Bayesian Networks and Belief Propagation: From Rumelhart to Pearl to Today

Rina Dechter
Donald Bren School of Computer Science
University of California, Irvine, USA

In The ELSC-ICNC Retreat 2012, Ein Gedi

The David E. Rumelhart Prize

For Contributions to the Theoretical Foundations of Human Cognition

Rumelhart Prize for Pearl, 2011

http://thesciencenetwork.org/programs/cogsci-2011/rumelhart-lecture-judea-pearl

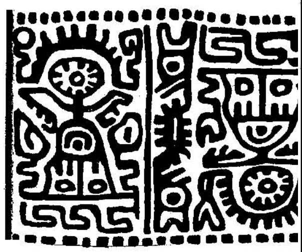
Dr. Judea Pearl has been a key researcher in the application of probabilistic methods to the understanding of intelligent systems, whether natural or artificial. He has pioneered the development of graphical models, and especially a class of graphical models known as Bayesian networks, which can be used to represent and to draw inferences from probabilistic knowledge in a highly transparent and computationally natural fashion. Graphical models have had a transformative impact across many disciplines, from statistics and machine learning to artificial intelligence; and they are the foundation of the recent emergence of Bayesian cognitive science. Dr. Pearl's work can be seen as providing a rigorous foundation for a theory of epistemology which is not merely philosophically defensible, but which can be mathematically specified and computationally implemented. It also provides one of the most influential sources of hypotheses about the function of the human mind and brain in current cognitive science.

Rumelhart 1976:
Towards an interactive model of Reading

TOWARD AN INTERACTIVE MODEL OF READING

David E. Rumelhart

Pearl: so we have a combination of a top down and a bottom up modes of reasoning which somehow coordinate their actions resulting in a friendly handshaking."



CENT

Jack and Jill went up the hill.

The pole vault was the last event.

UNIVERSITY OF CALIFOR
LA JOLLA, CALIFORNIA 9

Figure 3 The dependence of letter perception of context. (After Nash-Weber, 1975.)

Rumelhart's Proposed Solution

Rumelhart (1976) Figure 10

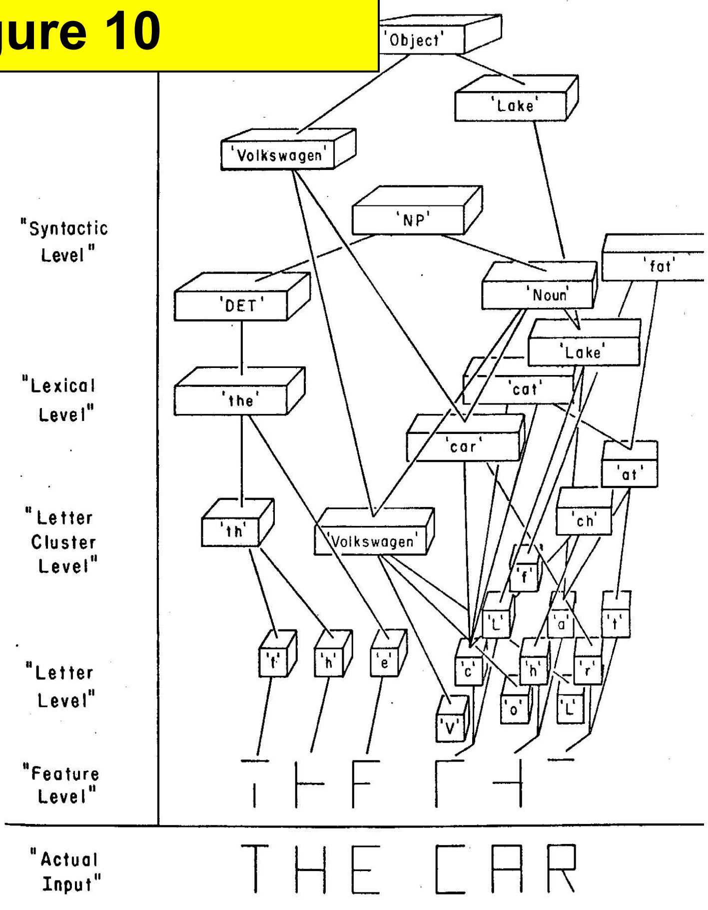


Figure 10 The message center well into the processing sequence.

and sisters of h_i. Equation (2) gives the value of the contextual strength of h_i:

$$(2) \qquad \beta_i = \begin{cases} Pr(h_i) & P_i = L_i = \phi \\ \dfrac{\Sigma s_k \cdot Pr(h_i \mid h_k)}{\nu_i} & \text{otherwise,} \end{cases}$$

where the sum is over all $h_k \in P_i$ or L_i. Thus, when h_i has no parents or left sisters, its contextual strength is given by its a priori probability. Otherwise, its contextual strength is given by the sum, over all of its left sisters and parents of the strength of the left sister or parent, h_k, times the conditional probability of the hypothesis given h_k. This sum is then

Pearl 1982: Reverend Bayes on Inference Engines

REVEREND BAYES ON INFERENCE ENGINES: A DISTRIBUTED HIERARCHICAL APPROACH(*)(**)

Judea Pearl
Cognitive Systems Laboratory
School of Engineering and Applied Science
University of California, Los Angeles
90024

ABSTRACT

This paper presents generalizations of Bayes likelihood-ratio updating rule which facilitate an asynchronous propagation of the impacts of new beliefs and/or new evidence in hierarchically organized inference structures with multi-hypotheses variables. The computational scheme proposed specifies a set of belief parameters, communica-

feature of hierarchical inference systems is th
the relation P(D|H) is computable as a cascade
local, more elementary probability relations in
volving <u>intervening variables</u>. Intervening var
ables, (e.g., organisms causing a disease) may
may not be directly observable. Their computat
al role, however, is to provide a conceptual
summarization for loosely coupled subsets of ob
vational data so that the computation of P(H|D

Bayes Net (1985)

