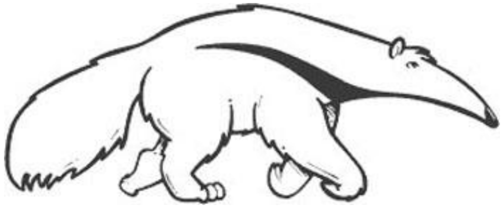
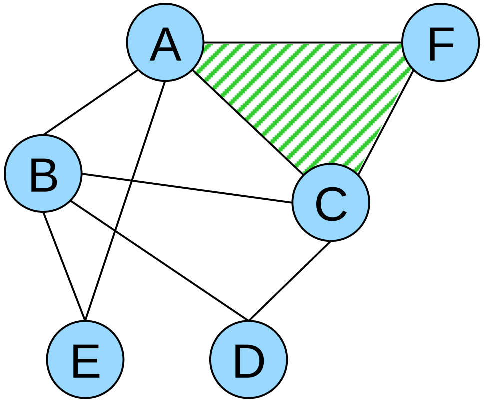
Probabilistic Reasoning Meets Heuristic Search

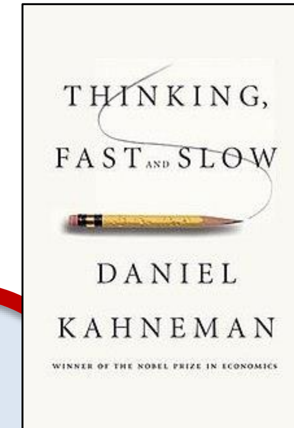
Rina Dechter

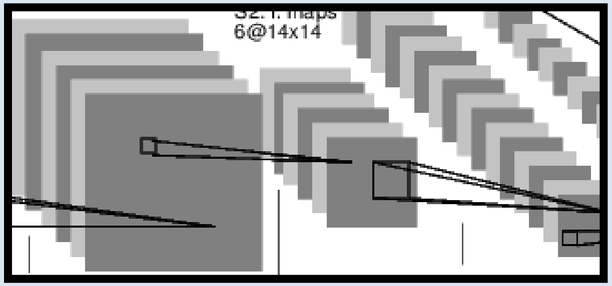
Dept. of Computer Science, UC Irvine





AI Renaissance

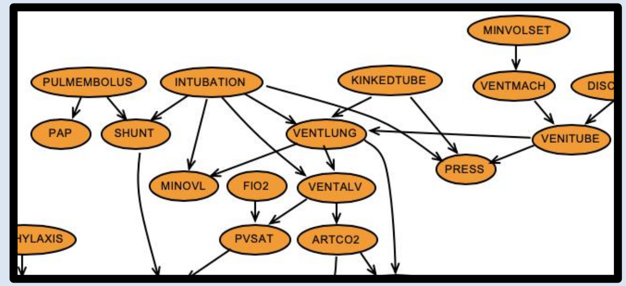




- Deep learning
  - Fast predictions
  - “Instinctive”
  - Model-free

Tools:

Tensorflow, PyTorch, …



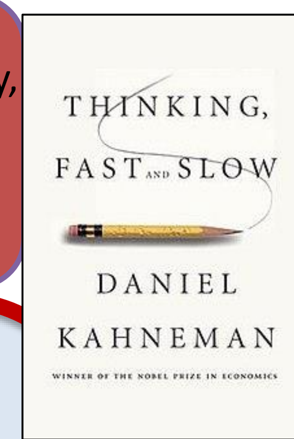
- Probabilistic models
  - Slow reasoning
  - “Logical / deliberative”
  - Model-based

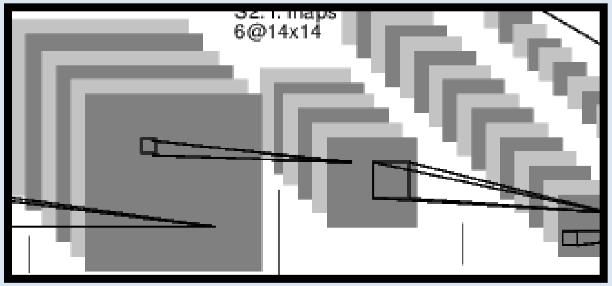
Tools:

Probabilistic programming, Markov Logic, …

AI Renaissance

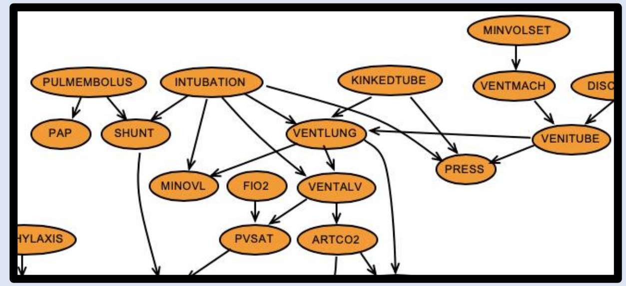
"If a machine does not have a model of reality, you cannot expect the machine to behave intelligently in that reality". (Pearl 2018)





- Deep learning
  - Fast predictions
  - "Instinctive"
  - Model-free

Tools:
Tensorflow, PyTorch, …

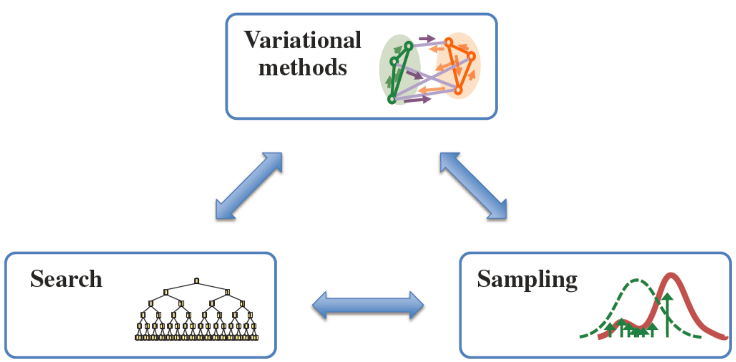


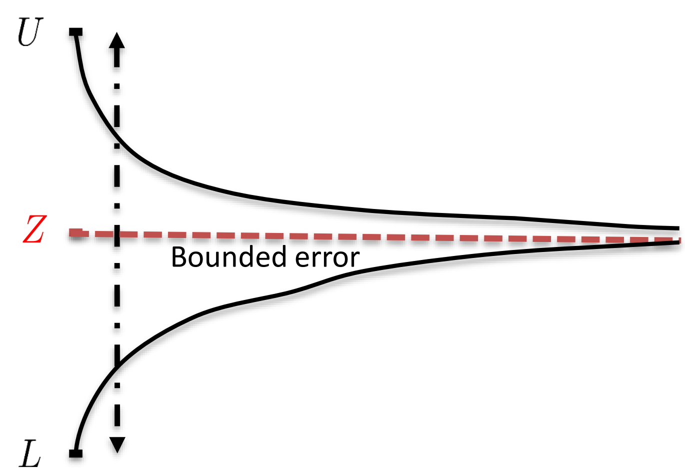
- Probabilistic models
  - Slow reasoning
  - "Logical / deliberative"
  - Model-based

Tools:
Probabilistic programming, Markov Logic, …

Outline

- Graphical models: definition, examples, methodology
- Inference and variational bounds
- AND/OR search spaces
- Variational bounds as search heuristics
- Combining methods: Heuristic Search for Marginal Map
- Conclusion





Graphical Models

- Describe structure in large problems
  - Large complex system $f(X)$
  - Made of "smaller", "local" interactions $f_\alpha(X_\alpha)$
  - Complexity emerges through interdependence
- More formally:

A *graphical model* consists of:

$X = \{X_1, \ldots, X_n\}$ -- variables (we'll assume discrete)

$D = \{D_1, \ldots, D_n\}$ -- domains

$F = \{f_{\alpha_1}, \ldots, f_{\alpha_m}\}$ -- (non-negative) functions or "factors"

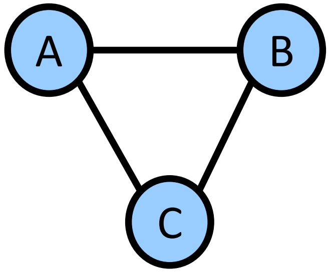
- Example:

$$f(A,B,C) = f(A,B)f(A,C)f(B,C)$$

A	B	f(A,B)
0	0	**0.24**
0	1	**0.56**
1	0	**1.1**
1	1	**1.2**

...

B	C	f(B,C)
0	0	**0.12**
0	1	**0.36**
1	0	**0.3**
1	1	**1.8**



Example:

$A \in \{0,1\}$

$B \in \{0,1\}$

$C \in \{0,1\}$

$f_{AB}(A,B), \quad f_{BC}(B,C)$

Graph Visualiization: Primal Graph

A *graphical model* consists of:

$X = \{X_1, \ldots, X_n\}$ -- variables

$D = \{D_1, \ldots, D_n\}$ -- domains

$F = \{f_{\alpha_1}, \ldots, f_{\alpha_m}\}$ -- functions or "factors"

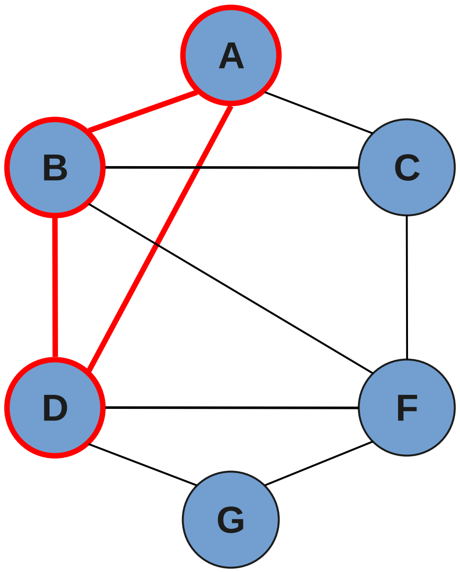
and a *combination operator (product, sum…)*

Primal graph:

variables → nodes

factors → cliques

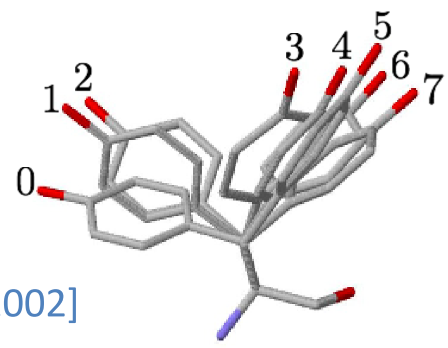
$$F(A,B,C,D,F,G) = f_1(A,B,D) + f_2(D,F,G) + f_3(B,C,F) + f_4(A,C)$$



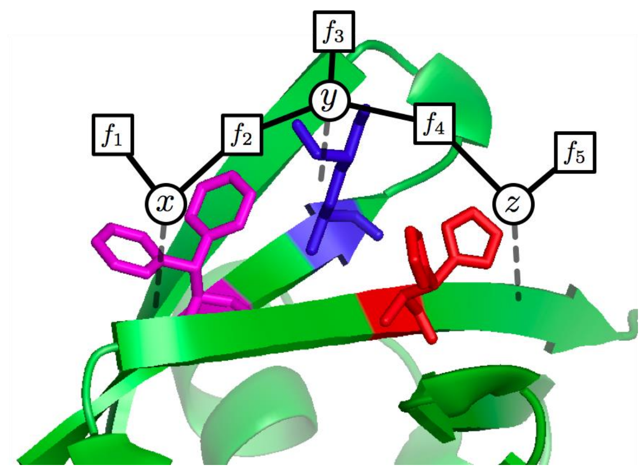
Graphical Models

- Describe structure in large problems
  - Large complex system $f(X)$
  - Made of "smaller", "local" interactions $f_\alpha(X_\alpha)$
  - Complexity emerges through interdependence
- Examples & Tasks
  - Maximization (**MAP**): compute the most probable configuration

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \prod_\alpha f_\alpha(\mathbf{x}_\alpha) \qquad f(\mathbf{x}^*) = \max_{\mathbf{x}} \prod_\alpha f_\alpha(\mathbf{x}_\alpha)$$



Phenylalanine



[Yanover & Weiss 2002]

Graphical Models

- Describe structure in large problems
  - Large complex system $f(X)$
  - Made of "smaller", "local" interactions $f_\alpha(X_\alpha)$
  - Complexity emerges through interdependence

- Examples & Tasks
  - Summation & marginalization

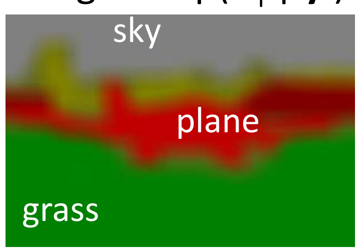
"partition function"

$$p(x_i) = \frac{1}{Z} \sum_{\mathbf{x} \setminus x_i} \prod_\alpha f_\alpha(\mathbf{x}_\alpha) \quad \text{and} \quad Z = \sum_{\mathbf{x}} \prod_\alpha f_\alpha(\mathbf{x}_\alpha)$$

Observation **y**

$\Longrightarrow$

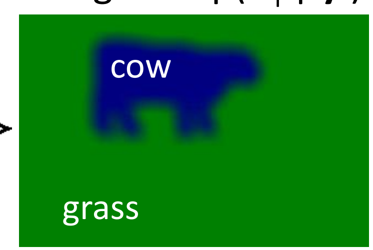
Marginals p(x_i | **y**)



Observation **y**

$\Longrightarrow$

Marginals p(x_i | **y**)



e.g., [Plath et al. 2009]

Graphical Models

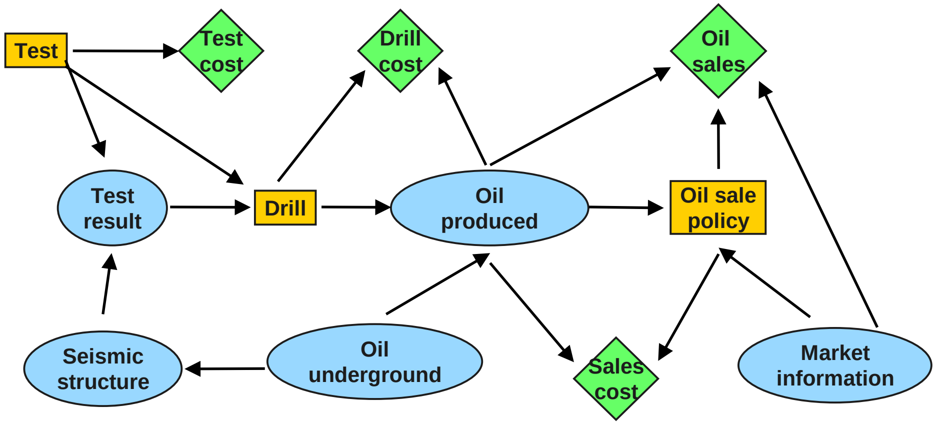
- Describe structure in large problems
  - Large complex system $f(X)$
  - Made of "smaller", "local" interactions $f_\alpha(X_\alpha)$
  - Complexity emerges through interdependence
- Examples & Tasks
  - Mixed inference (marginal MAP, MEU, …)

$$f(\mathbf{x}_M^*) = \max_{\mathbf{x}_M} \sum_{\mathbf{x}_S} \prod_\alpha f_\alpha(\mathbf{x}_\alpha)$$

Influence diagrams & optimal decision-making

(the "oil wildcatter" problem)

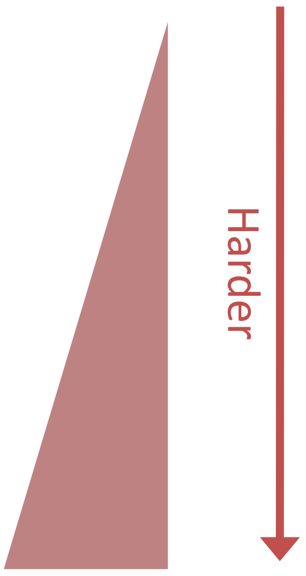
e.g., [Raiffa 1968; Shachter 1986]



Probabilistic Reasoning Problems

- Tasks:

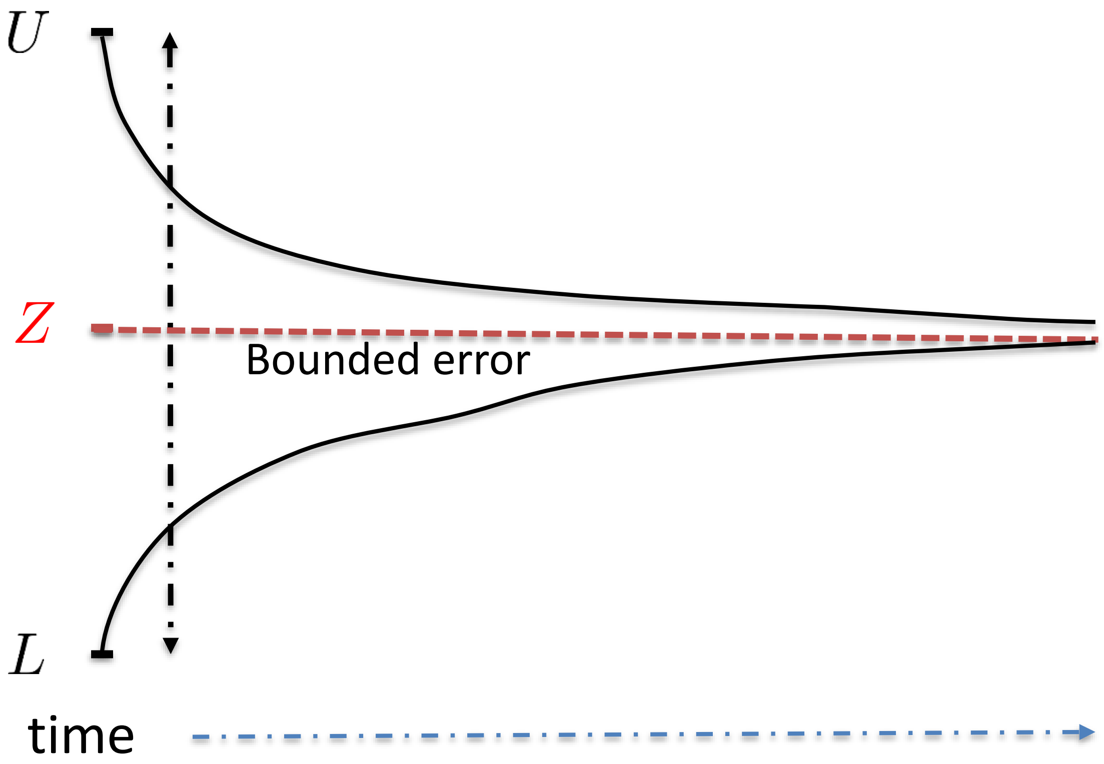
Max-Inference (most likely config.)	$f(\mathbf{x}^*) = \max_{\mathbf{x}} \prod_{\alpha} f_\alpha(\mathbf{x}_\alpha)$
Sum-Inference (data likelihood)	$Z = \sum_{\mathbf{x}} \prod_{\alpha} f_\alpha(\mathbf{x}_\alpha)$
Mixed-Inference (optimal prediction)	$f(\mathbf{x}_M^*) = \max_{\mathbf{x}_\mathrm{M}} \sum_{\mathbf{x}_S} \prod_{\alpha} f_\alpha(\mathbf{x}_\alpha)$



