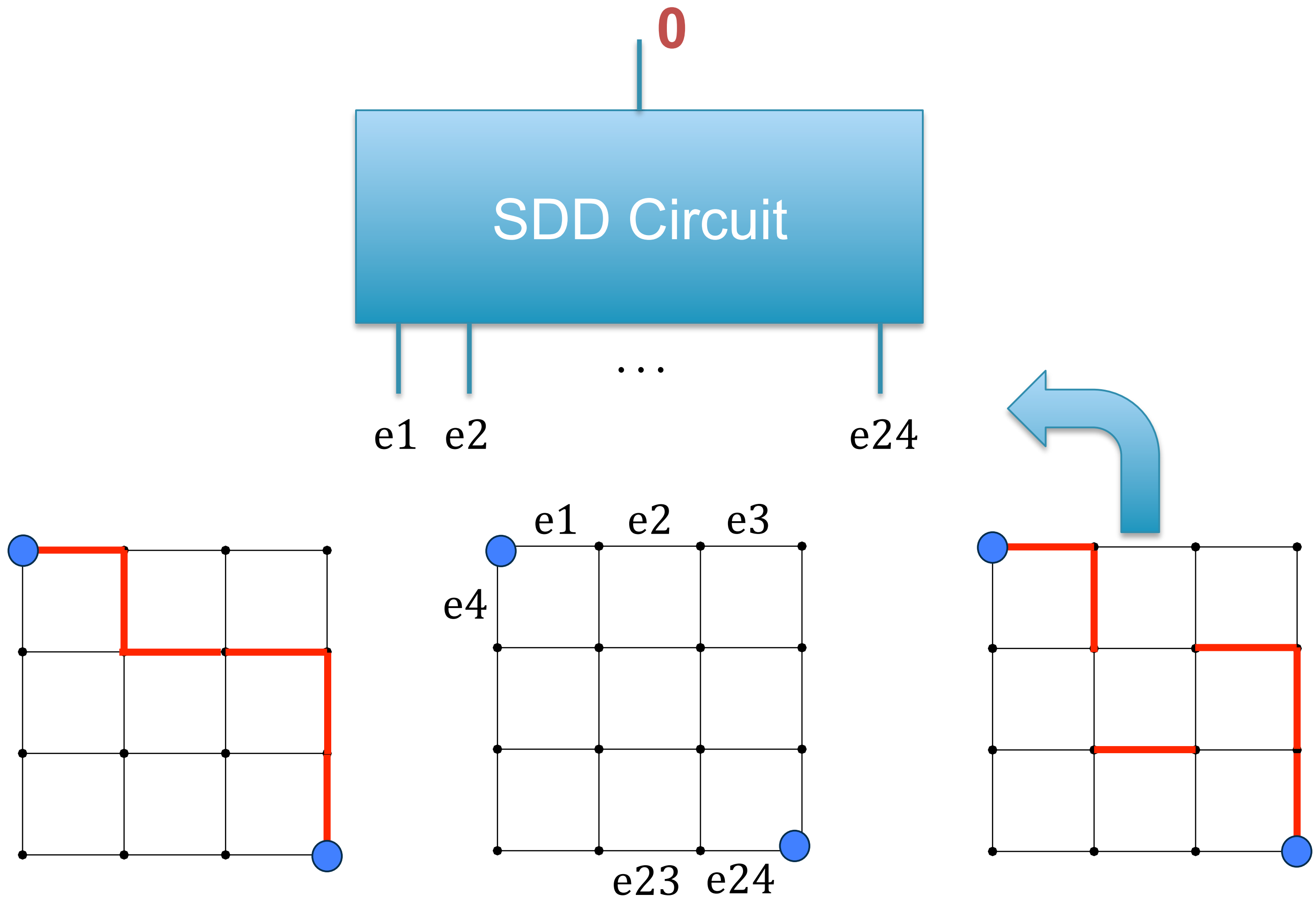
Combinatorial Objects: Routes

Choi et al, AAAI 2016



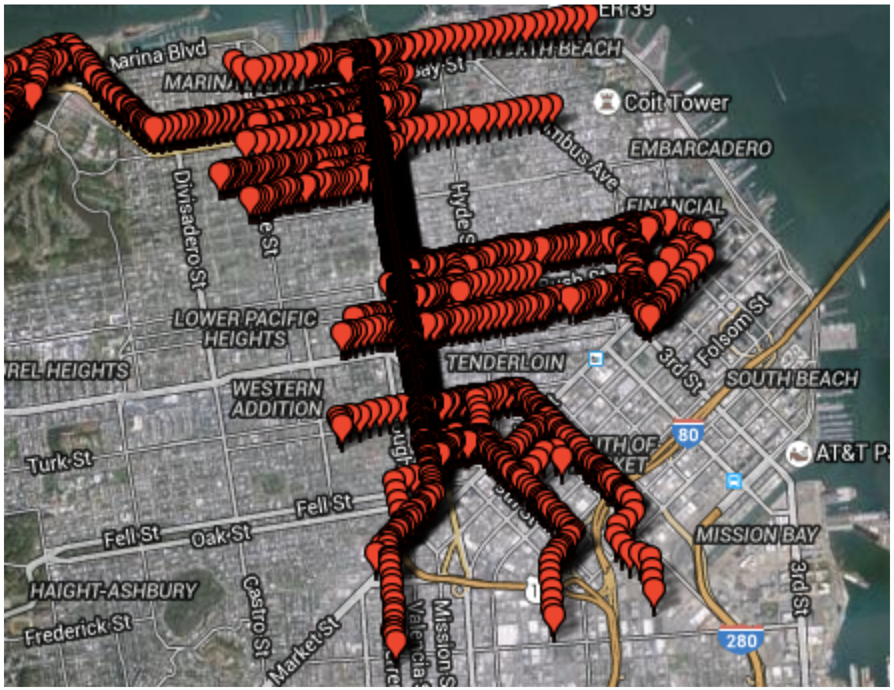
Combinatorial Objects: Routes

Choi et al, AAAI 2016



Combinatorial Objects: Routes

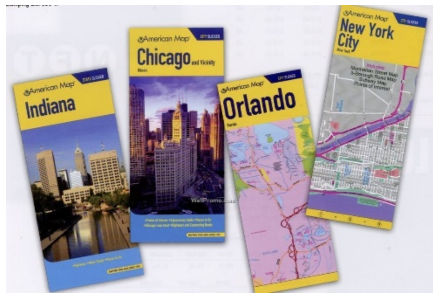
Choi et al, NIPS 2017



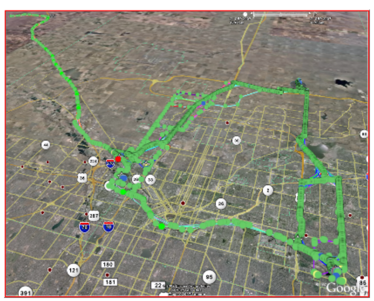
- Uber GPS data in SF
- Project GPS coordinates onto a grid/graph, then learn distributions over routes
- Applications:
  - Detect anomalies
  - Given a partial route, predict its most likely completions

Combining Knowledge & Data

Input: Knowledge (a map)



Input: Data (GPS routes)