BAYESIAN NETWORKS: A MODEL OF SELF-ACTIVATED MEMORY FOR EVIDENTIAL REASONING*

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California
Los Angeles, CA 90024
(judea@UCLA-locus)
(213) 825-3243

Topics: Memory Models
Belief Systems
Inference Mechanisms
Knowledge Representation

To be presented at
the 7th Conference of
the Cognitive Science Society
University of California, Irvine
August 15-17, 1985

Technical Report
CSD-850020
R-44
Revision II
October 1985

A CONSTRAINT - PROPAGATION APPROACH TO PROBABILISTIC REASONING*

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California
Los Angeles, CA 90024
(judea@UCLA-locus)
(213) 825-3243

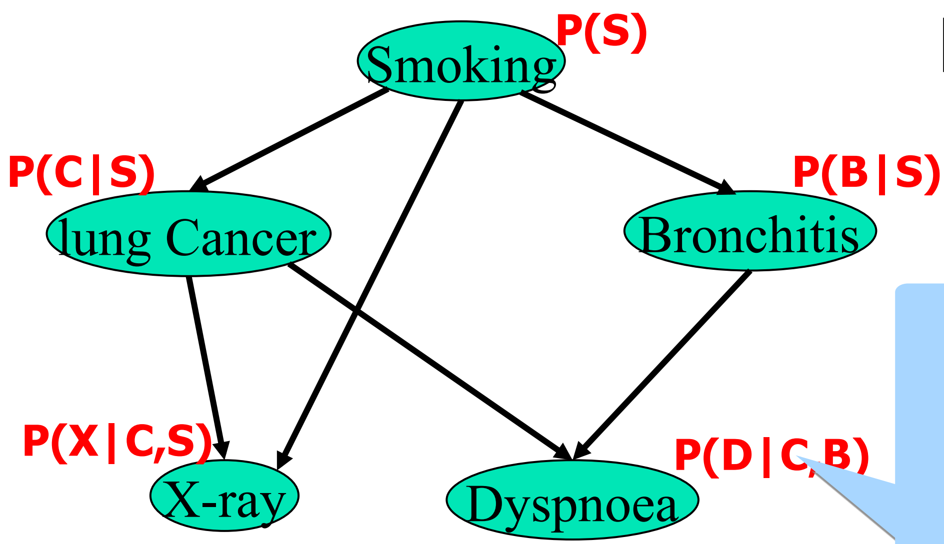
Outline

- Bayesian networks from historical perspective
- Bayesian networks, a short tutorial
- The belief propagation on trees
- From trees to graphs
- From Bayesian networks to graphical models
- Some observations on loopy belief propagation

Outline

- Bayesian networks from historical perspective
- Bayesian networks, a short tutorial
- The belief propagation on trees
- From trees to graphs
- From Bayesian networks to graphical models
- Some observations on loopy belief propagation

Bayesian Networks (Pearl 1985)
A Medical Diagnosis Example (Lauritsen and Spigelhalter, 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

$P(S, C, B, X, D) = P(S)\,P(C|S)\,P(B|S)\,P(X|C,S)\,P(D|C,B)$

The product of all these assessments constitue a joint-probability model which supporst the assessed quantities

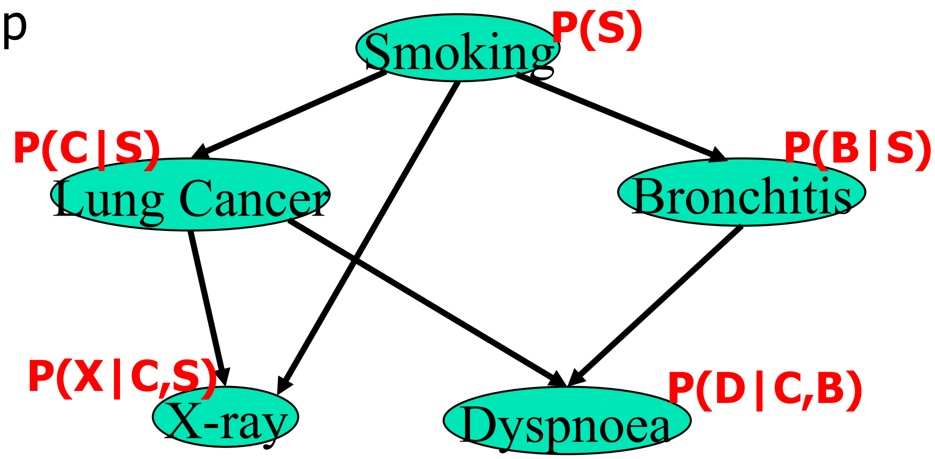
Belief Updating:

P (lung cancer=yes | smoking=no, dyspnoea=yes) = ?

MPE = find argmax P(S)· P(C|S)· P(B|S)· P(X|C,S)· P(D|C,B)