  – Combinatorial search / counting queries
  – Exact reasoning NP-complete (or worse)

Anytime Bounds

- Desiderata
  - Meaningful confidence interval
  - Responsive
  - Complete

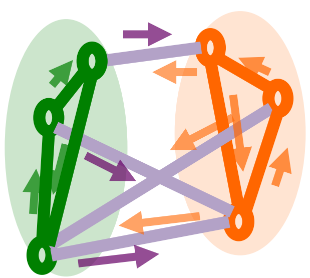


Approximate Reasoning

- Three major paradigms
  - Effective at different types of problems
  - Each can exploit the graph… but more

Variational methods

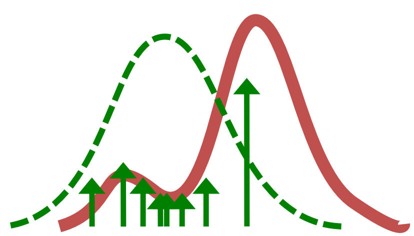
Reason over small subsets of variables at a time



- Bounds
- Responsive
- Complete

(Monte Carlo) Sampling

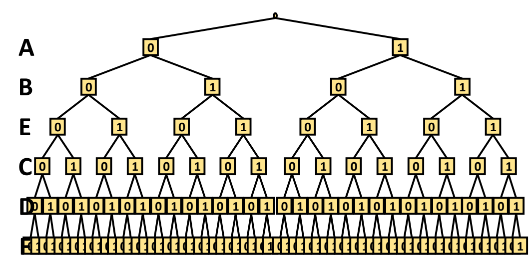
Use randomization to estimate averages over the state space



- Bounds
- Responsive
- Complete

(Heuristic) Search

Structured enumeration over all possible states

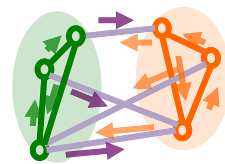


- Bounds
- Responsive
- Complete

Combining Approaches

A powerful tool:
The graphs

Inference, Variational methods



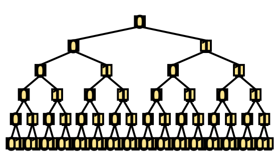
Bucket-elimination
weighted mini-bucket (WMB)
[Dechter 1999, Dechter and Rish, 2003
Liu and Ihler, ICML 2011]

provide heuristic

provide proposal

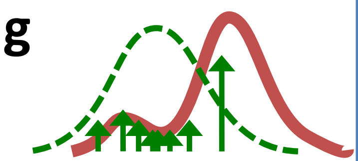
WMB-IS
[Liu et al., NIPS 2015]

Search



refine proposal

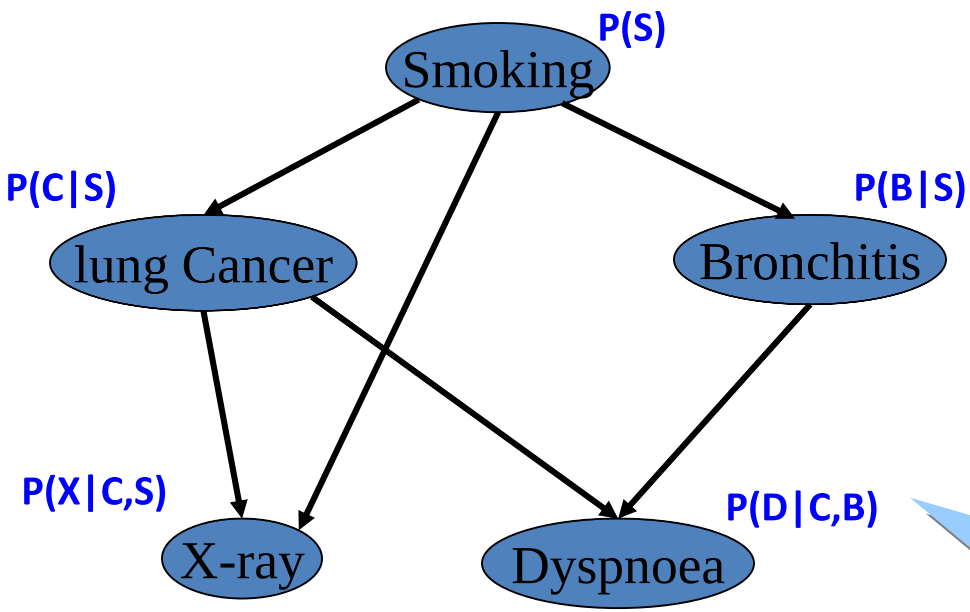
Sampling



AND/OR search (AODFS)
[Marinescu et al 2009, Lou et al., AAAI 2017]

dynamic importance sampling (DIS)
[Lou et al., NIPS 2017]

Bayesian Networks (Pearl 1988)



BN = (G, Θ)

CPD:

C	B	P(D|C,B)	
0	0	0.1	0.9
0	1	0.7	0.3
1	0	0.8	0.2
1	1	0.9	0.1

P(S, C, B, X, D) = ***P(S) P(C|S) P(B|S) P(X|C,S) P(D|C,B)***

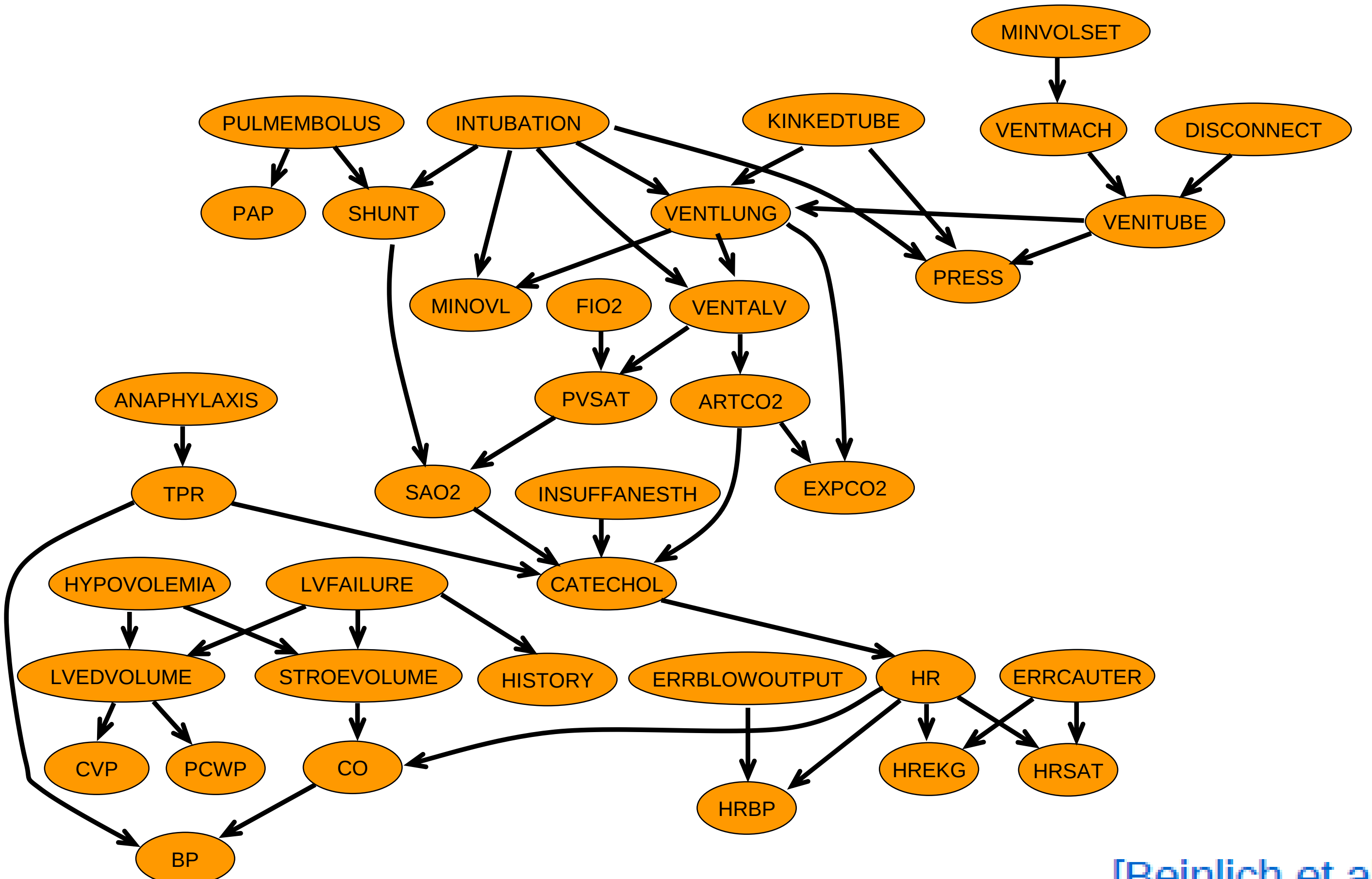
Combination: Product
Marginalization: sum/max

- Posterior marginals, probability of evidence, MPE
- P(D= 0) = $\sum_{S,L,B,X}$ P(S)· P(C|S)· P(B|S)· P(X|C,S)· P(D|C,B

MAP(P)= $max_{S,L,B,X}$ P(S)· P(C|S)· P(B|S)· P(X|C,S)· P(D|C,B)

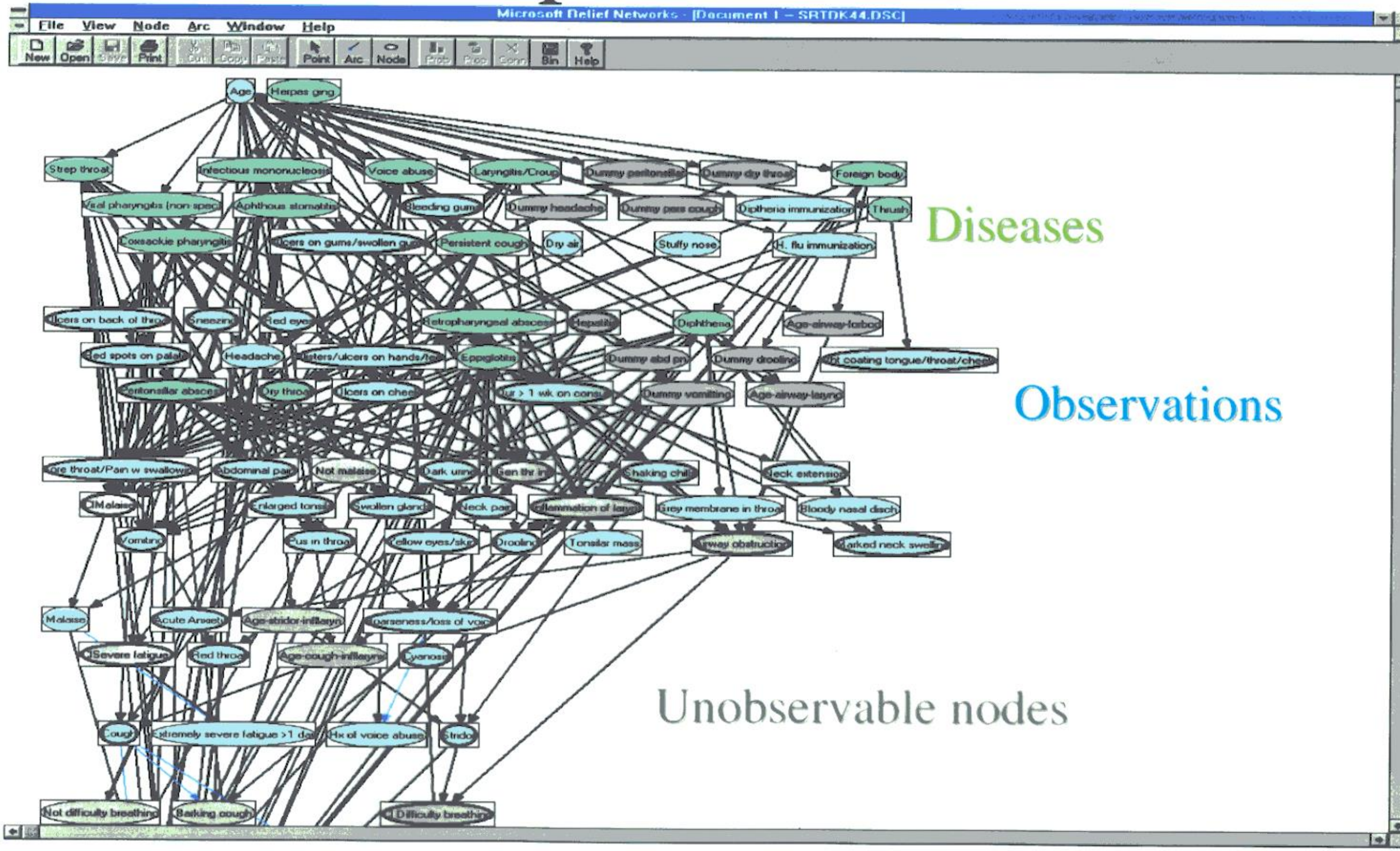
Monitoring Intensive-Care Patients

The “alarm” network - 37 variables, 509 parameters (instead of 2^{37})



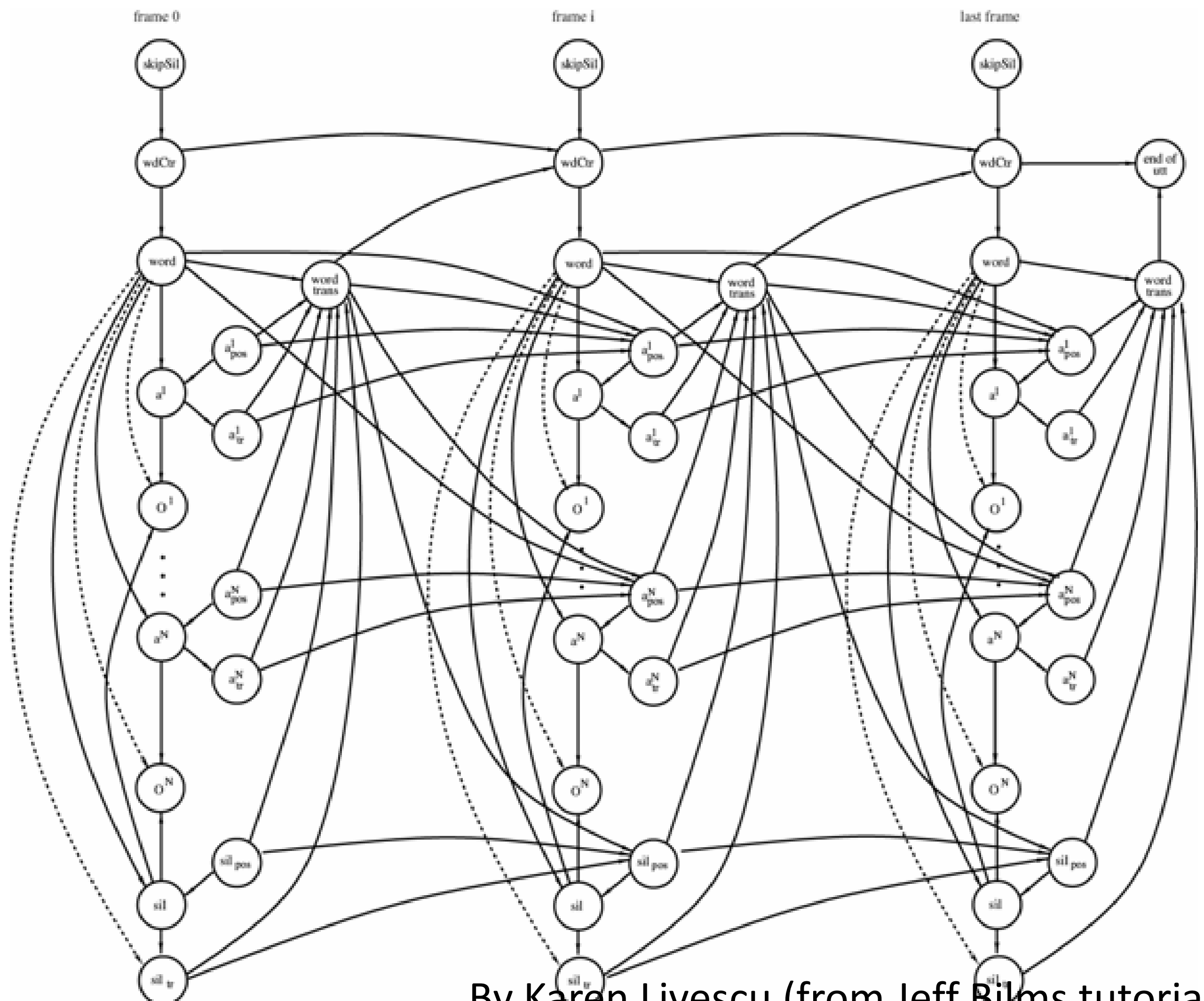
[Beinlich et al., 1989]

Chief Complaint: Sore Throat



KI

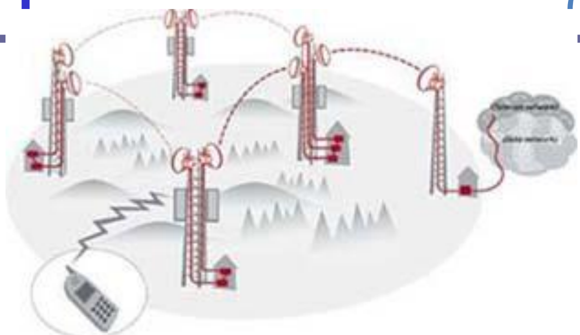
Phone-free Articulatory Graph



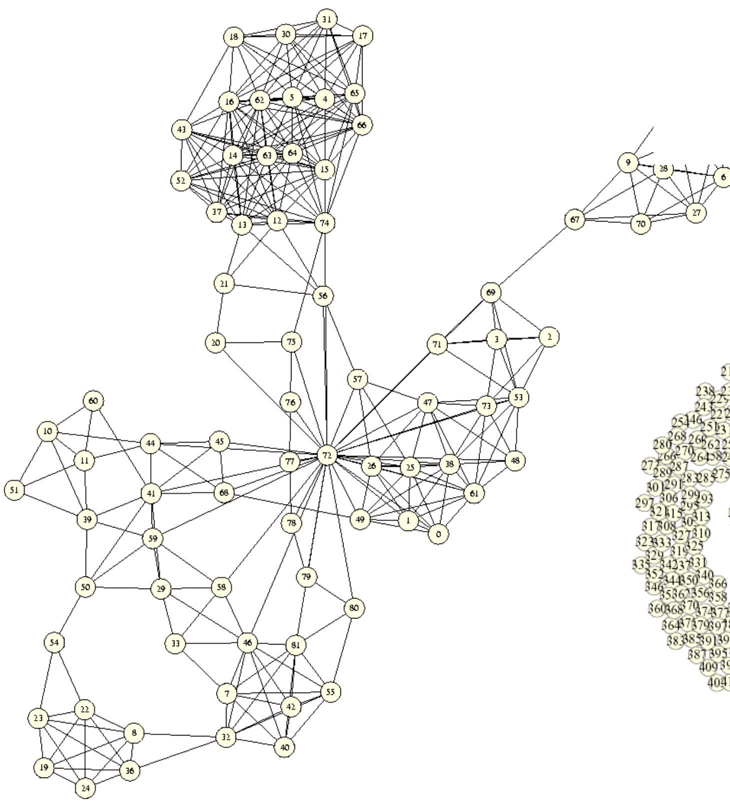
By Karen Livescu (from Jeff Bilms tutorial