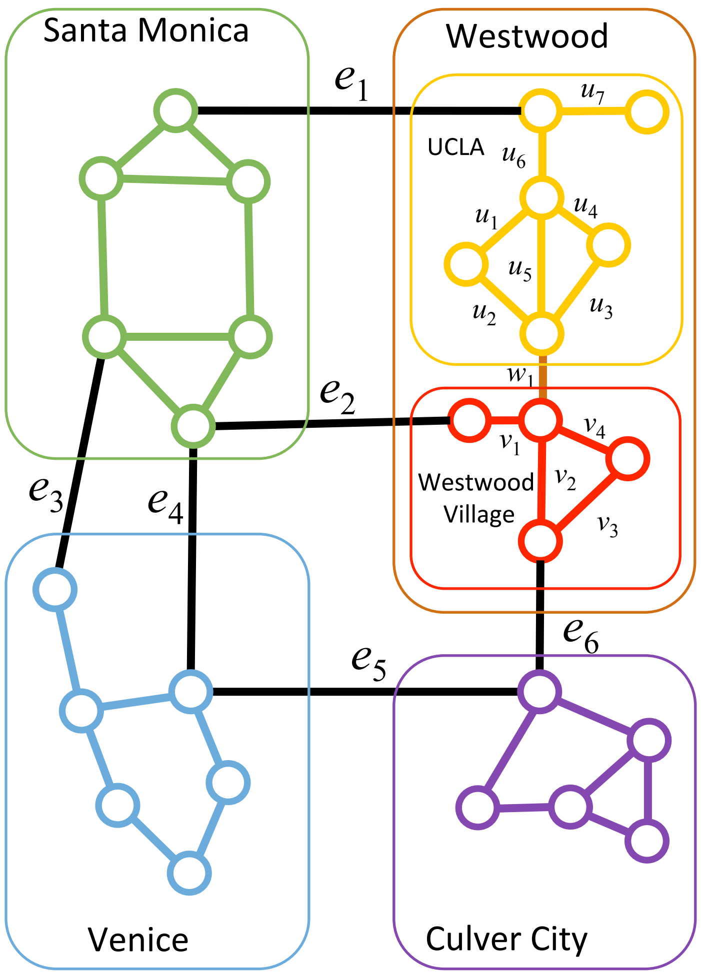
Output: Probabilistic Model over Routes

Estimate traffic
Predict routes
Predict the impact of an intervention

Hierarchical Maps

Choi et al, NIPS 2017

Shen et al, AAAI 2018



Combinatorial Objects: Rankings

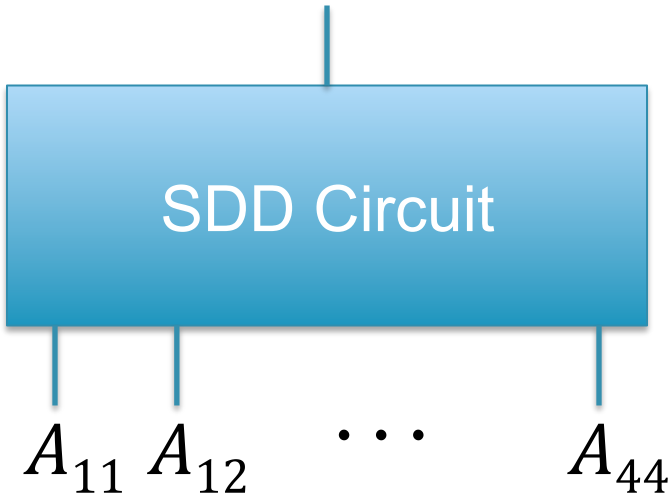
Choi et al, IJCAI 2015

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

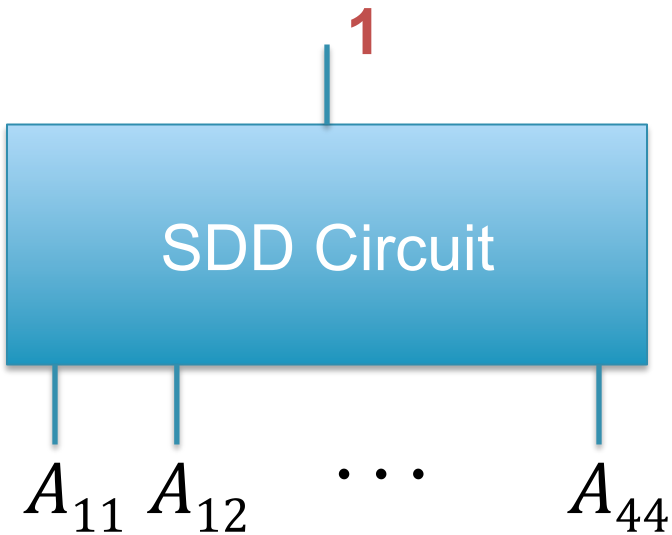
Choi et al, IJCAI 2015



	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

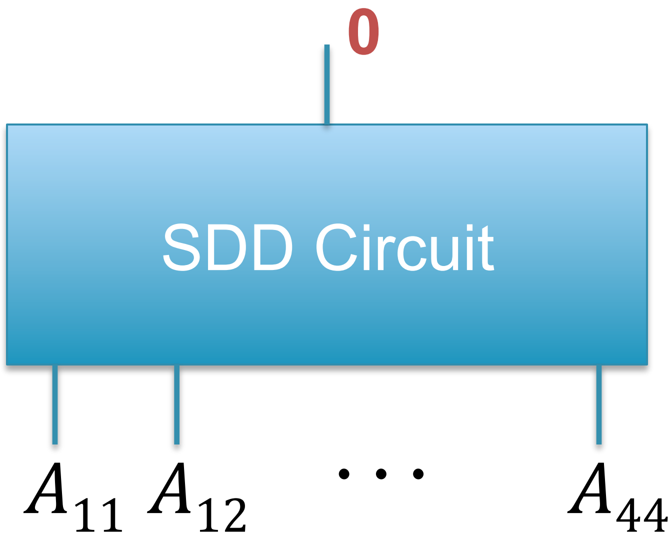
Choi et al, IJCAI 2015



	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

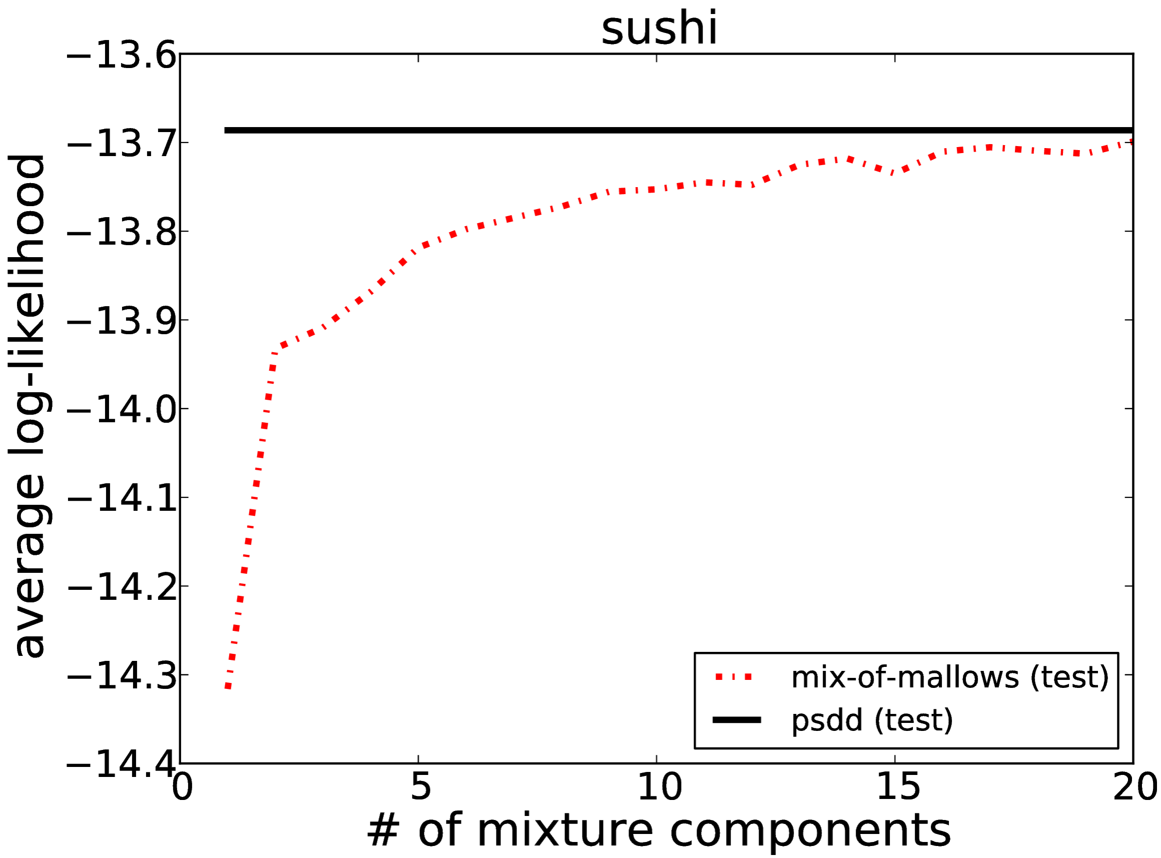
Choi et al, IJCAI 2015



	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	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	?	…