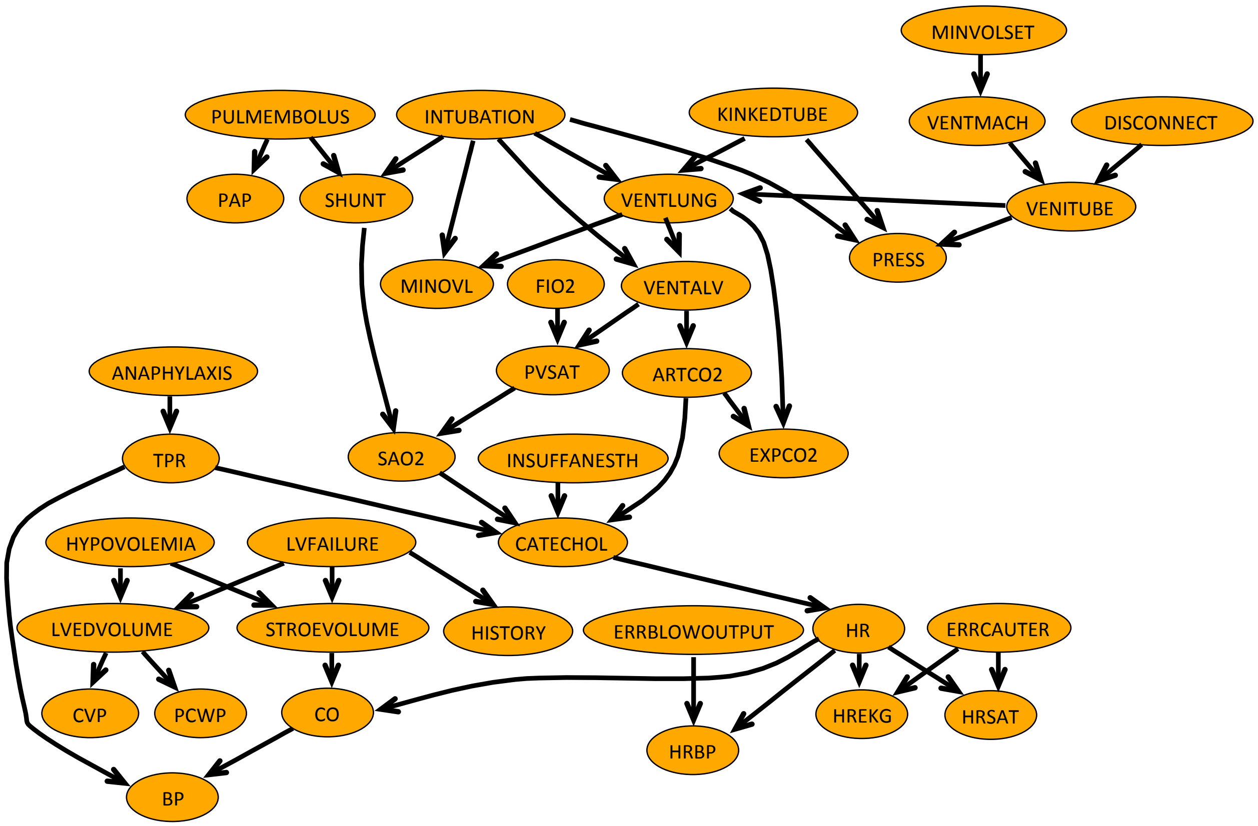
Bayesian Networks Encode Independencies

Causal relationship



An important feature of a Bayesian network is that it provides a clear visual representation for many independence relationships embedded in the underlying probabilistic model. The criterion for detecting these independencies is based on graph separation: namely, if all paths between x_i and x_j are "blocked" by a subset S of variables, then x_i is independent of x_j given the values of the variables in S. Thus, each variable x_i is independent of both its siblings and its grandparents, given the values of the variables in its parent set S_i. For this "blocking" criterion to hold in general, we must provide a special interpretation of separation for nodes that share common children. We say that the pathway along arrows meeting head-to-head at node x_k is "blocked", unless x_k or any of its descendants is in S. In Figure 1, for example, x_2 and x_3 are independent given $S_1 = \{x_1\}$ or $S_2 = \{x_1, x_4\}$, because the two paths between x_2 and x_3 will be blocked by either S_1 or S_2. However, x_2 and x_3 may not be independent given $S_3 = \{x_1, x_6\}$, because x_6, as a descendant of x_5, "unblocks" the head-to-head connection at x_5, thus opening a pathway between x_2 and x_3.

Monitoring Intensive Care Patients



Alarm network

37 variables
509 parameters

<< 2^{37}

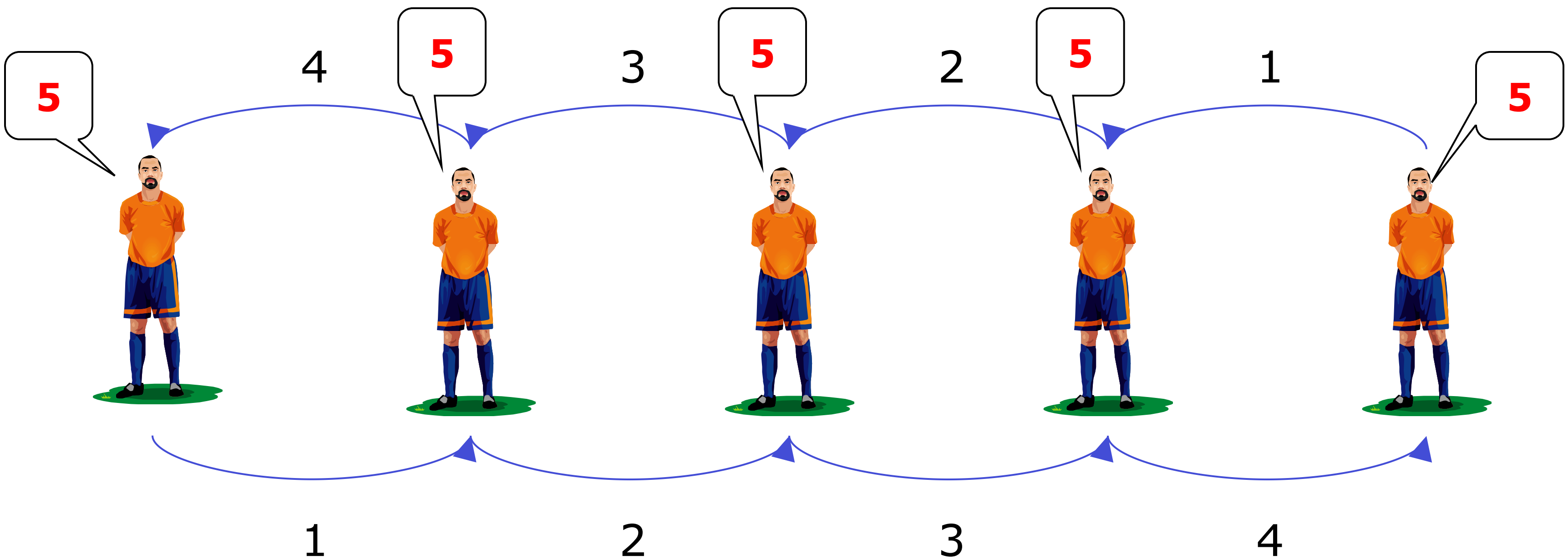
Outline

- Bayesian networks from historical perspective
- Bayesian Networks
- Belief propagation on trees
- From trees to graphs
- From Bayesian network to graphical models
- Some observations on loopy belief propagation