Radio Link Frequency Assignment Problem

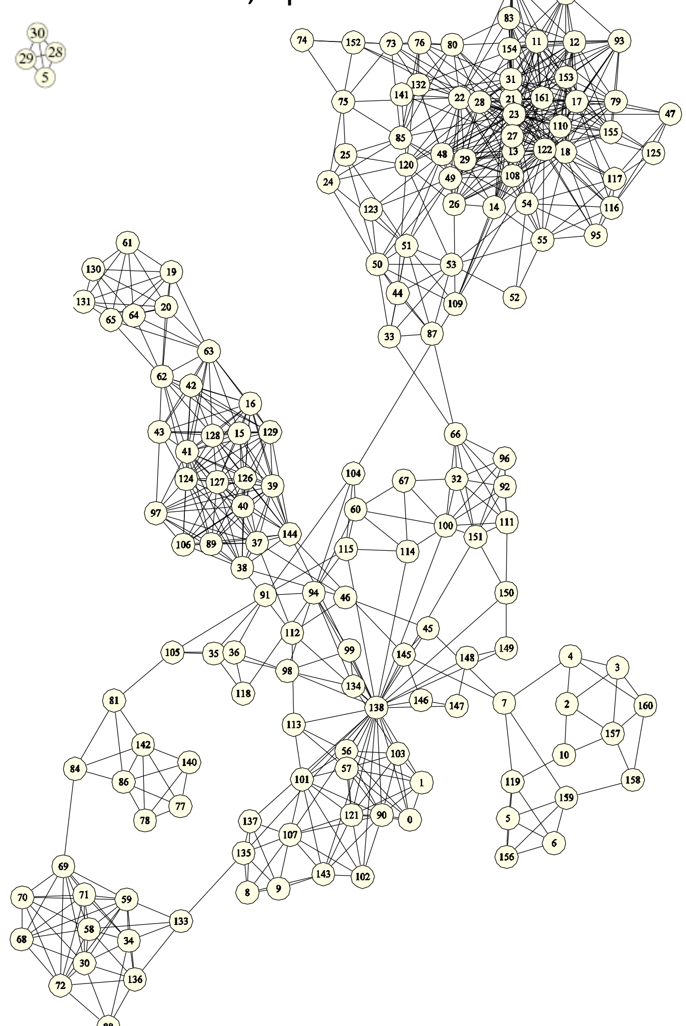
(Cabon et al., *Constraints* 1999) (Koster *et al.*, *4OR* 2003)



CELAR SCEN-06
n=100, d=44,
m=350, optimum=3389

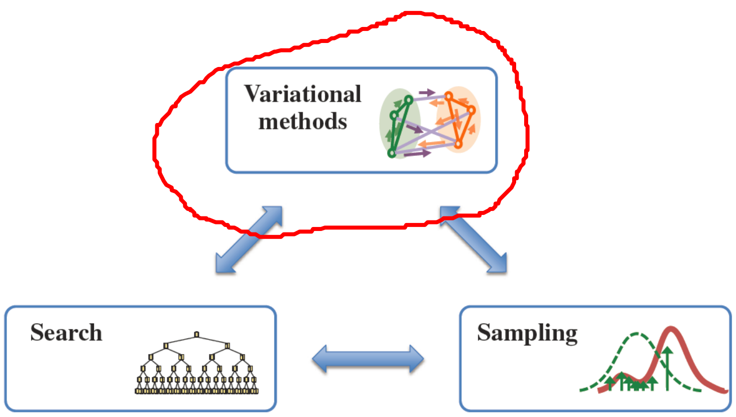


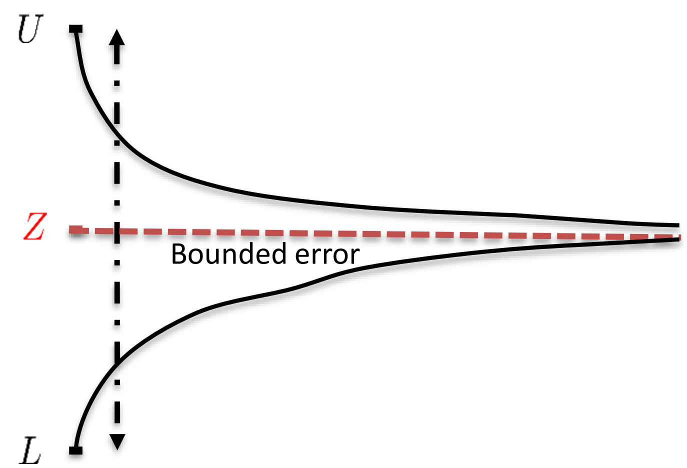
- CELAR SCEN-07r
  n=162, d=44,
  m=764, optimum=343592



Outline

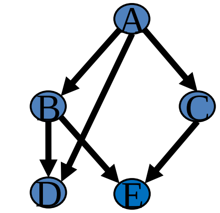
- Graphical models: definition, examples, methodology
- Inference and variational bounds
- AND/OR search spaces
- Variational bounds as search heuristics
- Combining methods: Heuristic Search for Marginal Map
- Conclusion





Marginals by Bucket Elimination

(Dechter 1999)



$$P(A\mid E=0)=\alpha\sum_{E=0,D,C,B} P(A)\cdot P(B\mid A)\cdot P(C\mid A)\cdot P(D\mid A,B)\cdot P(E\mid B,C)$$

$\sum_b \prod$ ← Elimination operator

bucket B: $P(b|a)$ $P(d|b,a)$ $P(e|b,c)$

bucket C: $P(c|a)$ $\lambda_{B\to C}(a,d,c,e)$

bucket D: $\lambda_{C\to D}(a,d,e)$

bucket E: $e=0$ $\lambda_{D\to A}(a,e)$

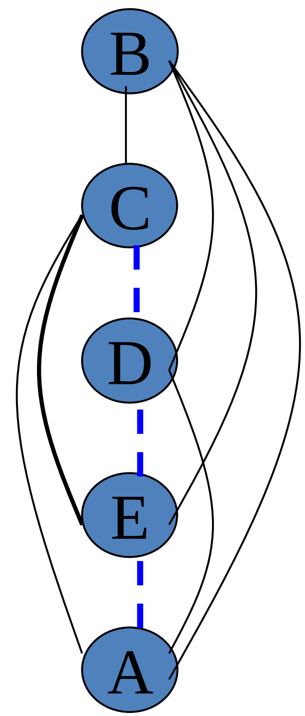
bucket A: $P(a)$ $\lambda_{E\to A}(a)$

$P(e=0)$

$P(a|e=0)$

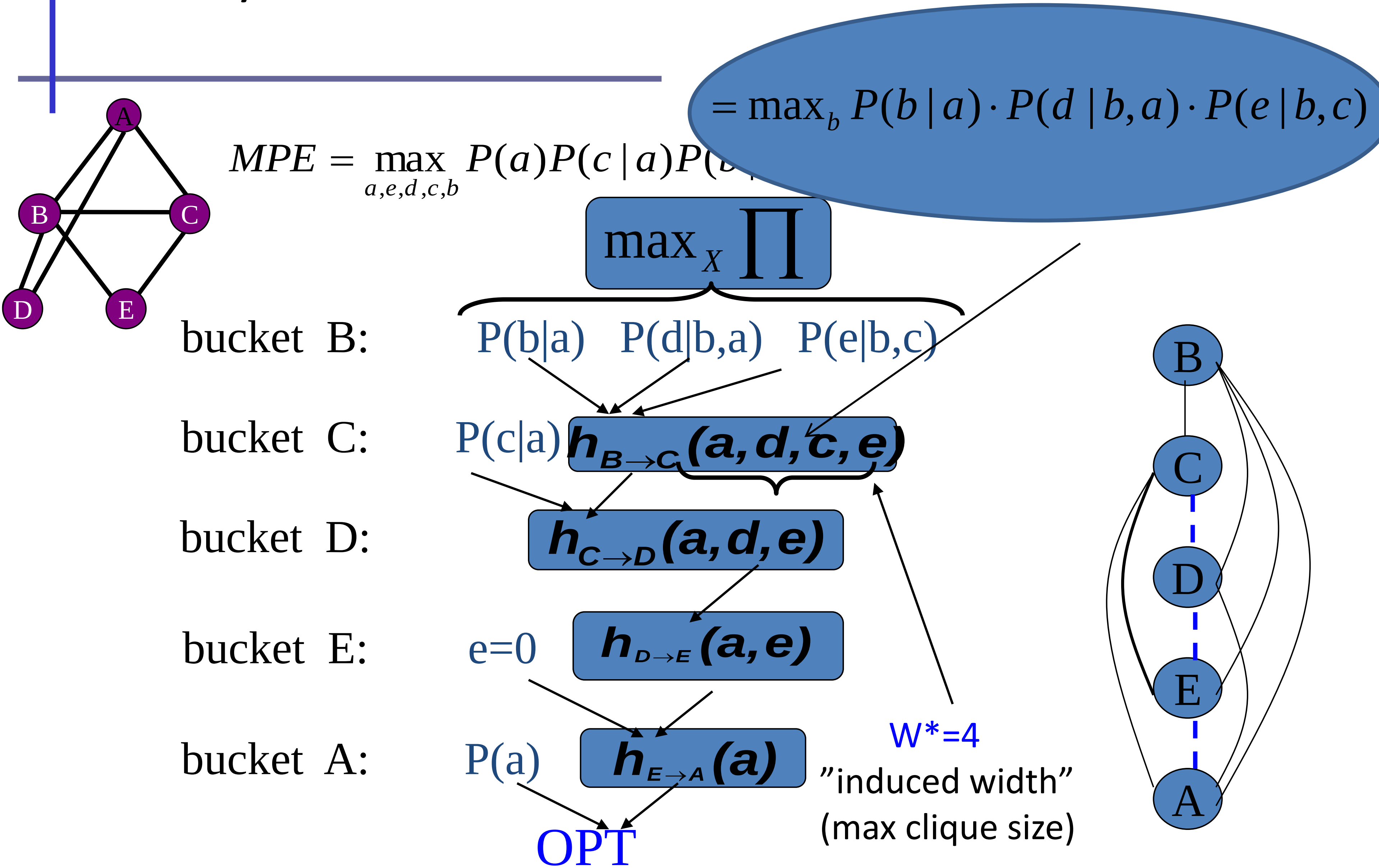
W*=4

"induced width" (max clique size)



Complexity time and space O(nk^{w*+1})

MAP by Bucket Elimination



Algorithm BE-mpe (Dechter 1996, Bertele and Briochi, 1977)

Decoding the Optimal-Tuple

5. $b' = \arg\max_b P(b \mid a') \times P(d' \mid b, a') \times P(e' \mid b, c')$

4. $c' = \arg\max_c P(c \mid a') \times h^B(a', d', c, e')$

3. $d' = \arg\max_d h^C(a', d, e')$

2. $e' = 0$

1. $a' = \arg\max_a P(a) \cdot h^E(a)$

B: P(b|a) P(d|b,a) P(e|b,c)

C: P(c|a) $h^B(a,d,c,e)$

D: $h^C(a,d,e)$

E: e=0 $h^D(a,e)$

A: P(a) $h^E(a)$

Return (a', b', c', d', e')

Complexity of Bucket Elimination;

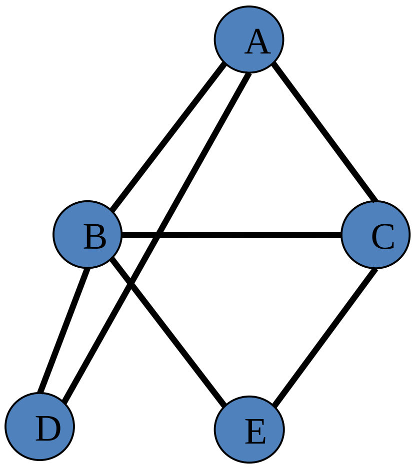
Bucket Elimination is **time** and **space**

$$O\left(r \exp(w^*(d))\right)$$

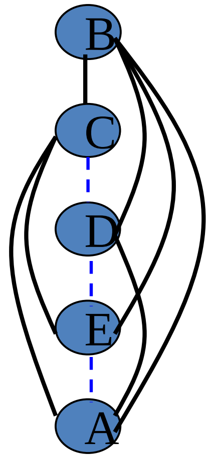
$w^*(d)$ – the induced width of graph along ordering d

r = number of functions

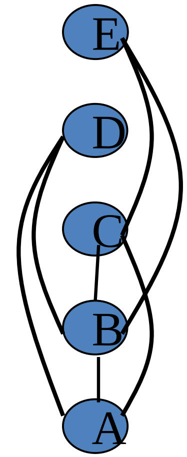
The effect of the ordering:



"Moral" graph



$w^*(d_1) = 4$



$w^*(d_2) = 2$

Finding the smallest induced width is hard**!**

Bucket and Mini-Bucket Elimination

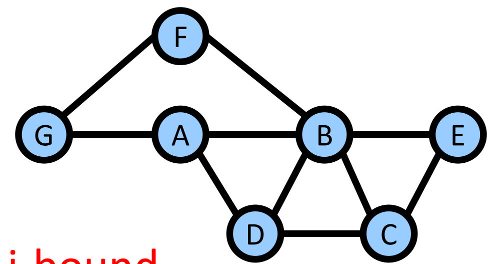
[Dechter 1999; Dechter & Rish, 2003, Liu & Ihler 2011]

$\lambda^D(A) = \max_D f(A,D)$

$\lambda^D(B,C) = \max_D [f(B,D) * f(C,D)]$

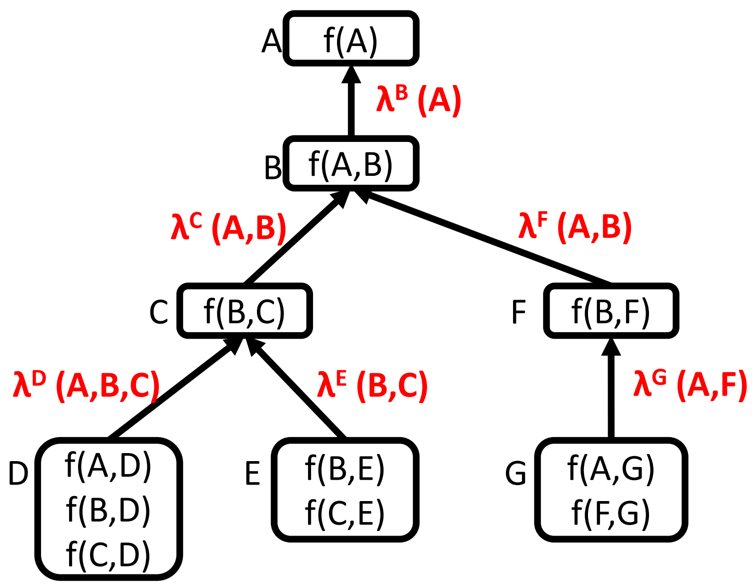
$$\sum_X F(X)$$

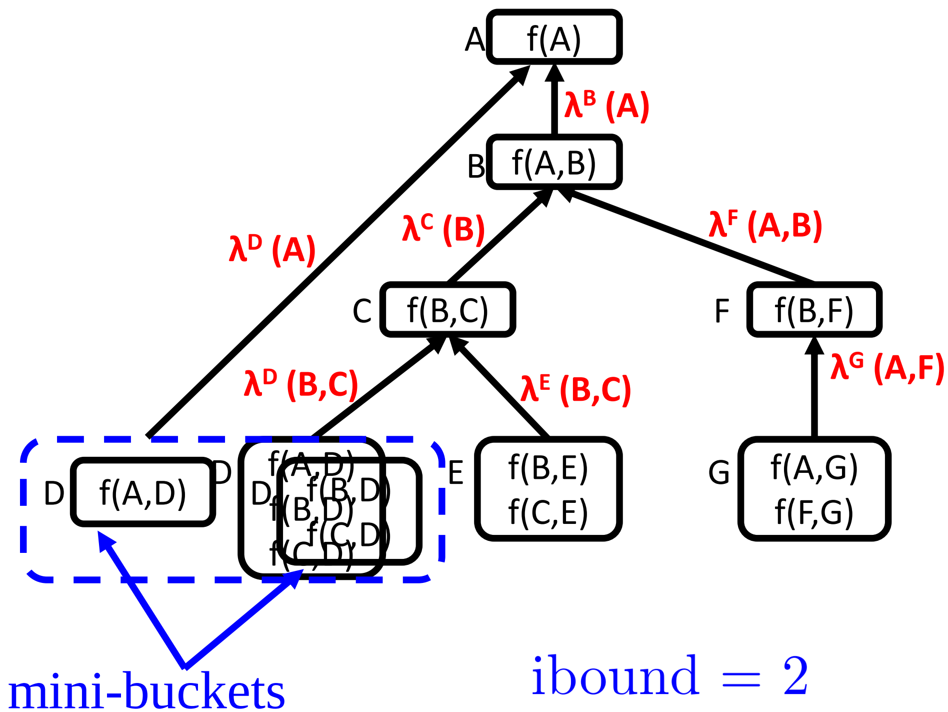
$$F(\mathbf{X}) = f(A)f(A,B)f(A,D)f(A,G)f(B,C)f(B,D)$$
$$f(B,E)f(B,F)f(C,D)f(C,E)f(E,G)$$



Time and space exponential in the induced-width/tree-width

Exponential in i-bound





$\lambda^D(A,B,C) =$
$\max_D [f(A,D) * f(B,D) * f(C,D)]$

$\text{ibound} = 2$

Mini-Bucket Approximation

For optimization

Split a bucket into mini-buckets —> bound complexity

bucket (X) =

$$\left\{ f_1,\ f_2,\ \ldots\ f_r,\ f_{r+1},\ \ldots\ f_n \right\}$$

$$\lambda_X(\cdot) = \max_x \prod_{i=1}^{n} f_i(x,\ldots)$$

$$\left\{ f_1,\ \ldots\ f_r \right\} \qquad \left\{ f_{r+1},\ \ldots\ f_n \right\}$$

$$\lambda_{X,1}(\cdot) = \max_x \prod_{i=1}^{r} f_i(x,\ldots) \qquad \lambda_{X,2}(\cdot) = \max_x \prod_{i=r+1}^{n} f_i(x,\ldots)$$