2	fatty tuna	?	?	…
3	tuna	tuna roll	?	…
4	fatty tuna	salmon roe	?	…
5	egg	?	?	…

Structured 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) and (tuna > sea eel)			…
2	(fatty tuna is 1st) and (salmon roe > egg)			…
3	tuna > squid			…
4	egg is last			…
5	egg > squid > shrimp			…

Learning Distributions over Partial Rankings

Choi et al, IJCAI 2015

- Movielens Dataset:
  - 3,900 movies, 6,040 users, 1m ratings
  - take ratings from 64 most rated movies
  - ratings 1-5 converted to pairwise prefs.
- PSDD for **partial** rankings
  - 4 tiers
  - 18,711 parameters

movies by expected tier

rank	movie
1	The Godfather
2	The Usual Suspects
3	Casablanca
4	The Shawshank Redemption
5	Schindler's List
6	One Flew Over the Cuckoo's Nest
7	The Godfather: Part II
8	Monty Python and the Holy Grail
9	Raiders of the Lost Ark
10	Star Wars IV: A New Hope

PSDD Sizes

items	tier size	Size		
n	k	SDD	Structured Space	Unstructured Space
8	2	443	840	$1.84 \cdot 10^{19}$
27	3	4,114	$1.18 \cdot 10^{9}$	$2.82 \cdot 10^{219}$
64	4	23,497	$3.56 \cdot 10^{18}$	$1.04 \cdot 10^{1233}$
125	5	94,616	$3.45 \cdot 10^{31}$	$3.92 \cdot 10^{4703}$
216	6	297,295	$1.57 \cdot 10^{48}$	$7.16 \cdot 10^{14044}$
343	7	781,918	$4.57 \cdot 10^{68}$	$7.55 \cdot 10^{35415}$

Structured Queries

rank	movie
1	Star Wars V: The Empire Strikes Back
2	Star Wars IV: A New Hope
3	The Godfather
4	The Shawshank Redemption
5	The Usual Suspects

- no other Star Wars movie in top-5
- at least one comedy in top-5

rank	movie
1	Star Wars V: The Empire Strikes Back
2	American Beauty
3	The Godfather
4	The Usual Suspects
5	The Shawshank Redemption