Distributed Belief Propagation

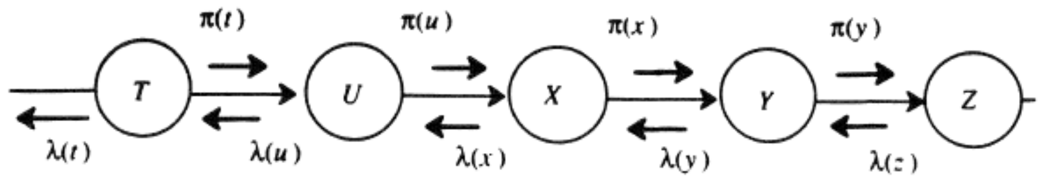
The essence of belief propagation is to make global information be shared locally by every entity

How many people?



Belief Prop

BIDIRECTIONAL



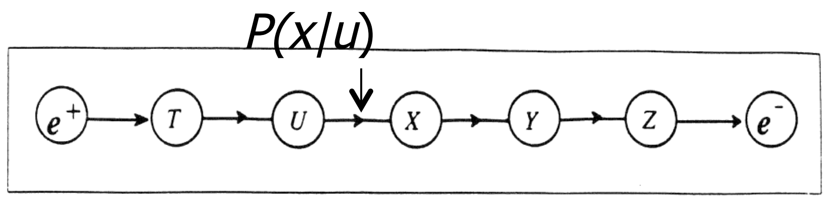


Figure 4.7. A causal chain with evidential data at its head (e^+) and tail (e^-).

$$BEL(x) \triangleq P(x \mid e^-, e^+) = \alpha\, P(e^- \mid x, e^+)\, P(x \mid e^+) = \alpha\, P(e^- \mid x)\, P(x \mid e^+)$$

$$= \alpha\, \pi(x)\, \lambda(x),$$

Where

$$\lambda(x) = P(e^- \mid x) \quad \text{(4.6a)}$$

← Diagnostic support

Causal support →

$$\pi(x) = P(x \mid e^+). \quad \text{(4.6b)}$$

Information about $\pi(x)$ propagates from e^+ down the chain

$$\pi(x) = P(x \mid e^+) = \sum_u P(x \mid u, e^+)\, P(u \mid e^+).$$

Since U separates X from e^+, we obtain

$$\pi(x) = \sum_y P(x \mid u)\, \pi(u) = \pi(u) \cdot M_{x \mid u}. \quad \text{(4.7)}$$

Belief Propagation in Trees

DATA FUSION

$$BEL(x) = P(x \mid e_X^+, e_X^-) = \alpha\, P(e_X^- \mid e_X^+, x)\, P(x \mid e^+)$$
$$= \alpha\, P(e_X^- \mid x)\, P(x \mid e_X^+)\,, \tag{4.13}$$

The probability distribution of every variable in the tree can be computed if the node corresponding to that variable contains the vectors

$$\lambda(x) = P(e_X^- \mid x) \tag{4.15}$$

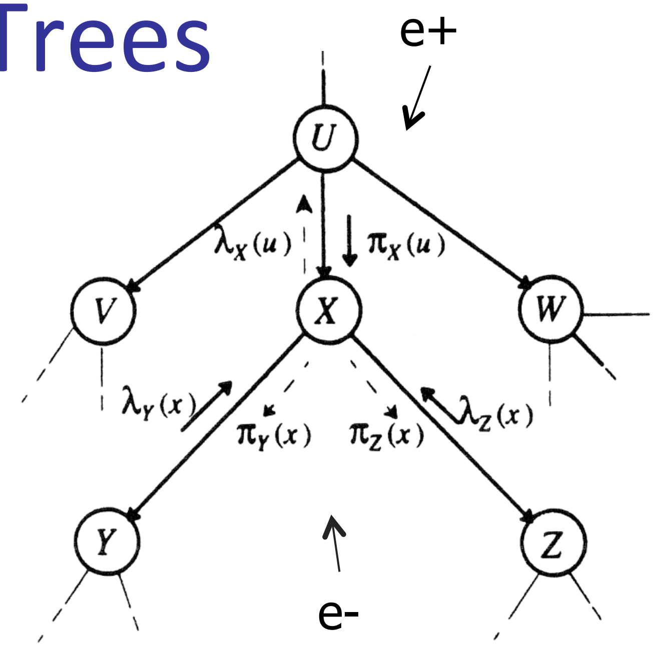
and

$$\pi(x) = P(x \mid e_X^+)\,. \tag{4.16}$$

$\pi(x)$ = causal or *predictive* support attributed to the assertion "$X = x$" by all non-descendants of X.

$\lambda(x)$ = diagnostic or *retrospective* support that "$X = x$" receives from X's descendants.

$$BEL(x) = \alpha \lambda(x)\, \pi(x). \tag{4.17}$$



How information from several descendants fuses at node X.

Write $e_X^- = e_Y^- \cup e_Z^-$, and since X separates its children, we have

$$\lambda(x) = P(e_X^- \mid x)$$
$$= P(e_Y^-, e_Z^- \mid x)$$
$$= P(e_Y^- \mid x) P(e_Z^- \mid x)\,. \tag{4.18}$$

So $\lambda(x)$ can be formed as a product of terms such as $P(e_Y^- \mid x)$, if these terms are delivered to X as messages from its children.