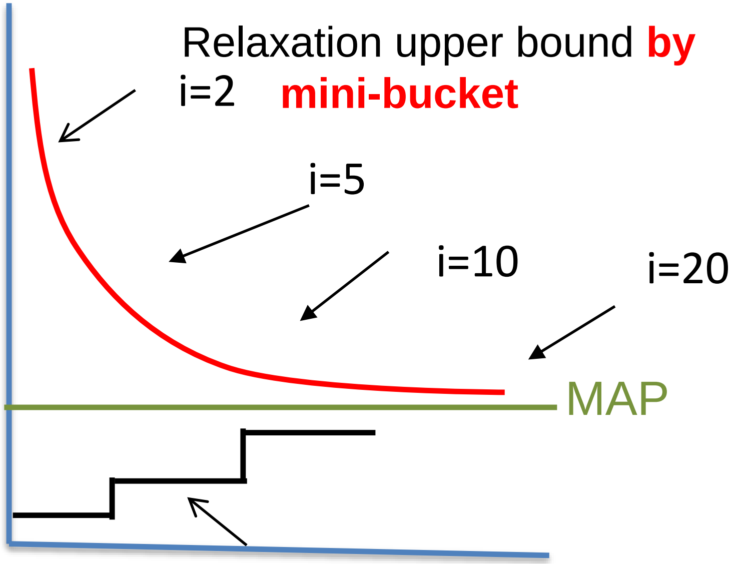
$$\lambda_X(\cdot) \ \leq\ \lambda_{X,1}(\cdot)\ \lambda_{X,2}(\cdot)$$

Exponential complexity decrease: $O(e^n) \longrightarrow O(e^r) + O(e^{n-r})$

Properties of Mini-Bucket Eliminaton

(For optimization)

- Bounding from above and below

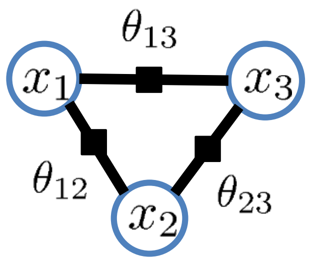


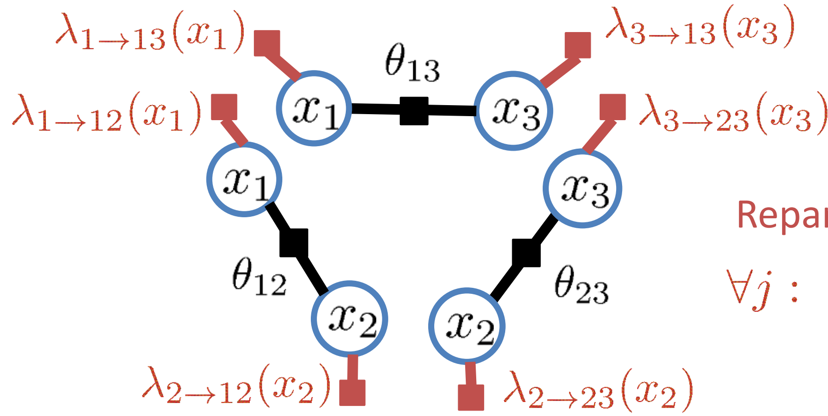
Consistent solutions (greedy search)

- Complexity: O(r exp(i)) time and O(exp(i)) space.
- Accuracy: determined by Upper/Lower bound.
- As i-bound increases, both accuracy and complexity increase.
- Possible use of mini-bucket approximations:
  - As anytime algorithms
  - As heuristics in search
-

Tightening the Bound

Add factors that "adjust" each local term, but cancel out in total





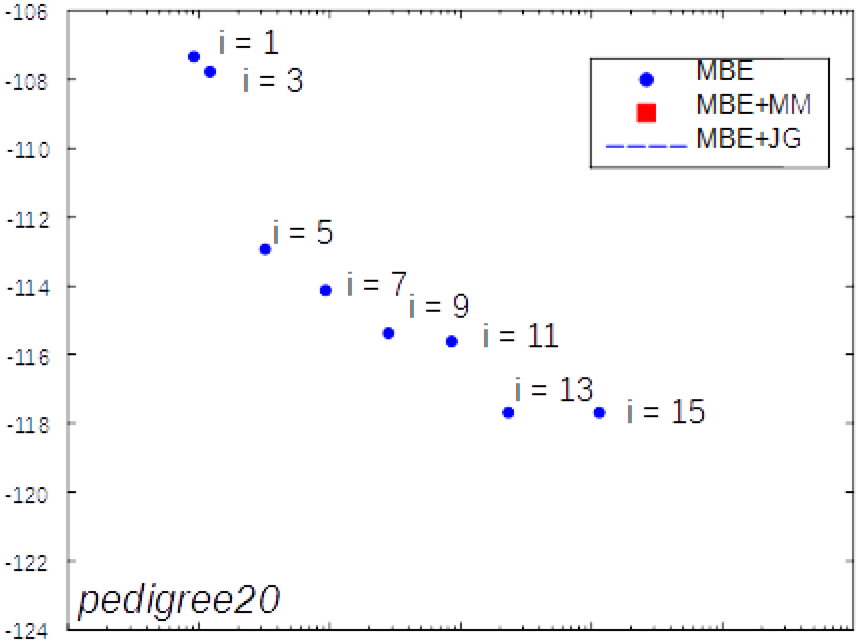
Reparameterization:

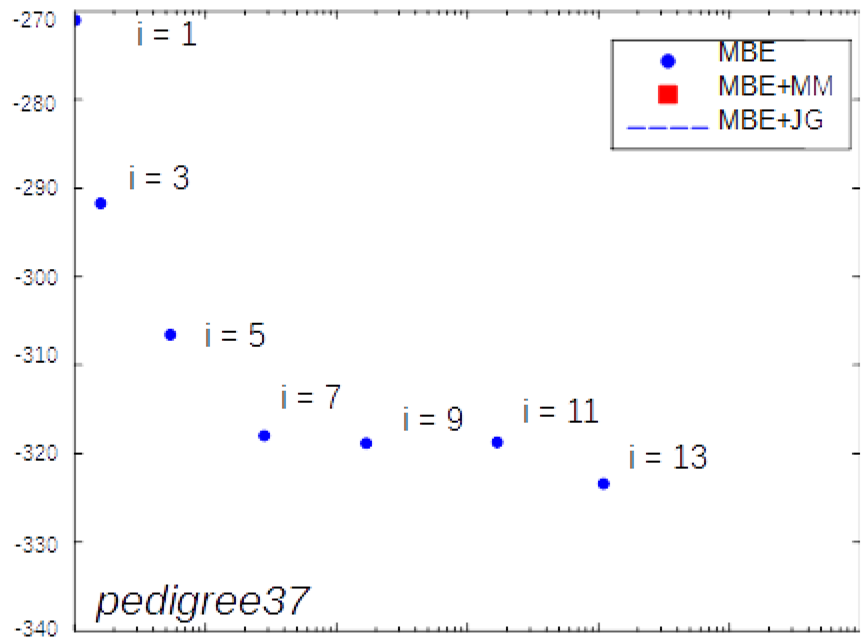
$$\forall j : \sum_{\alpha \ni j} \lambda_{j\to\alpha}(x_j) = 0$$

$$\log f(\mathbf{x}^*) = \max_{\mathbf{x}} \sum_{\alpha} \theta_\alpha(\mathbf{x}_\alpha) \quad \leq \min_{\{\lambda_{i\to\alpha}\}} \sum_{\alpha} \max_{\mathbf{x}_\alpha} \left[\theta_\alpha(\mathbf{x}_\alpha) + \sum_{i\in\alpha} \lambda_{i\to\alpha}(x_i) \right]$$

- Bound solution using decomposed optimization
- Solve independently: optimistic bound
- Tighten the bound by re-parameterization
  - Enforces lost equality constraints using Lagrange multipliers

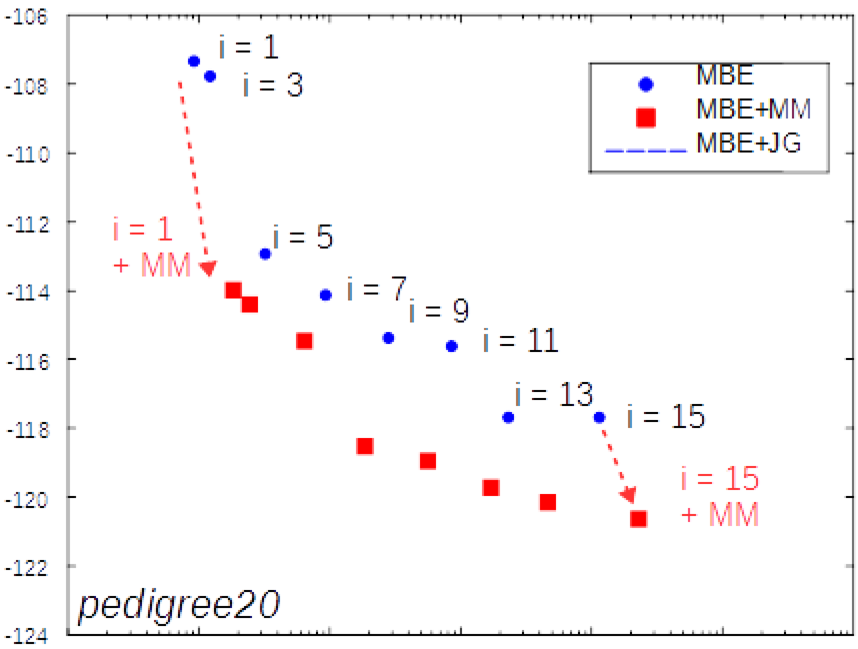
Anytime Approximation

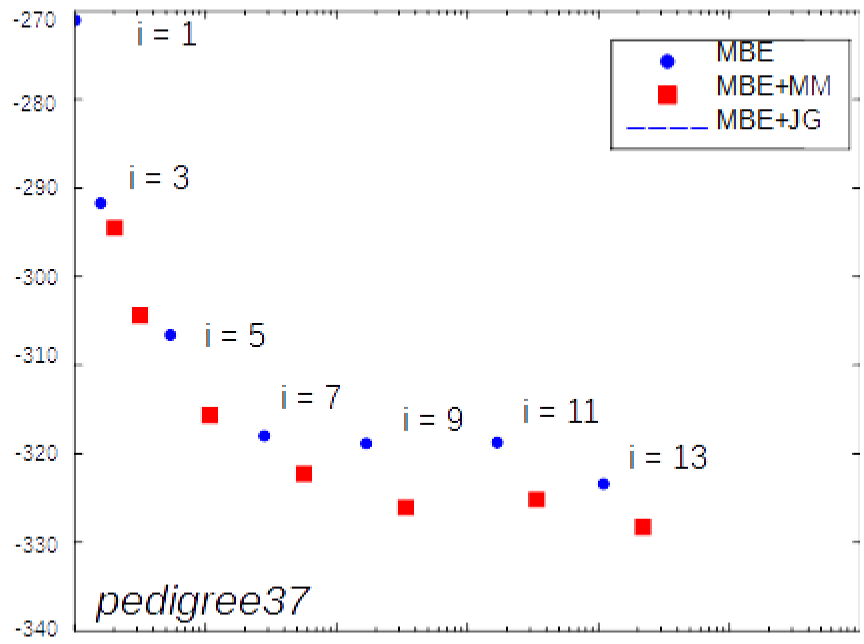




- Can tighten the bound in various ways
  - Cost-shifting (improve consistency between cliques)
  - Increase i-bound (higher order consistency)
- Simple moment-matching step improves bound significantly

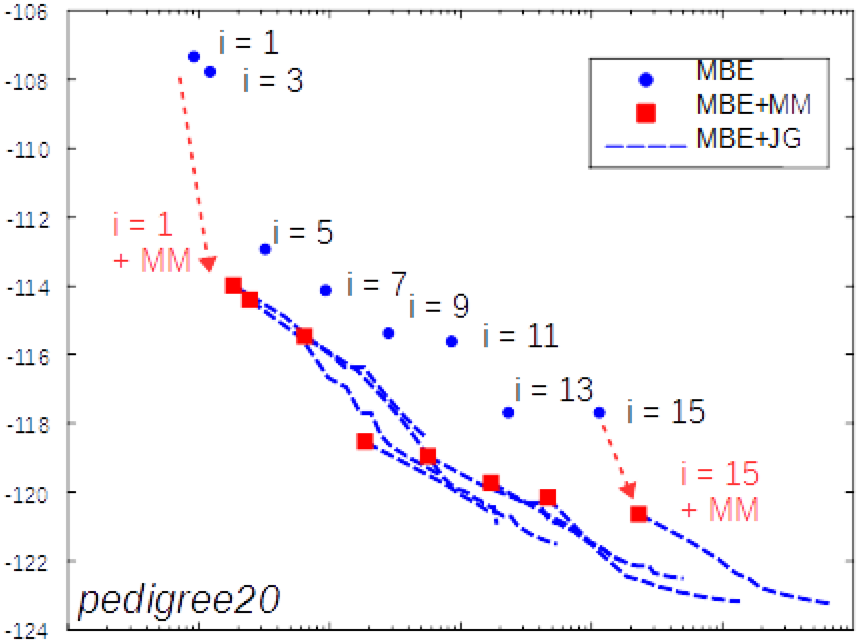
Anytime Approximation

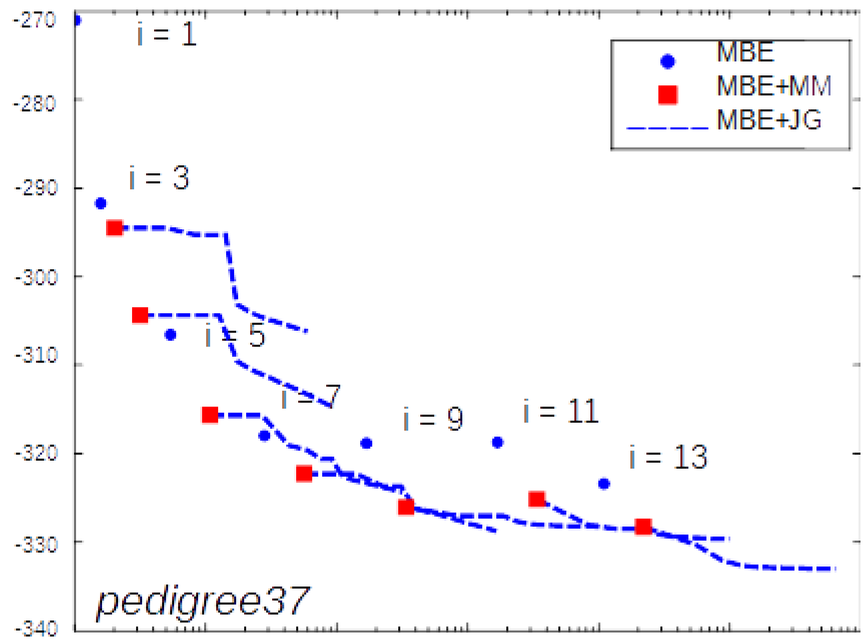




- Can tighten the bound in various ways
  - Cost-shifting (improve consistency between cliques)
  - Increase i-bound (higher order consistency)
- Simple moment-matching step improves bound significantly

Anytime Approximation





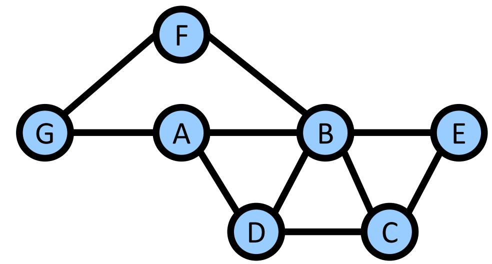
- Can tighten the bound in various ways
  - Cost-shifting (improve consistency between cliques)
  - Increase i-bound (higher order consistency)
- Simple moment-matching step improves bound significantly

Bucket Elimination (BE) and WMB

Holder inequality facilitated weighted MB for summation

[Dechter 1999, Ihler et. Al. 2013]

$$F(\mathbf{X}) = f(A)f(A,B)f(A,D)f(A,G)f(B,C)f(B,D)$$
$$f(B,E)f(B,F)f(C,D)f(C,E)f(E,G)$$



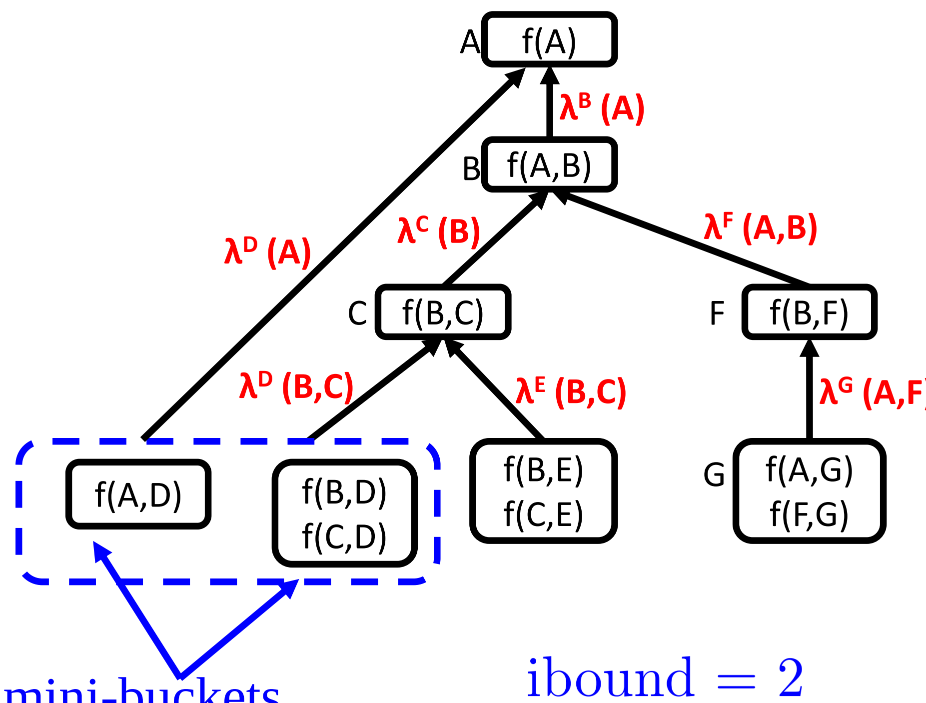
Pros:

- Computationally bounded
- Gives upper or lower bound
- Cost-shifting Message passing
- improves bound

Cons:

- Not anytime!