diversified recommendations via
logical constraints

Conclusion

Logic in Probabilistic Inference

Beyond NP: PP, NP^{PP} & PP^{PP}

Four queries

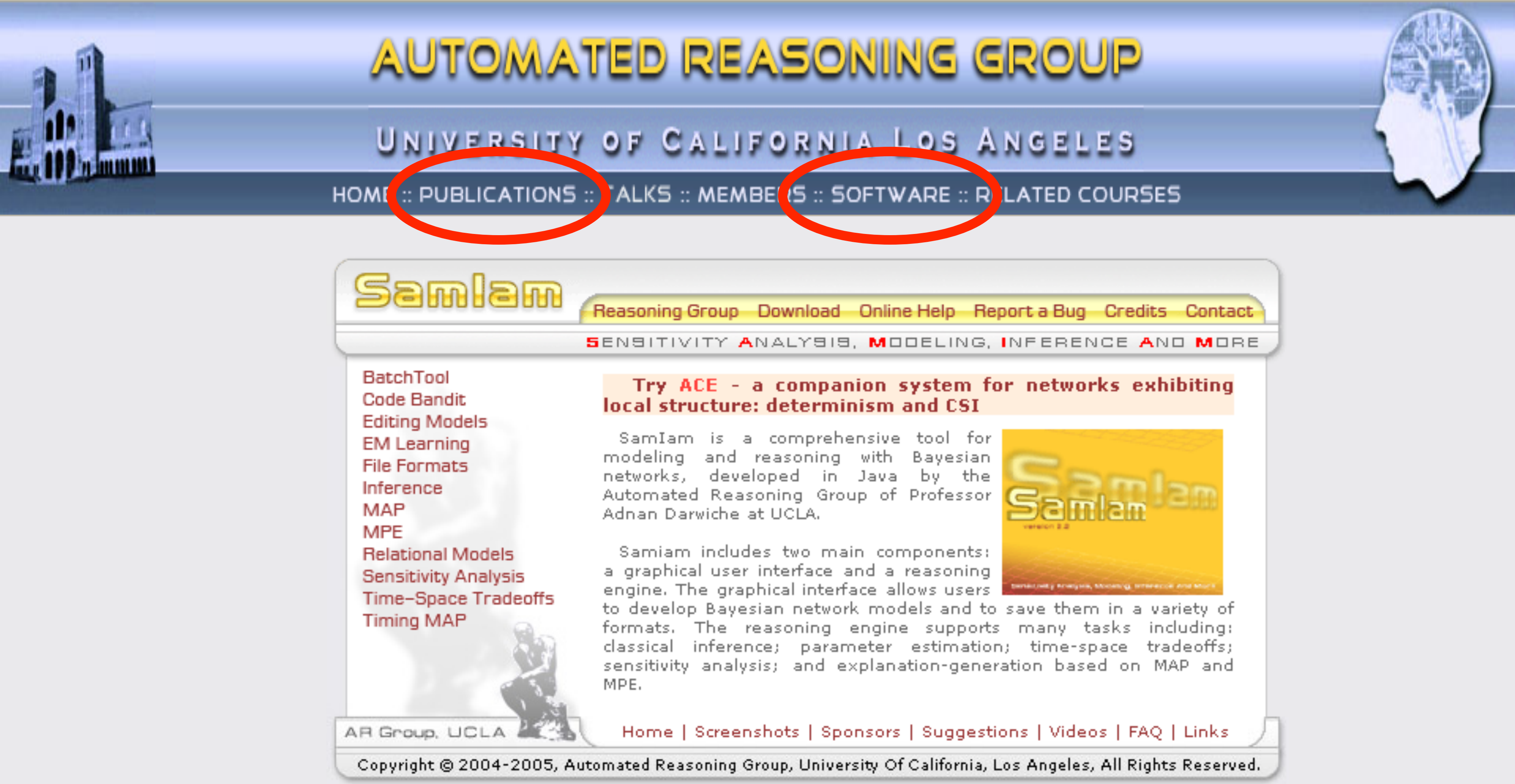
Logic in Machine Learning

Beyond data: Background knowledge

Beyond classical datasets

Tractable learning

PSDD



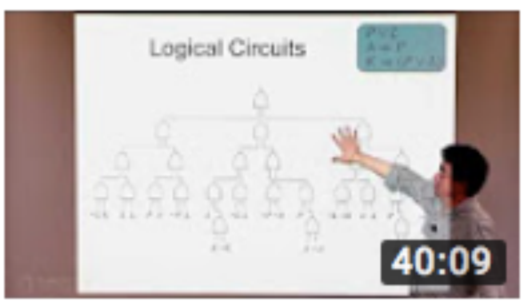
c2d: compiled cnf to decomposable and deterministic circuits
mini-c2d: open source + modular design (sat box)
sdd library: SDD generalized OBDD (similar to CUDD)
Ace: Compiles Bayesian networks into circuits

http://reasoning.cs.ucla.edu/samiam/

Position Talk/Paper on AI

Adnan Darwiche – On Model-Based versus Model-Blind
UCR School of Public Policy
2.1K views

Paper Title: Human-Level Intelligence or Animal-Like Abilities?



Tractable Learning in Structured Probability Spaces
Simons Institute
1.5K views