Denoting these messages by subscripted λ's,

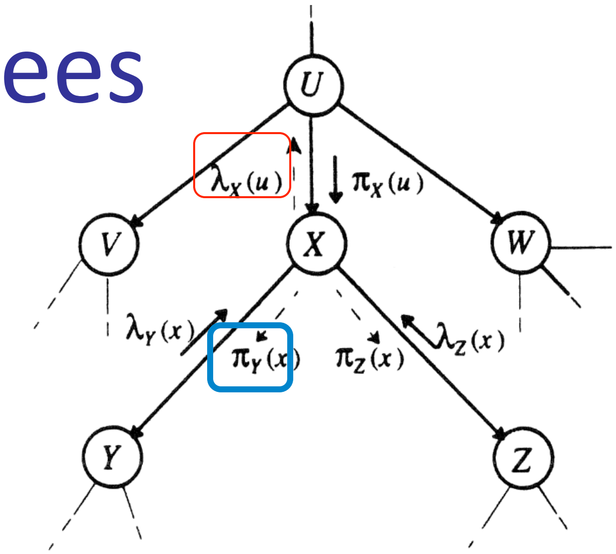
$$\lambda_Y(x) = P(e_Y^- \mid x) \tag{4.19a}$$

$$\lambda_Z(x) = P(e_Z^- \mid x)\,, \tag{4.19b}$$

we have the product rule:

$$\lambda(x) = \lambda_Y(x)\, \lambda_Z(x)\,. \tag{4.20}$$

Belief Propagation in Trees



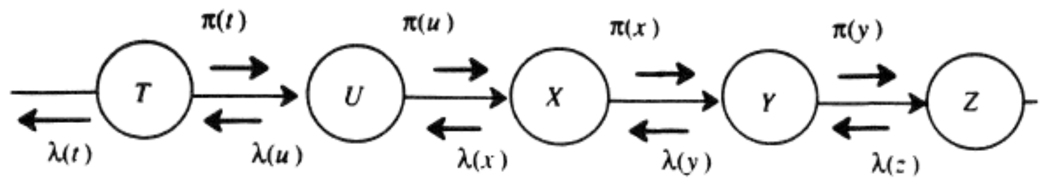
Step 2 - Bottom-up propagation: X computes $\lambda_X(u)$, which is sent to its parent U:

$$\lambda_x(u) = \sum_x \prod_k \lambda_{Y_k}(x) \cdot P(x \mid u)$$

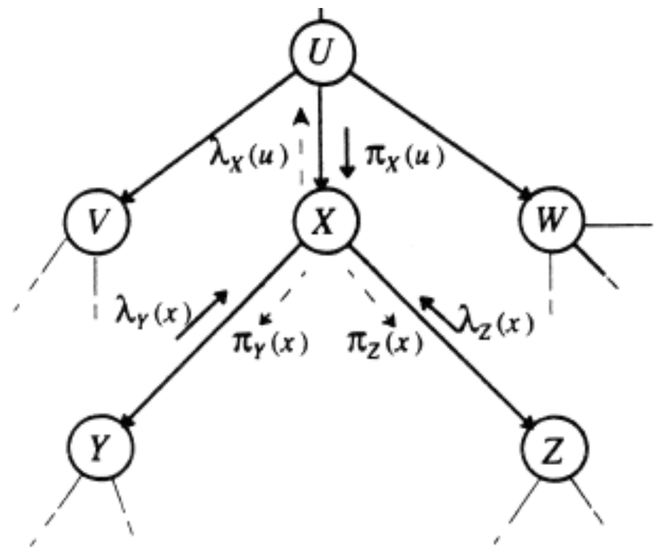
Step 3 - Top-down propagation: X the new $\pi_{Y_j}(x)$ message that X sends to its j-th child Y_j is computed by

$$\pi_{Y_j}(x) = \alpha\, \pi(x) \prod_{k \neq j} \lambda_{Y_k}(x) . \qquad (4.29)$$

Distributed Belief Propagation



Causal support ↓



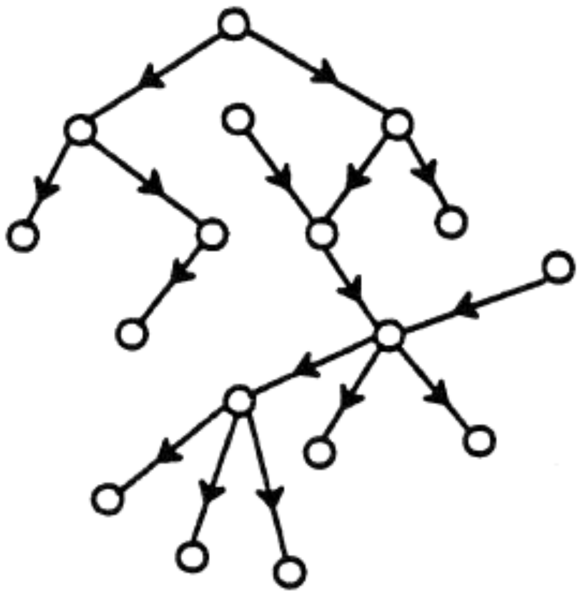
↑ Diagnostic support

Figure 4.14. Fragment of causal tree, showing incoming (solid arrows) and outgoing (broken arrows) messages at node X.