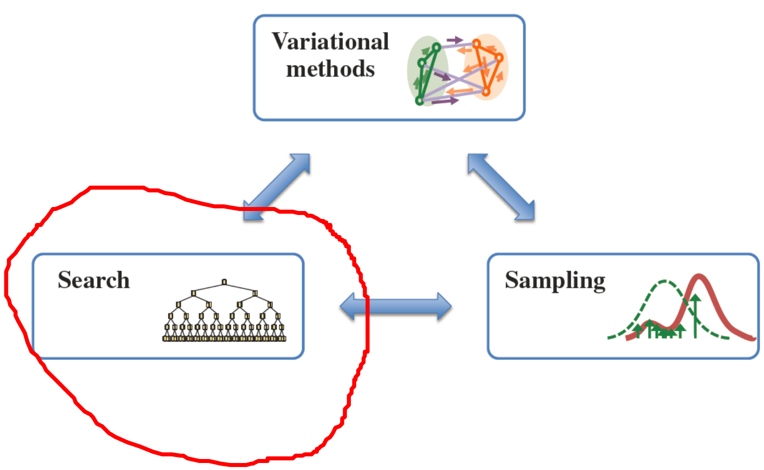
not asymp. tight w/o more memory

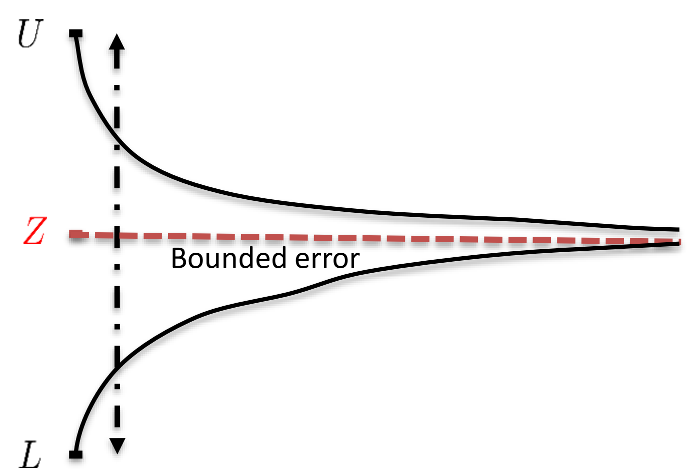


$\text{ibound} = 2$

Outline

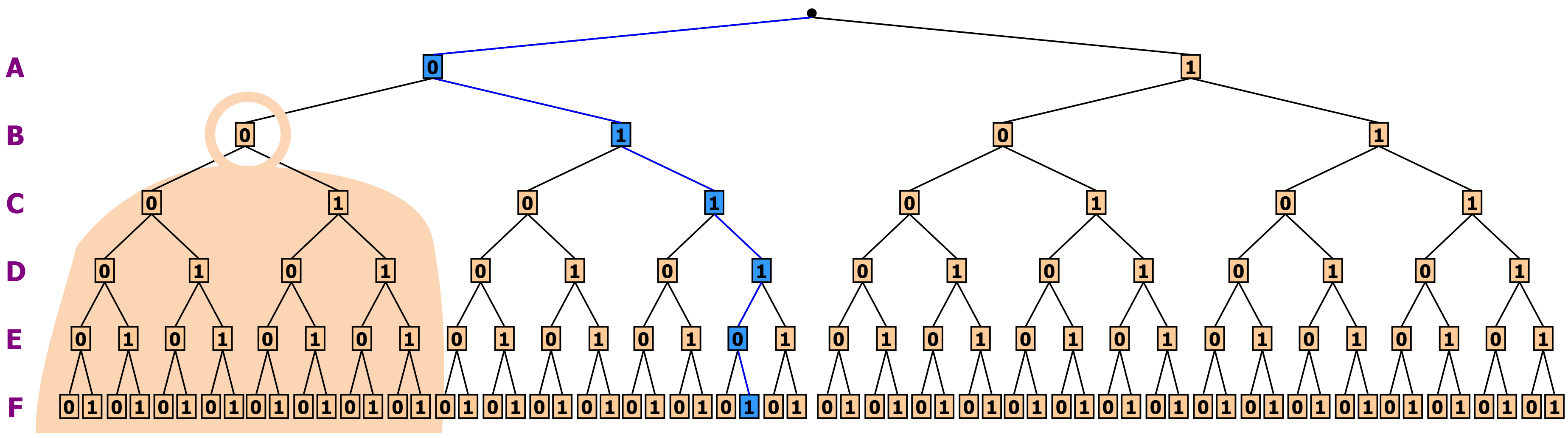
- Graphical models: definition, examples, methodology
- Inference and variational bounds
- AND/OR search spaces
- Variational bounds as search heuristics
- Combining methods: Heuristic Search for Marginal Map
- Conclusion





Search Trees and Tasks

- Organize / structure the state space
  - Leaf nodes = model configurations
  - “Value” of a node = optimal of sumr of configurations below



Potential search spaces

$$F^* = \min_x \sum_\alpha f_\alpha(x_\alpha)$$

A	B	f_1
0	0	2
0	1	0
1	0	1
1	1	4

A	C	f_2
0	0	3
0	1	0
1	0	0
1	1	1

A	E	f_3
0	0	0
0	1	3
1	0	2
1	1	0

A	F	f_4
0	0	2
0	1	0
1	0	0
1	1	2

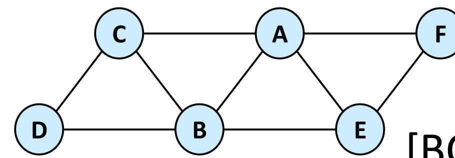
B	C	f_5
0	0	0
0	1	1
1	0	2
1	1	4

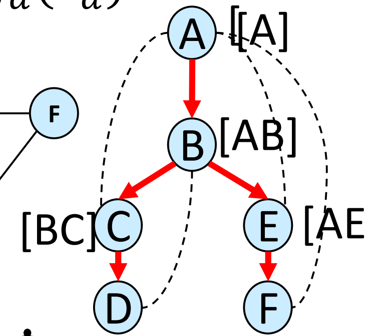
B	D	f_6
0	0	4
0	1	2
1	0	1
1	1	0

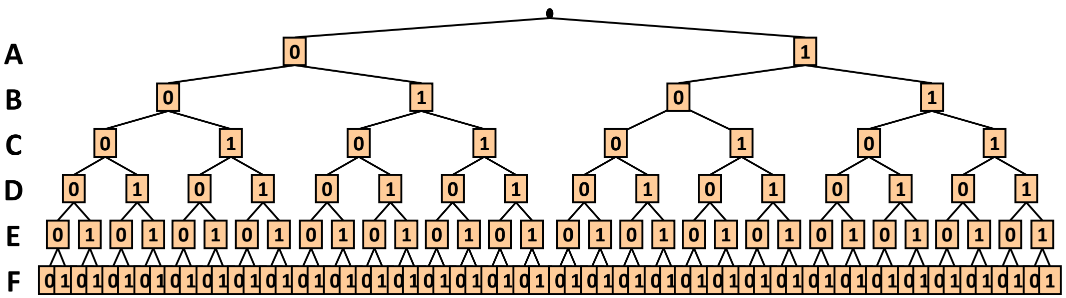
B	E	f_7
0	0	3
0	1	2
1	0	1
1	1	0

C	D	f_8
0	0	1
0	1	4
1	0	0
1	1	0

E	F	f_9
0	0	1
0	1	0
1	0	0
1	1	2

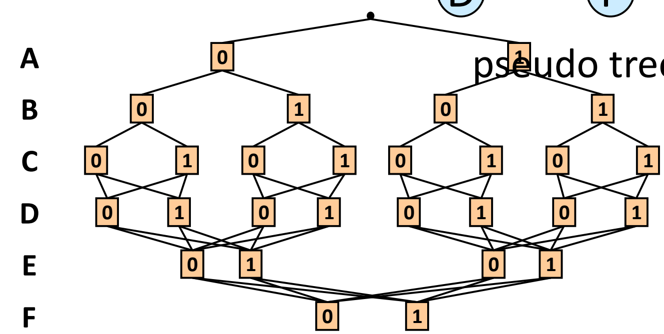






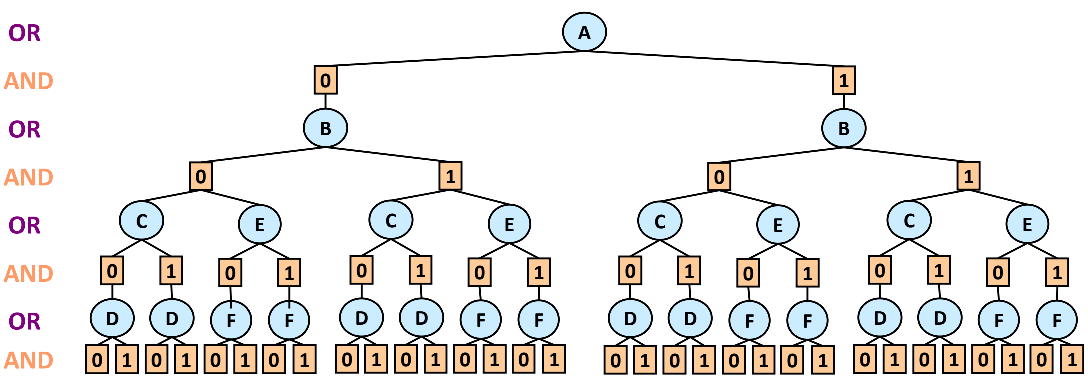
Full OR search tree

126 nodes



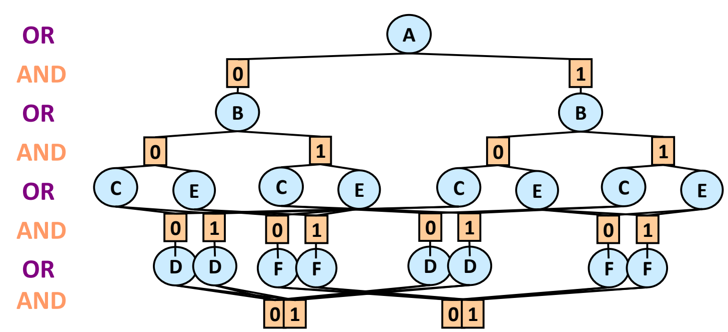
Context minimal OR search graph

28 nodes



Full AND/OR search tree

54 AND nodes



Context minimal AND/OR search graph

18 AND nodes

Any query is best computed
Over the c-minimal AO search space

Potential search spaces

$$F^* = \min_x \sum_\alpha f_\alpha(x_\alpha)$$

A	B	f_1
0	0	2
0	1	0
1	0	1
1	1	4

A	C	f_2
0	0	3
0	1	0
1	0	0
1	1	1

A	E	f_3
0	0	0
0	1	3
1	0	2
1	1	0

A	F	f_4
0	0	2
0	1	0
1	0	0
1	1	2

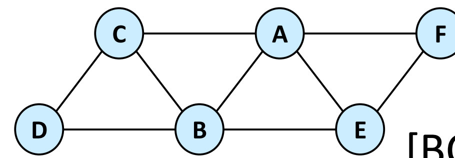
B	C	f_5
0	0	0
0	1	1
1	0	2
1	1	4

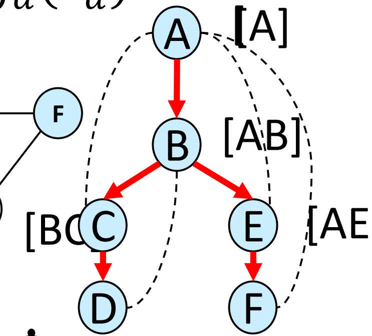
B	D	f_6
0	0	4
0	1	2
1	0	1
1	1	0

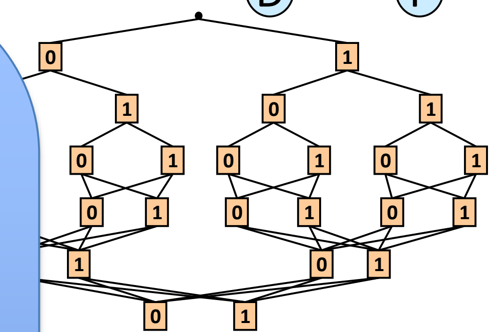
B	E	f_7
0	0	3
0	1	2
1	0	1
1	1	0

C	D	f_8
0	0	1
0	1	4
1	0	0
1	1	0

E	F	f_9
0	0	1
0	1	0
1	0	0
1	1	2

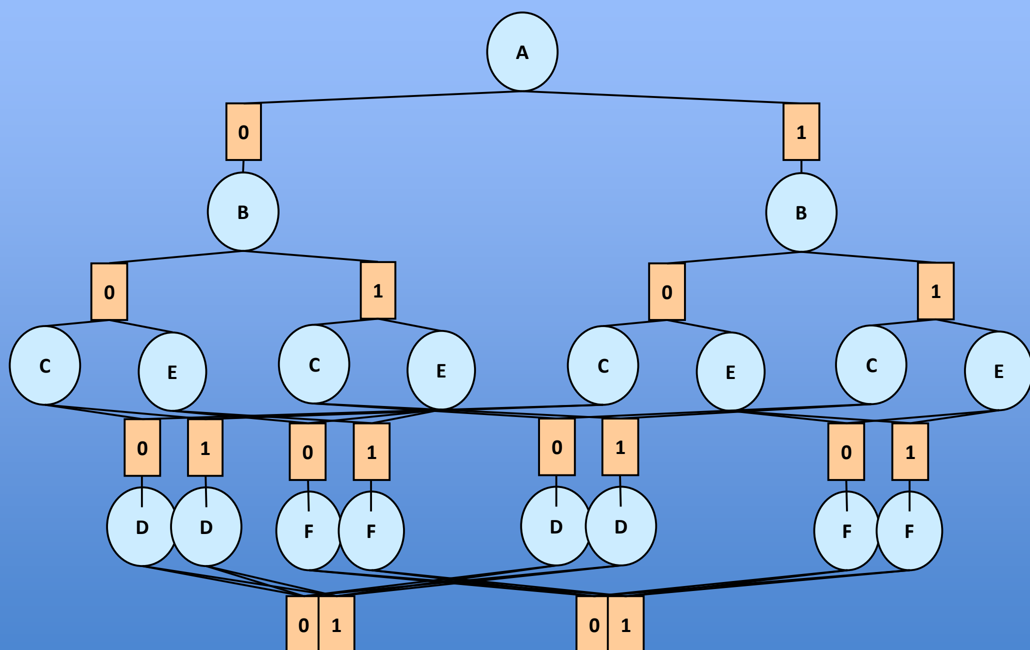






ext minimal OR search graph

28 nodes



Context minimal AND/OR search graph

18 AND nodes

...s best computed

over the c-minimal AO search space

Potential search spaces

$$F^* = \min_x \sum_\alpha f_\alpha(x_\alpha)$$

A	B	f_1
0	0	2
0	1	0
1	0	1
1	1	4

A	C	f_2
0	0	3
0	1	0
1	0	0
1	1	1

A	E	f_3
0	0	0
0	1	3
1	0	2
1	1	0

A	F	f_4
0	0	2
0	1	0
1	0	0
1	1	2

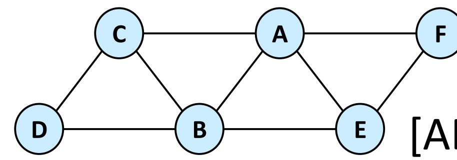
B	C	f_5
0	0	0
0	1	1
1	0	2
1	1	4

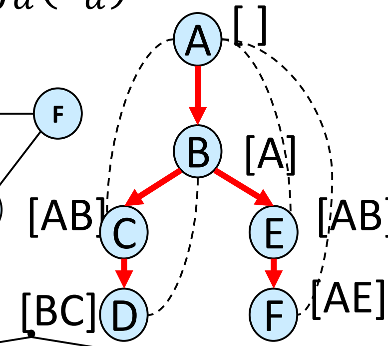
B	D	f_6
0	0	4
0	1	2
1	0	1
1	1	0

B	E	f_7
0	0	3
0	1	2
1	0	1
1	1	0

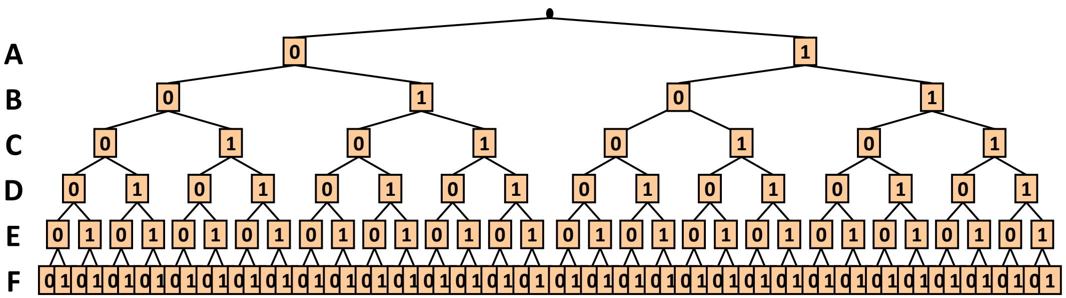
C	D	f_8
0	0	1
0	1	4
1	0	0
1	1	0

E	F	f_9
0	0	1
0	1	0
1	0	0
1	1	2



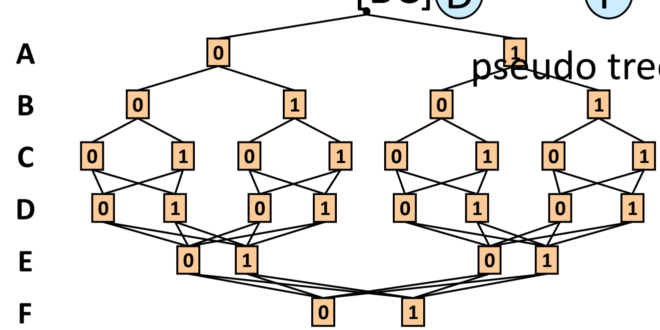


pseudo tree



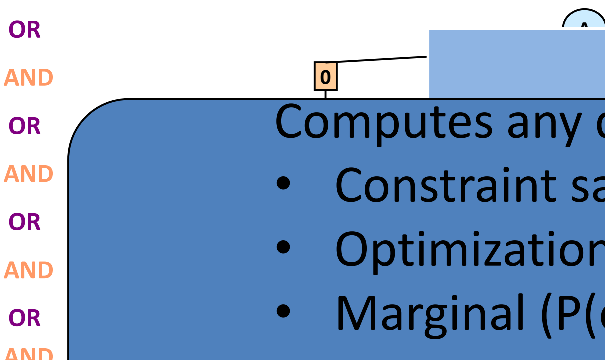
Full OR search tree

126 nodes

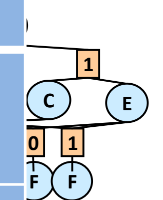


Context minimal OR search graph

28 nodes



	OR tree	AND/OR	OR graph	AND/OR graph
			$O(n\,k^{pw^*})$	$O(n\,k^{w^*})$
			$O(n\,k^{pw^*})$	$O(n\,k^{w^*})$



Computes any query:

- Constraint satisfaction
- Optimization (MAP)
- Marginal (P(e))
- Marginal map

54 AND nodes

18 AND nodes

Any query is best computed
Over the c-minimal AO search space

Cost of a Solution Tree

$P(E|A,B)$

A	B	E=0	E=1
0	0	.4	.6
0	1	.5	.5
1	0	.7	.3
1	1	.2	.8

Evidence: E=0

$P(B|A)$

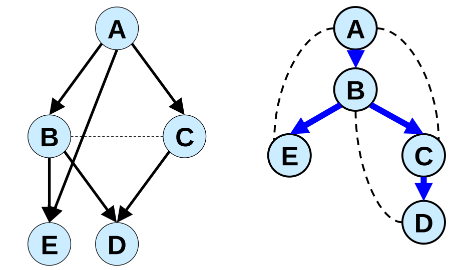
A	B=0	B=1
0	.4	.6
1	.1	.9

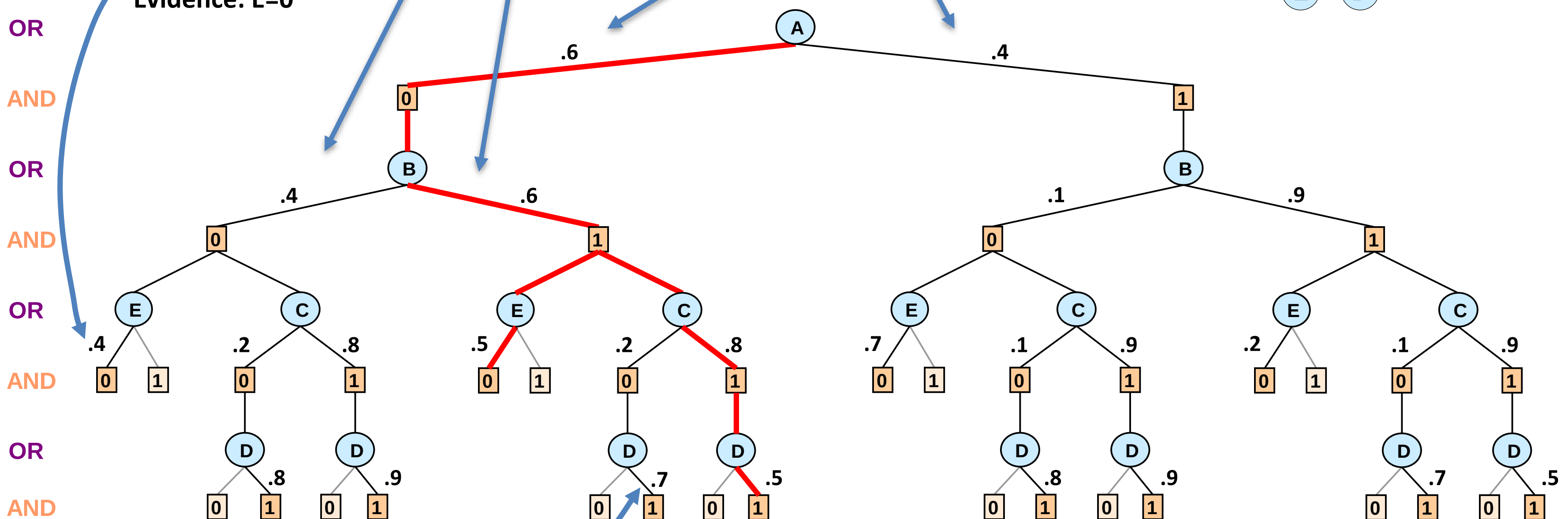
$P(C|A)$

A	C=0	C=1
0	.2	.8
1	.7	.3

$P(A)$

A	P(A)
0	.6
1	.4





$P(D|B,C)$

B	C	D=0	D=1
0	0	.2	.8
0	1	.1	.9
1	0	.3	.7
1	1	.5	.5

Evidence: D=1

Cost of the solution tree: the product of weights on its arcs

Cost of (A=0,B=1,C=1,D=1,E=0) = $0.6 \cdot 0.6 \cdot 0.5 \cdot 0.8 \cdot 0.5 = 0.0720$

Value of a Node (e.g., Probability of Evidence)

$P(E \mid A,B)$

A	B	E=0	E=1
0	0	.4	.6
0	1	.5	.5
1	0	.7	.3
1	1	.2	.8

Evidence: E=0

$P(B \mid A)$

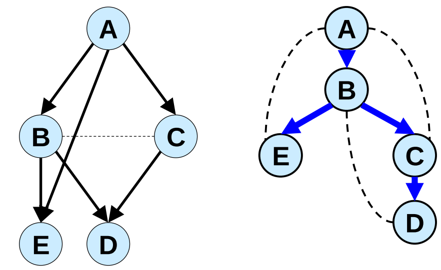
A	B=0	B=1
0	.4	.6
1	.1	.9

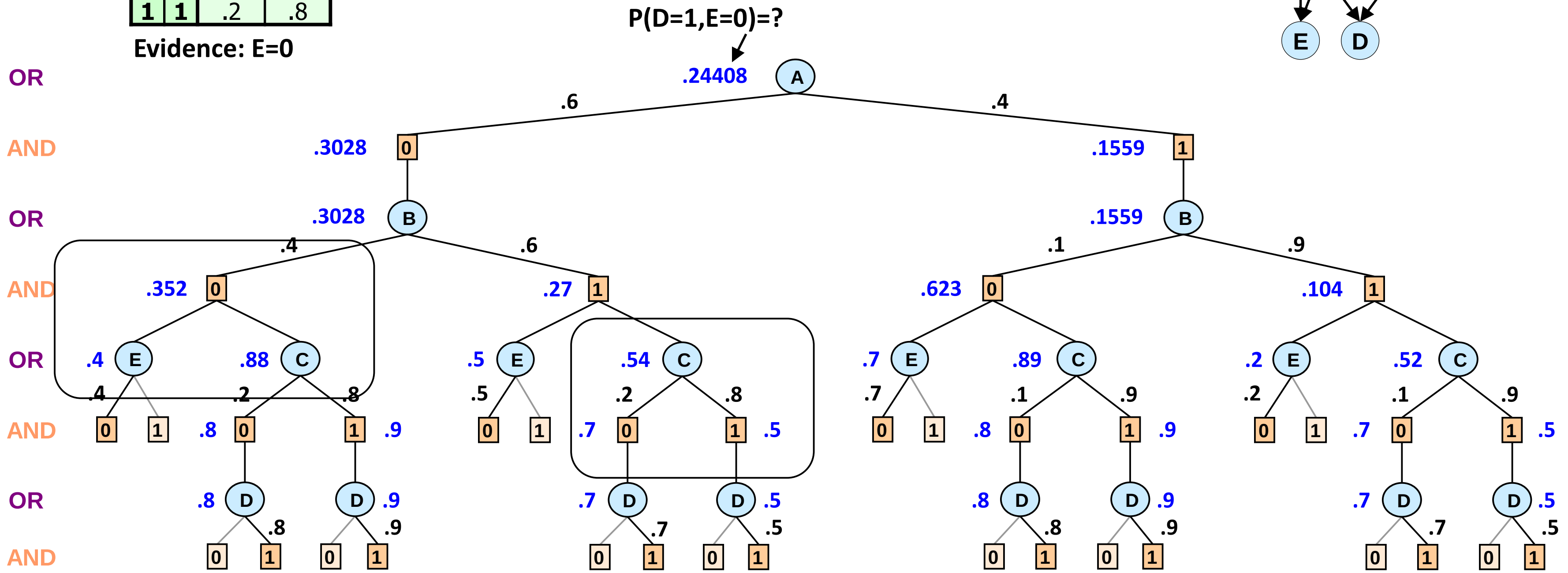
$P(C \mid A)$

A	C=0	C=1
0	.2	.8
1	.7	.3

$P(A)$

A	P(A)
0	.6
1	.4





$P(D \mid B,C)$

B	C	D=0	D=1
0	0	.2	.8
0	1	.1	.9
1	0	.3	.7
1	1	.5	.5

Evidence: D=1

Value of node = updated belief for sub-problem below

AND node: product $\prod_{n' \in children(n)} v(n')$

OR node: Marginalization by summation $\sum_{n' \in children(n)} w(n,n')\, v(n')$

Answering Queries: Sum-Product (Belief Updating)

$P(E|A,B)$

A	B	E=0	E=1
0	0	.4	.6
0	1	.5	.5
1	0	.7	.3
1	1	.2	.8

Evidence: E=0

$P(B|A)$

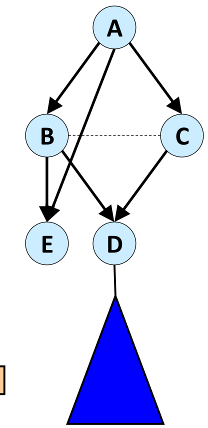
A	B=0	B=1
0	.4	.6
1	.1	.9

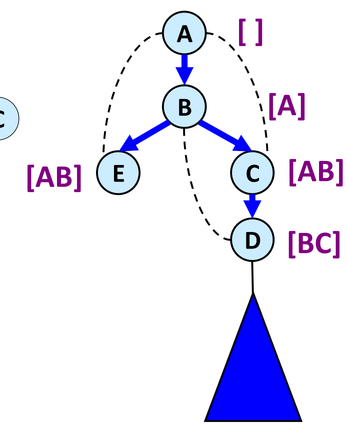
$P(C|A)$

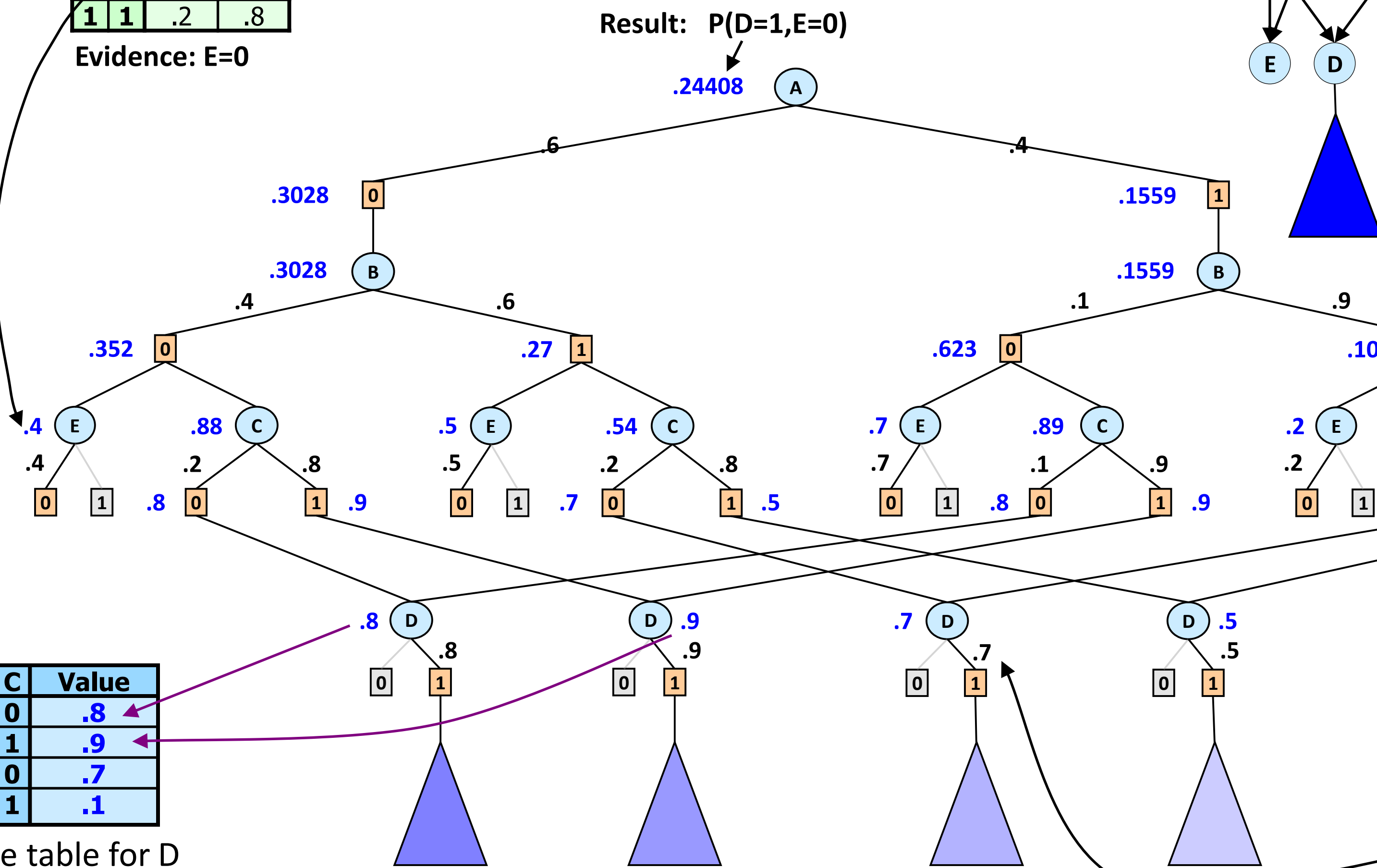
A	C=0	C=1
0	.2	.8
1	.7	.3

$P(A)$

A	P(A)
0	.6
1	.4







B	C	Value
0	0	.8
0	1	.9
1	0	.7
1	1	.1

Cache table for D

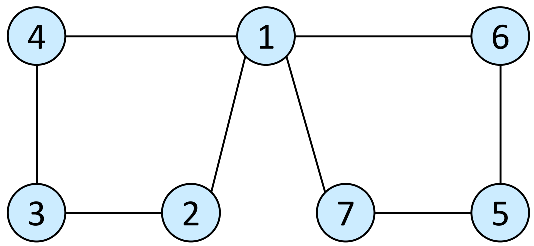
$P(D|B,C)$

B	C	D=0	D=1
0	0	.2	.8
0	1	.1	.9
1	0	.3	.7
1	1	.5	.5

Evidence: D=1

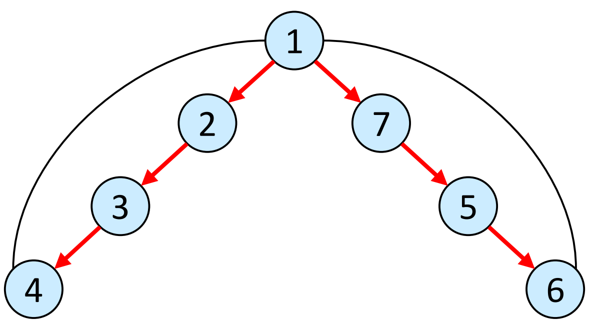
Pseudo-Trees

(Freuder 85, Bayardo 95, Bodlaender and Gilbert, 91)

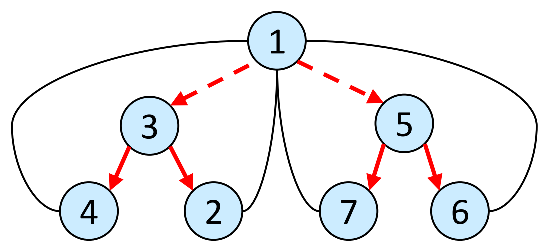


(a) Graph

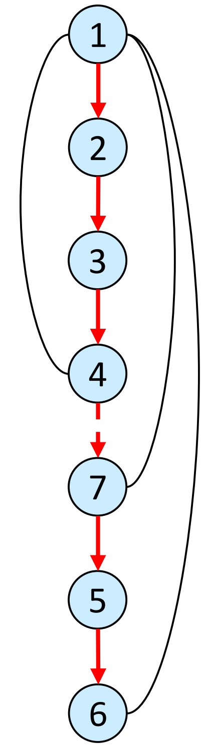
$h \le w^* \log n$



(b) DFS tree
height=3



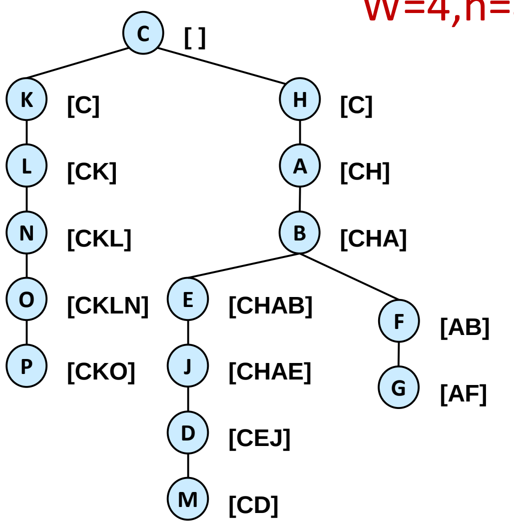
(c) pseudo- tree
height=2



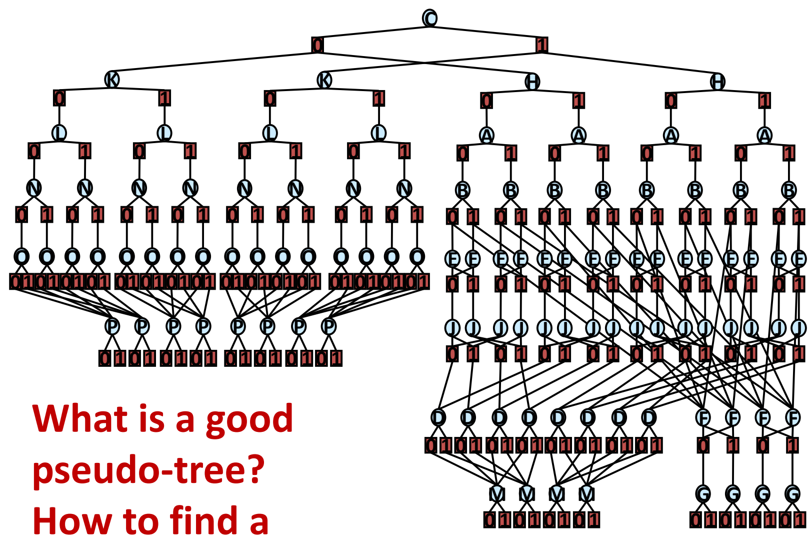
(d) Chain
height=6

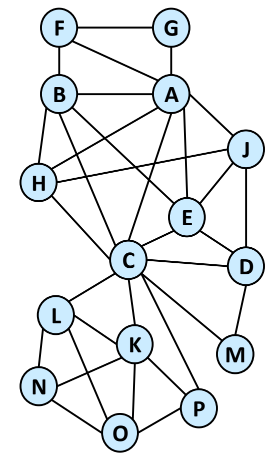
The Impact of the Pseudo-Tree

W=4,h=8



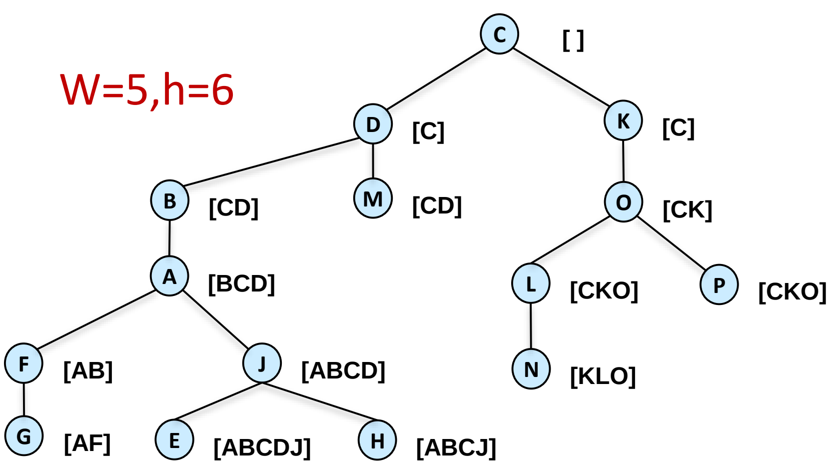
(C K H A B E J L N O D P M F G)

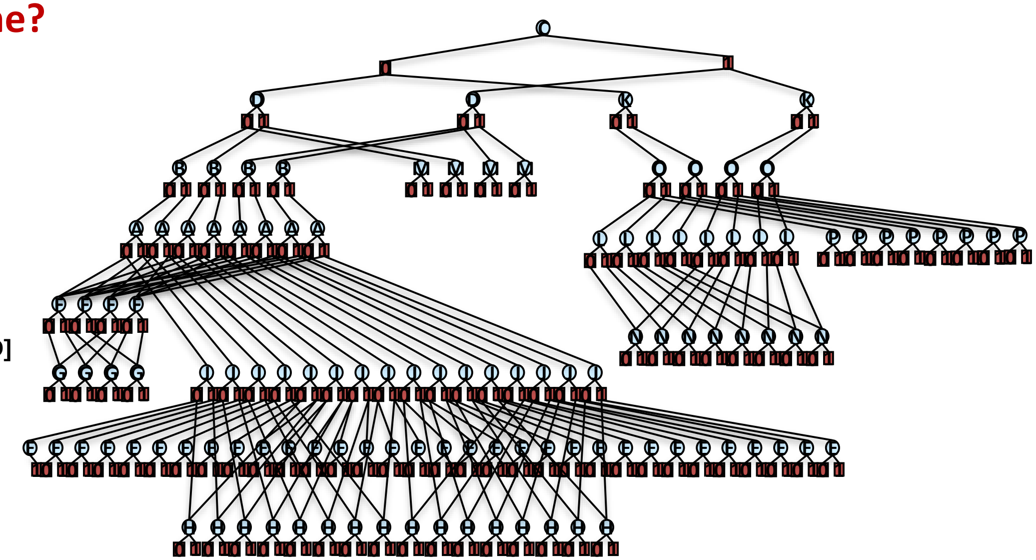




What is a good pseudo-tree? How to find a good one?