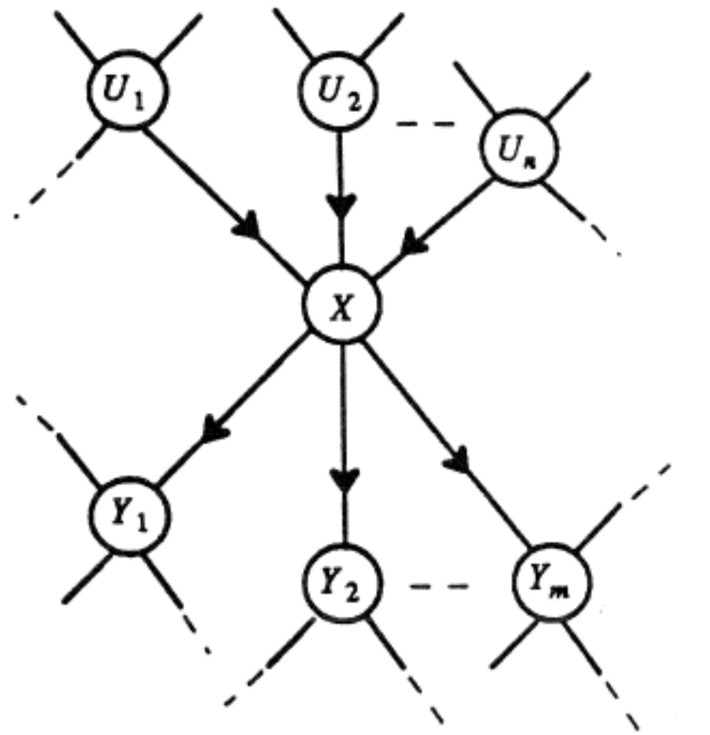
$$\mathbf{e} = \mathbf{e}_X^- \cup \mathbf{e}_X^+$$

e_X^- stands for the evidence contained in the tree rooted at X. e_X^+ stands for the evidence contained in the rest of the network.

Belief Propagation on Polytrees



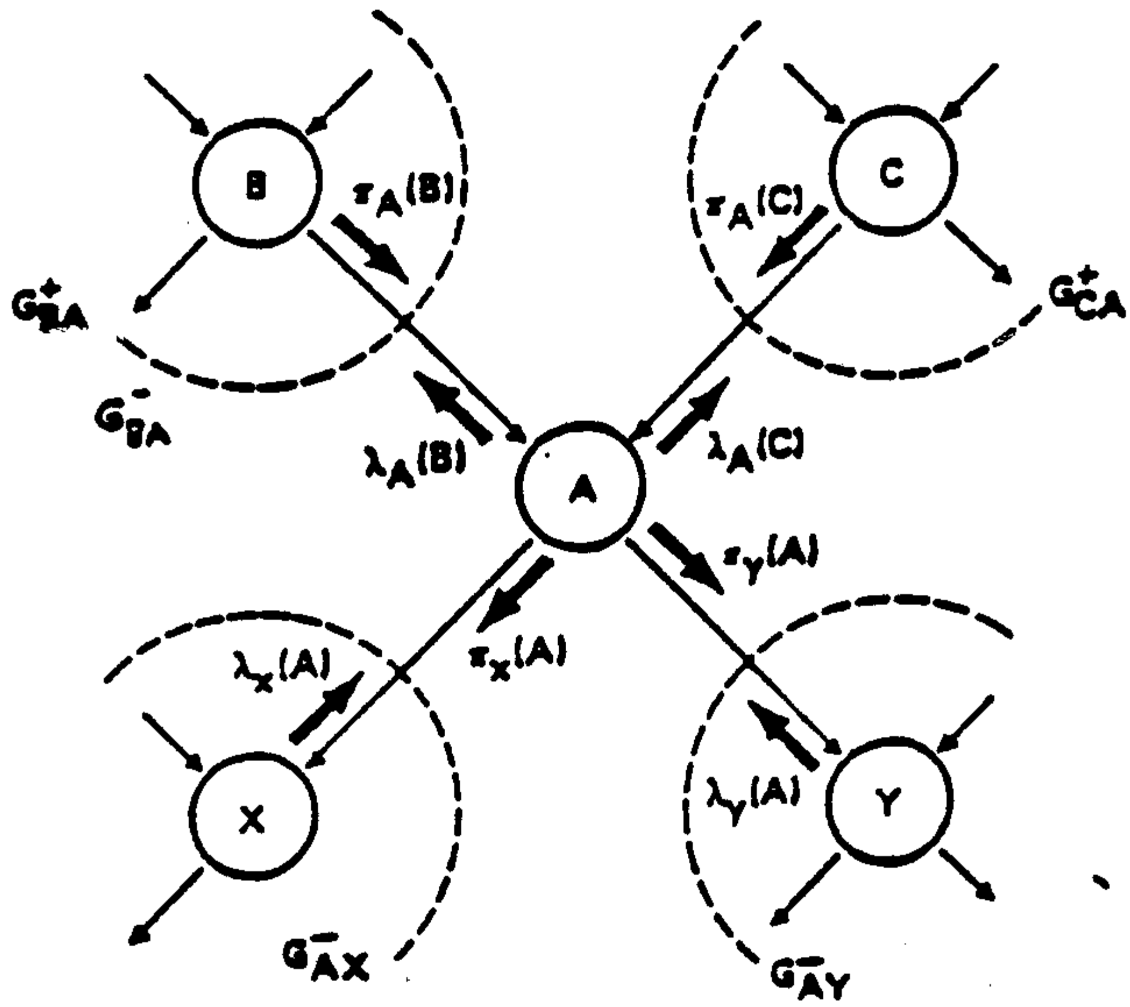
(a)



(b)

(a) Fragment of a polytree, (b) the parents and children of a typical r

Belief Propagation on Polytrees



Pearl (1982), (Belief Propogation)

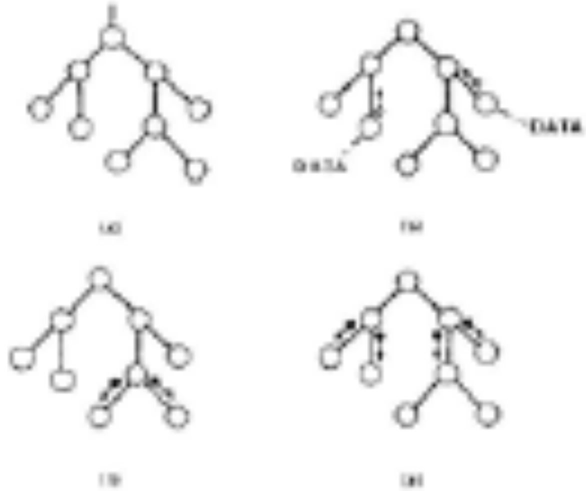


Figure 2

Properties of the Updating Scheme

1. The local computations required by the proposed scheme are efficient in both storage and time. For an m-ary tree with n states per node, each processor should store $n^2+mn+2n$ real numbers, and perform $2n^2+mn+2n$ multiplications per update. These expressions are on the order of the number of rules which each variable invokes.