W=5,h=6

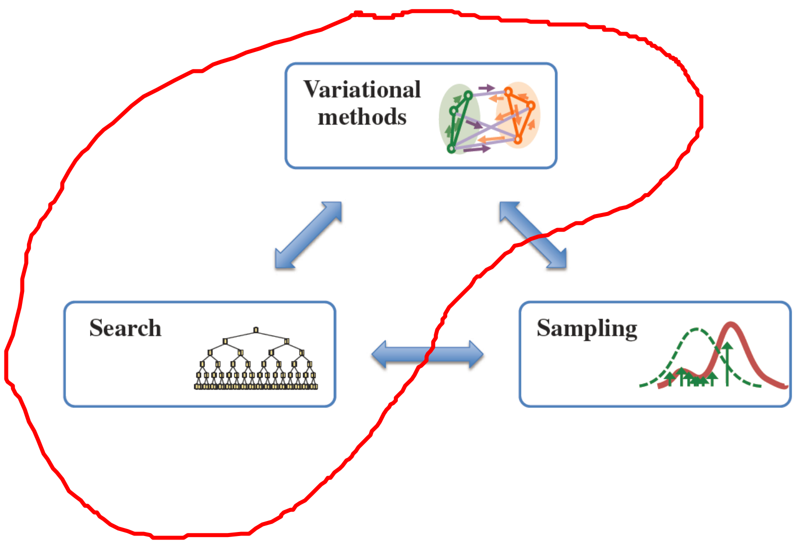


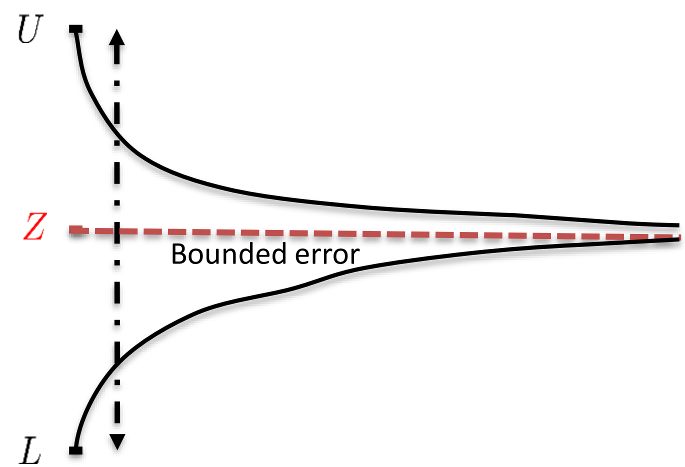


(C D K B A O M L N P J H E F G)

Outline

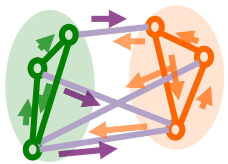
- Graphical models: definition, examples, methodology
- Inference and variational bounds
- AND/OR search spaces
- Variational bounds as heuristics for search
- Combining methods: Heuristic Search for Marginal Map
- Conclusion





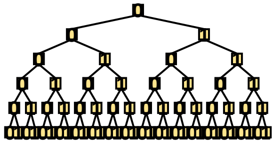
Search Aided by Variational Heuristics

Variational methods



provide pre-compiled heuristics

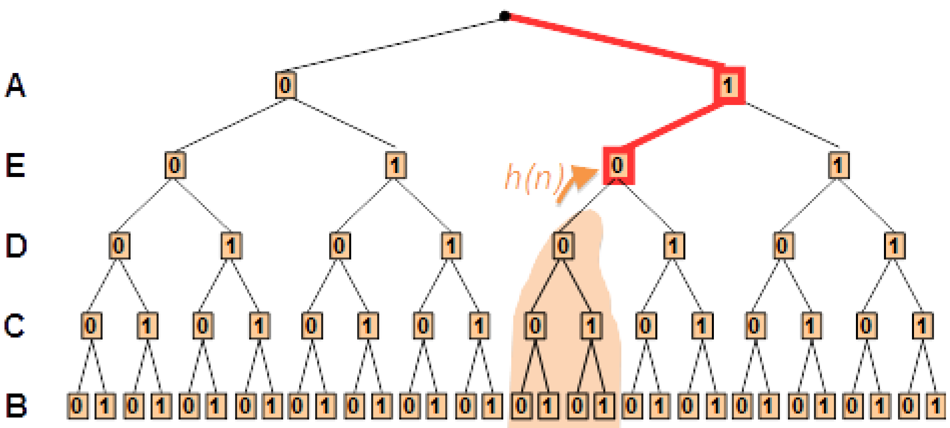
Search

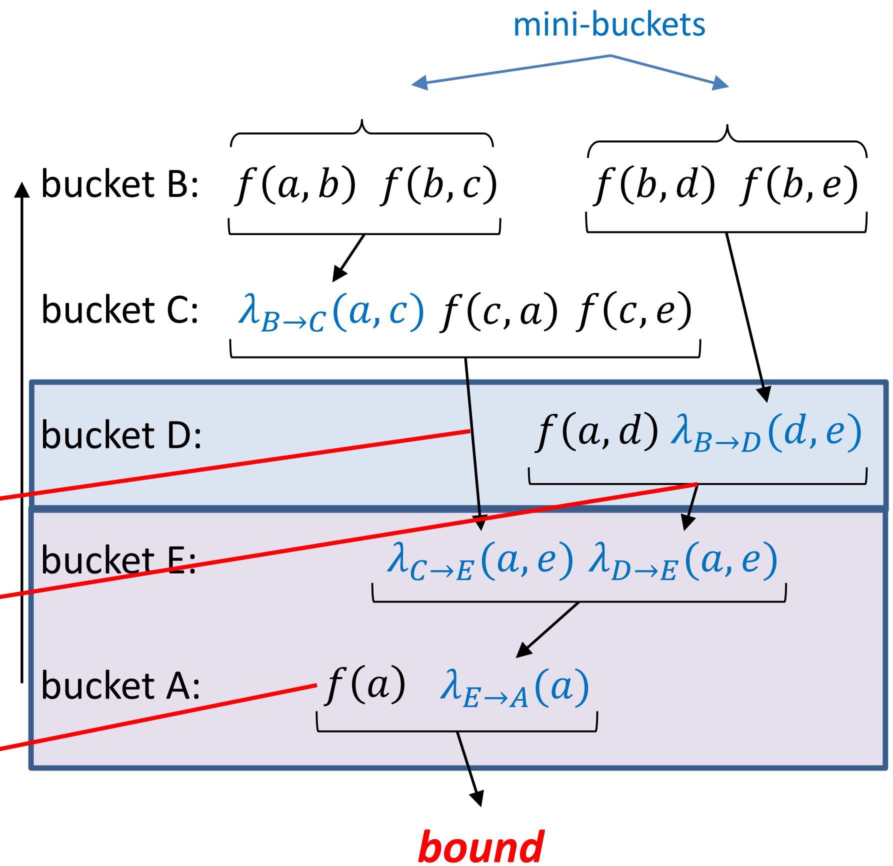


For MAP, marginal map and partition function

Static Mini-Bucket Heuristics

Given a partial assignment, $[\hat{a} = 1, \hat{e} = 0]$
(weighted) mini-bucket gives an admissible heuristic:





"cost to go":

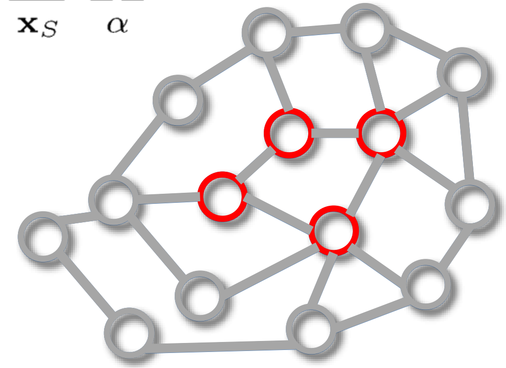
$$\tilde{h}(\hat{a},\hat{e},D) = \lambda_{C\to E}(\hat{a},\hat{e}) + f(\hat{a},D) + \lambda_{B\to D}(D,\hat{e})$$

"cost so far":

$$g(\hat{a},\hat{e},D) = f(A=\hat{a})$$

Why Marginal MAP?

$$f(\mathbf{x}_M^*) = \max_{\mathbf{x}_M} \sum_{\mathbf{x}_S} \prod_{\alpha} f_\alpha(\mathbf{x}_\alpha)$$

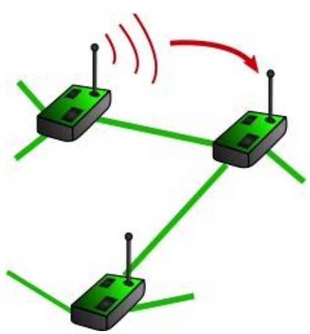


- Often, Marginal MAP is the "right" task:
  - We have a model describing a large system
  - We care about predicting the state of some part

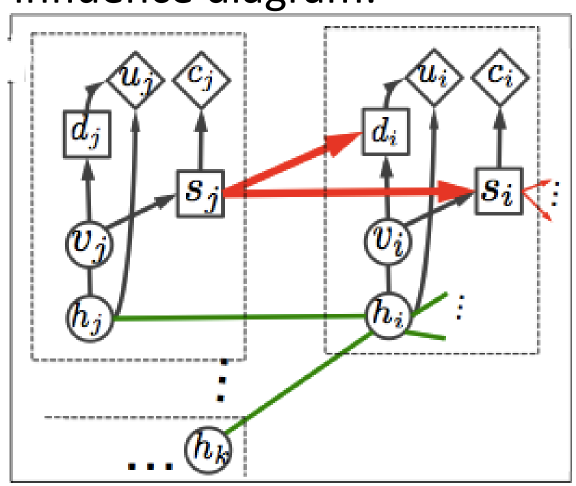
- Example: decision making
  - Sum over random variables (
  - Max over decision variables (specify action policies)

> – Complexity: NP^{PP} complete
> – Not necessarily easy on trees

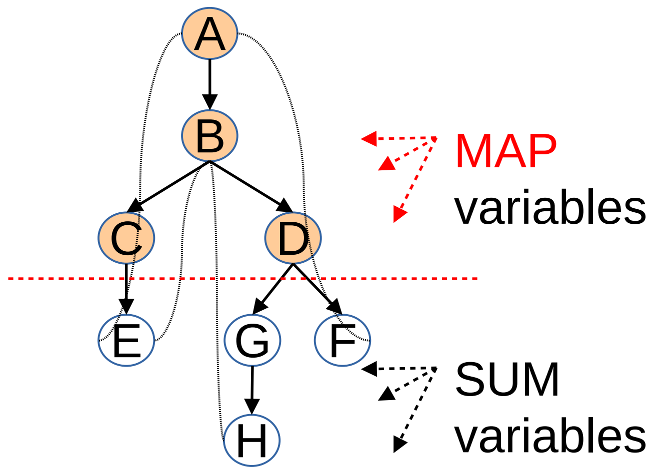
Sensor network



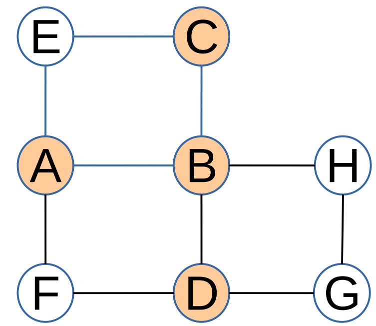
Influence diagram:



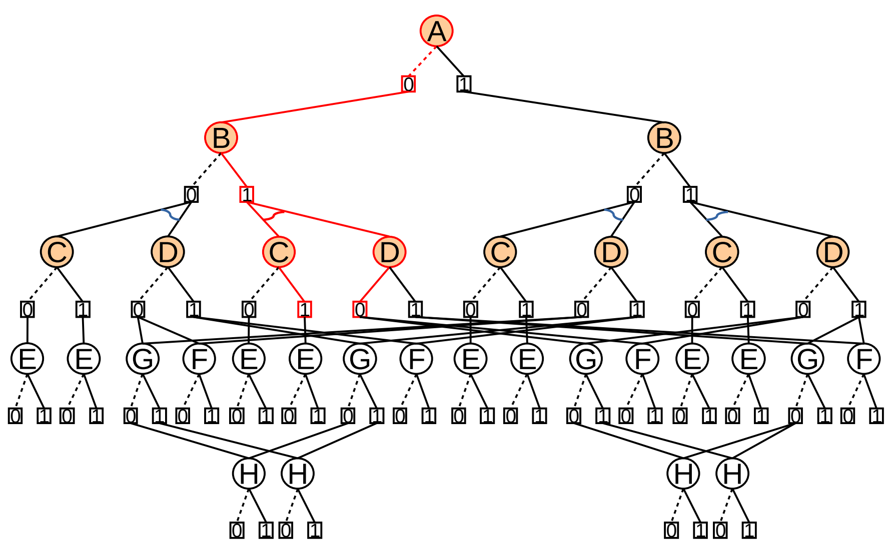
AND/OR Search for Marginal MAP

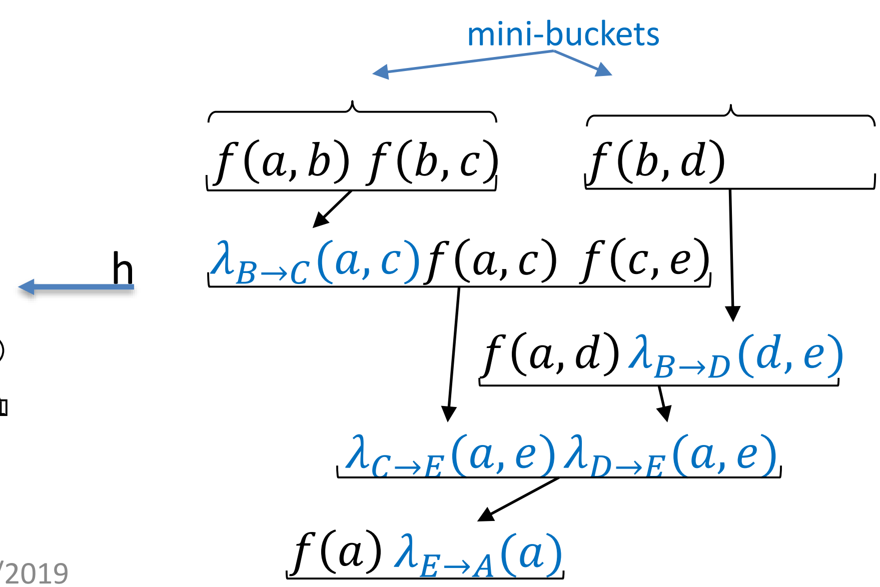


constrained
pseudo tree



primal graph





[Marinescu, Dechter and Ihler, 2014]

AO search for MAP winning UAI Probabilistic Inference Competitions

- **2006**  (aolib)
- **2008**  (aolib)
- **2011**  (daoopt)
- **2014**  (daoopt) |  (daoopt)  (merlin)

MPE/MAP | MMAP

Anytime AND/OR solvers for MMAP

- **Weighted Heuristic:** [Lee et. al. AAAI-2016]
  - Weighted Restarting AOBF (WAOBF)
  - Weighted Restarting RBFAOO (WRBFAOO)
  - Weighted Repairing AOBF (WRAOBF)

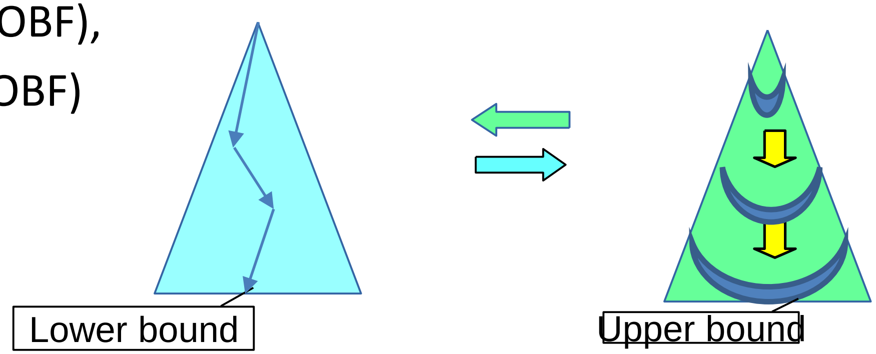
Weighted A* search [Pohl 1970]

- non-admissible heuristic
- Evaluation function:

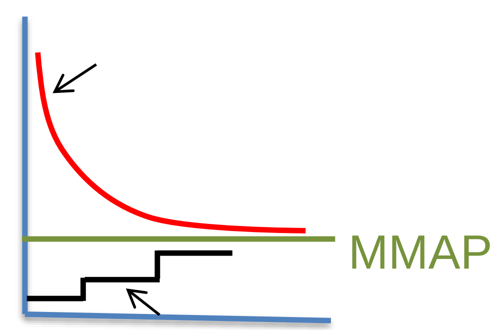
  $f(n)=g(n)+w \cdot h(n)$

- **Guaranteed w-optimal solution, cost $C \leq w \cdot C^*$**

- **Interleaving Best and depth-first search**: (Marinescu et. al AAAI-2017)
  - Look-ahead (LAOBF),
  - alternating (AAOBF)