Outline

- Bayesian networks from historical perspective
- Bayesian networks
- Belief propagation on trees
- From trees to graphs
- From Bayesian network to graphical models
- Some observation on loopy belief propagation

A Loopy Bayesian network

$$P(x_1,x_2,x_3,x_4,x_5,x_6) = P(x_6|x_5)\,P(x_5|x_2,x_3)\,P(x_4|x_1,x_2)\,P(x_3|x_1)\,P(x_2|x_1)\,P(x_1).$$

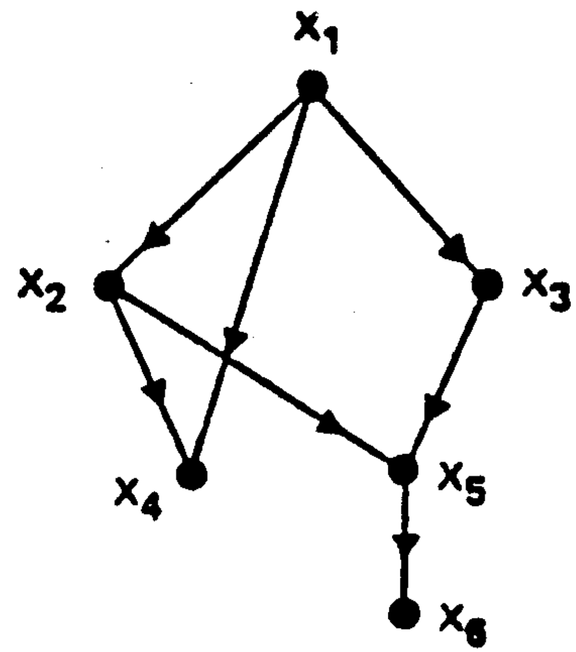
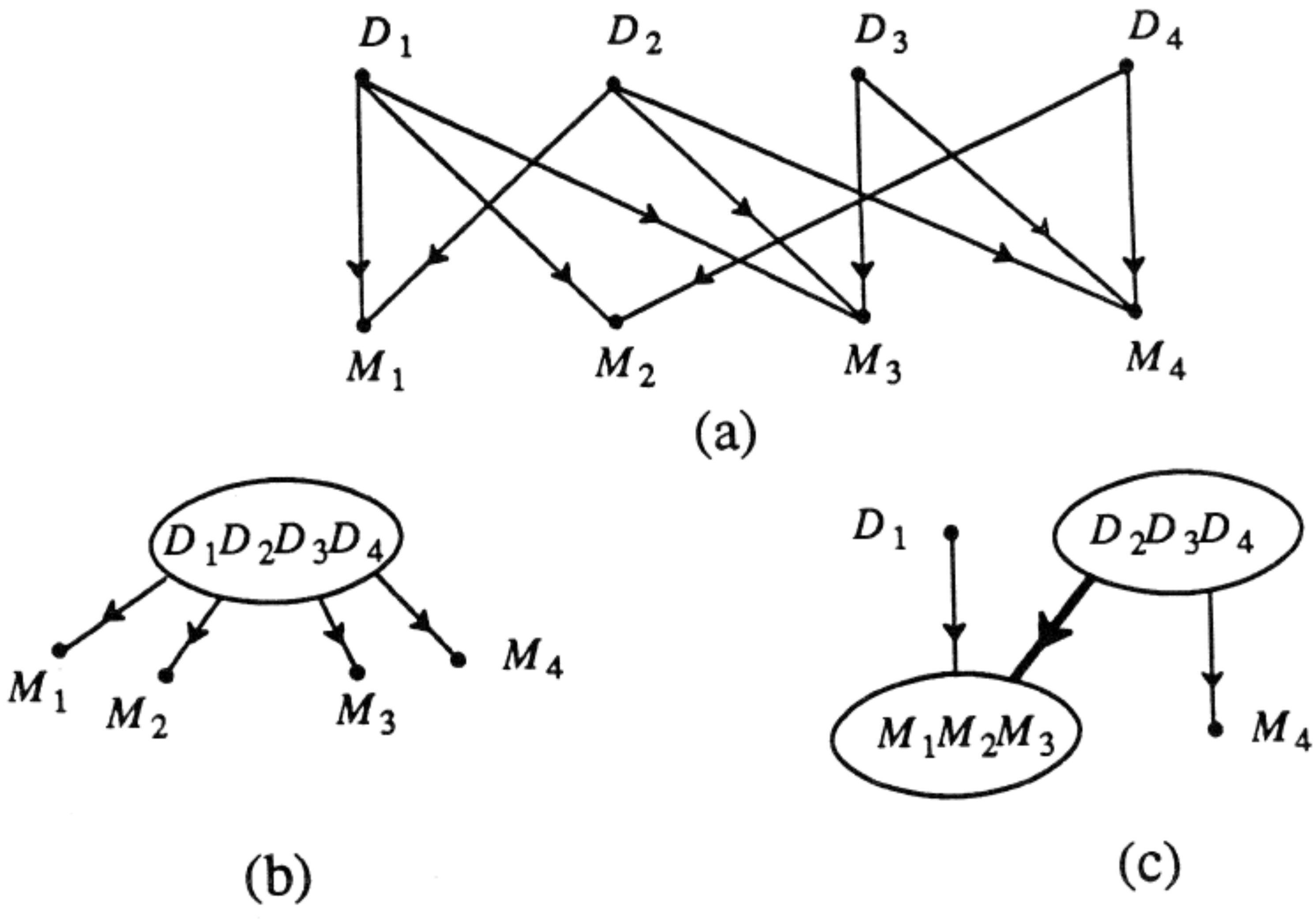