Goal: anytime bounds
And anytime solution



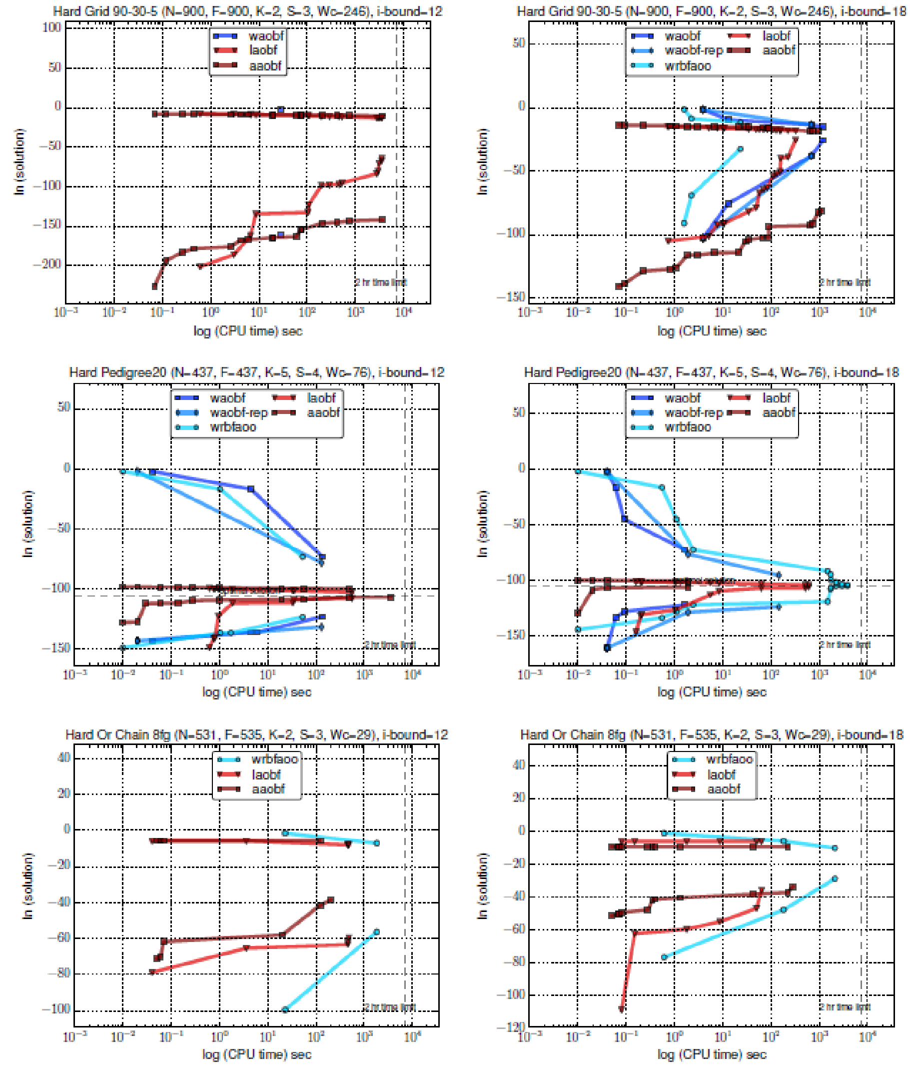
Anytime Bounding of Marginal MAP

(UAI’14, IJCAI’15, AAAI’16, AAAI’17, (Marinescu, Lee, Ihler, Dechter)

- Search: LAOBF, AAOBF, BRAOBB, WAOBF,WAOBF-rep
- heuristic: WMB-MM (20)
- memory: 24 GB

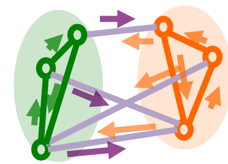
- Anytime lower and upper bounds from hard problem instances with i-bound 12 (left) and 18 (right).

- The horizontal axis is the CPU time in log scale and the vertical axis is the value of marginal MAP in log scale.



Combining Approaches

Variational methods



weighted mini-bucket (WMB)

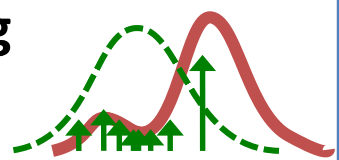
[Liu and Ihler, ICML 2011]

provide proposal

WMB-IS

[Liu et al., NIPS 2015]

Sampling



dynamic importance sampling (DIS)

[Lou et al., NIPS 2017]

Choosing a proposal

[Liu, Fisher, Ihler 2015]

- Can use WMB upper bound to define a proposal $q_{\text{wmb}}(x)$

$$\tilde{\mathbf{b}} \sim w_1\, q_1(b|\tilde{a},\tilde{c}) \;+\; w_2\, q_2(b|\tilde{d},\tilde{e})$$

Weighted mixture:

use minibucket 1 with probability w_1

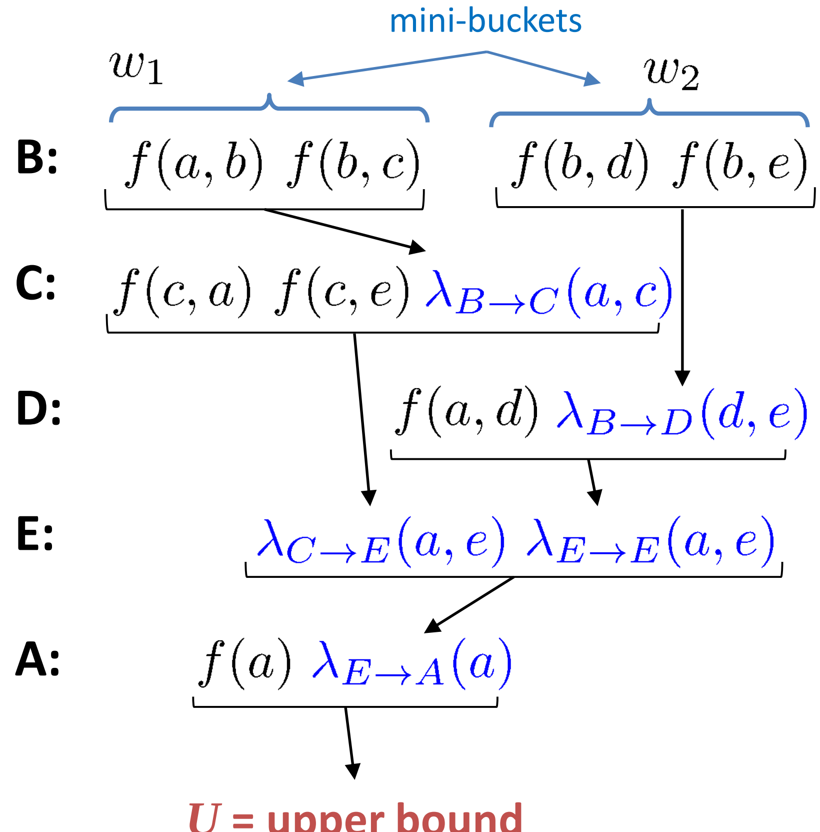
or, minibucket 2 with probability $w_2 = 1 - w_1$

where

$$q_1(b|a,c) = \left[\frac{f(a,b)\cdot f(b,c)}{\lambda_{B\to C}(a,c)}\right]^{\frac{1}{w_1}}$$

⋮

$$\tilde{\mathbf{a}} \sim q(A) = f(a)\cdot\lambda_{E\to A}(a)/U$$

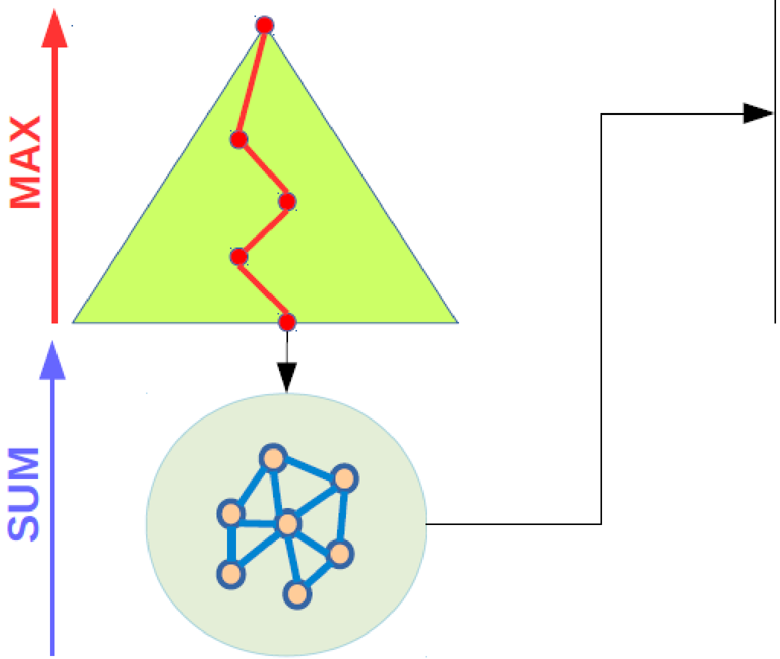


Key insight: provides bounded importance weights!

$$0 \le f(x)/q_{\text{wmb}}(x) \le U \quad \forall x$$

Probabilistic Lower Bounds

$$\max_{\mathbf{X}_M} \sum_{\mathbf{X}_S} P(\mathbf{X})$$



$$Z = \sum_{\mathbf{X}_S} P(X)|_{\bar{x}_M}$$

Solving the conditioned SUM subproblem is hard!

$$\#P - complete$$

[Liu *et al*. 2015]

Compute a (probabilistic) lower bound on the conditioned sum subproblem

$$Pr(\hat{Z} - \Delta(n, \delta) \leq Z) \geq (1 - \delta)$$

WMB based importance sampling scheme:

n - number of samples

δ - confidence value

Z_{wmb} - result of WMB

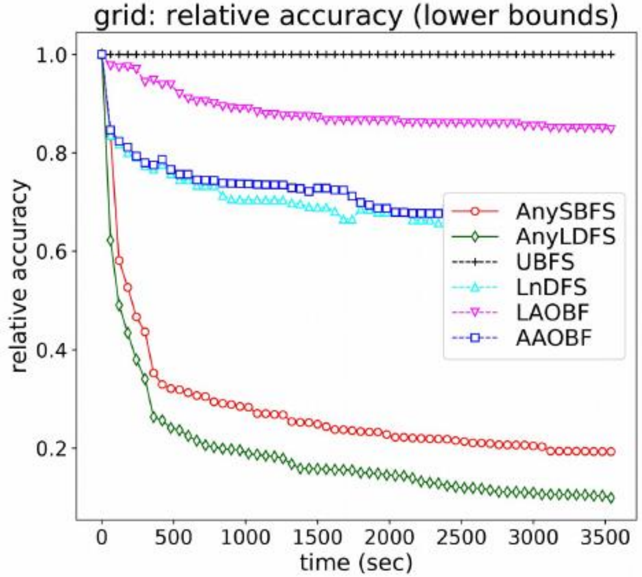
$\hat{Z}$ - Importance Sampling estimate

$$\Delta(n, \delta) = \sqrt{\frac{2\hat{var}(w(x))\log(2/\delta)}{n}} + \frac{7Z_{wmb}\log(2/\delta)}{3(n-1)}$$

Combining Methods: +Sampling

[Lou, Dechter, Ihler, AAAI-2018: "Anytime Anyspace AND/OR Best-first Search for Bounding Marginal MAP"]
[Lou, Dechter, Ihler, UAI-2018: "Finite Sample Bounds for Marginal MAP", UAI 2018]
[Marinescu, Ihler, Dechter: IJCAI-2018 "Stochastic Anytime Search for Bounding Marginal MAP"]

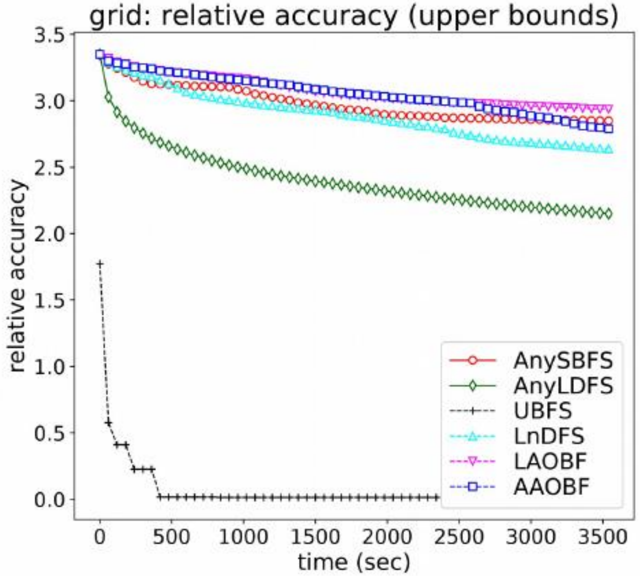
$$acc_{lb} = \frac{|l_t - l^*|}{l^*}$$



l_t – lower bound at time t
l^* – tightest lower bound found

Average over 150 instances

$$acc_{ub} = \frac{|u_t - u^*|}{u^*}$$

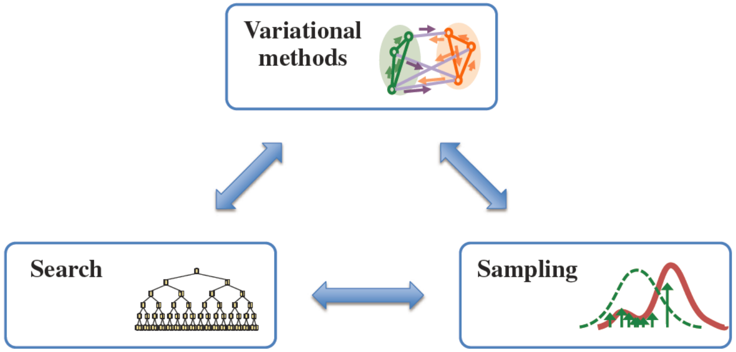


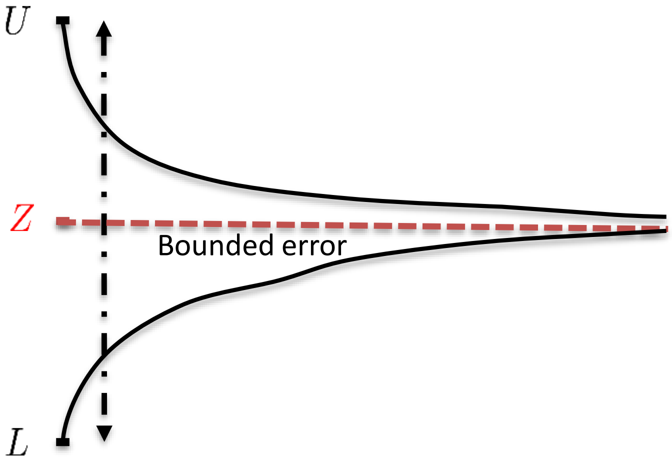
u_t – upper bound at time t
u^* – tightest upper bound found

Average over 150 instances

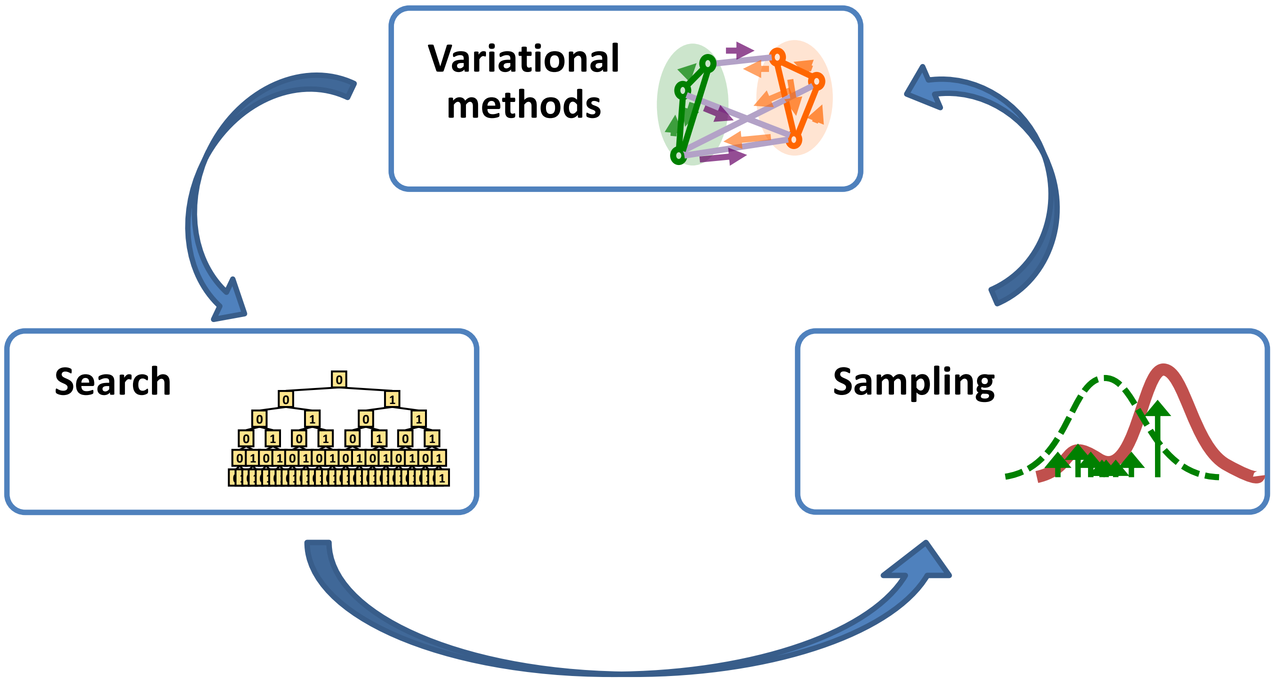
Outline

- Graphical models: definition, examples, methodology
- Inference and variational bounds
- AND/OR search spaces
- Variational bounds as search heuristics
- Combining methods: Heuristic Search for Marginal Map
- Conclusion





Continuing work

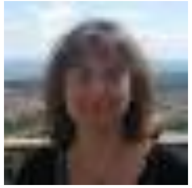


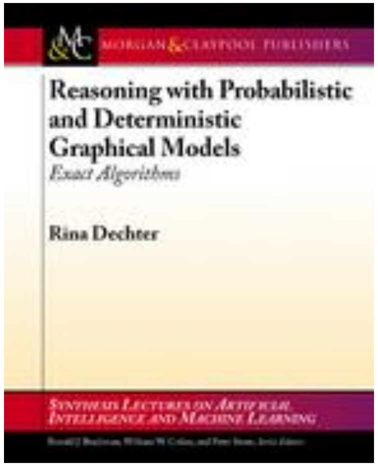
Thank You !

For publication see:
http://www.ics.uci.edu/~dechter/publications.html





Alex Ihler
Kalev Kask
Irina Rish
Bozhena Bidyuk
Robert Mateescu
Radu Marinescu
Vibhav Gogate
Emma Rollon
Lars Otten
Natalia Flerova
Andrew Gelfand
William Lam
Junkyu Lee
Qi Liu