Figure 1

Coping with Loops (Pearl 1988)

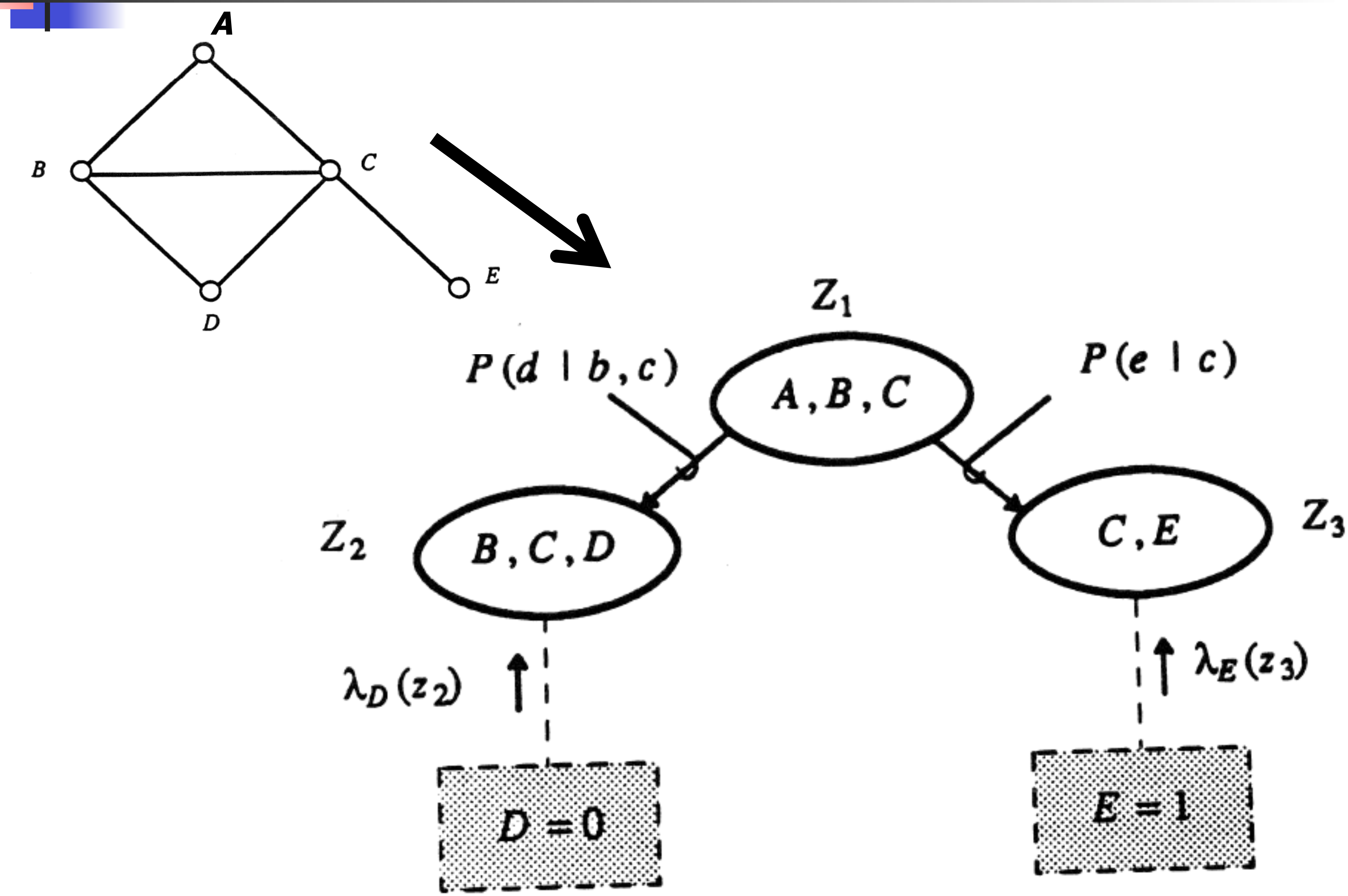
- Clustering methods (4.4.1)
  - Spigelhalter and Lauritsen: Junction-tree propagation (1988), Join-tree propagation (Pearl 1988)
- Conditioning schemes (4.4.2)
  - Loop-cutset scheme
- Stochastic simulation (Gibbs sampling) 4.4.3
- Loopy belief propagation (exercise)

(Poly)-Tree Clustering

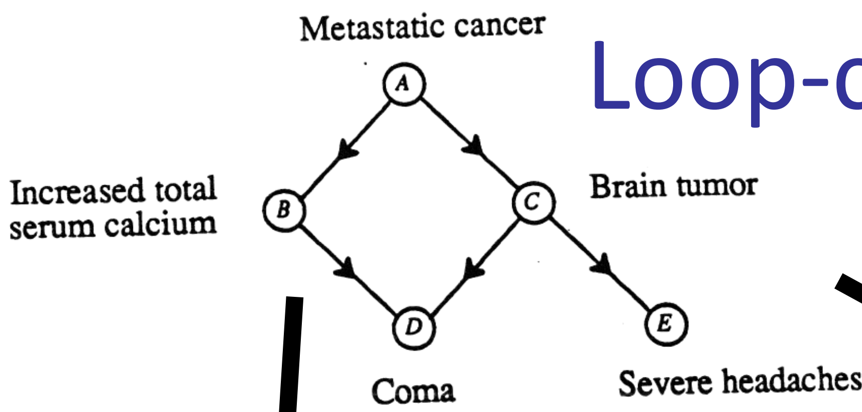


Clustering network (a) into a (b) tree or (c) polytree

Join-tree Clustering



Loop-cutset Conditioning



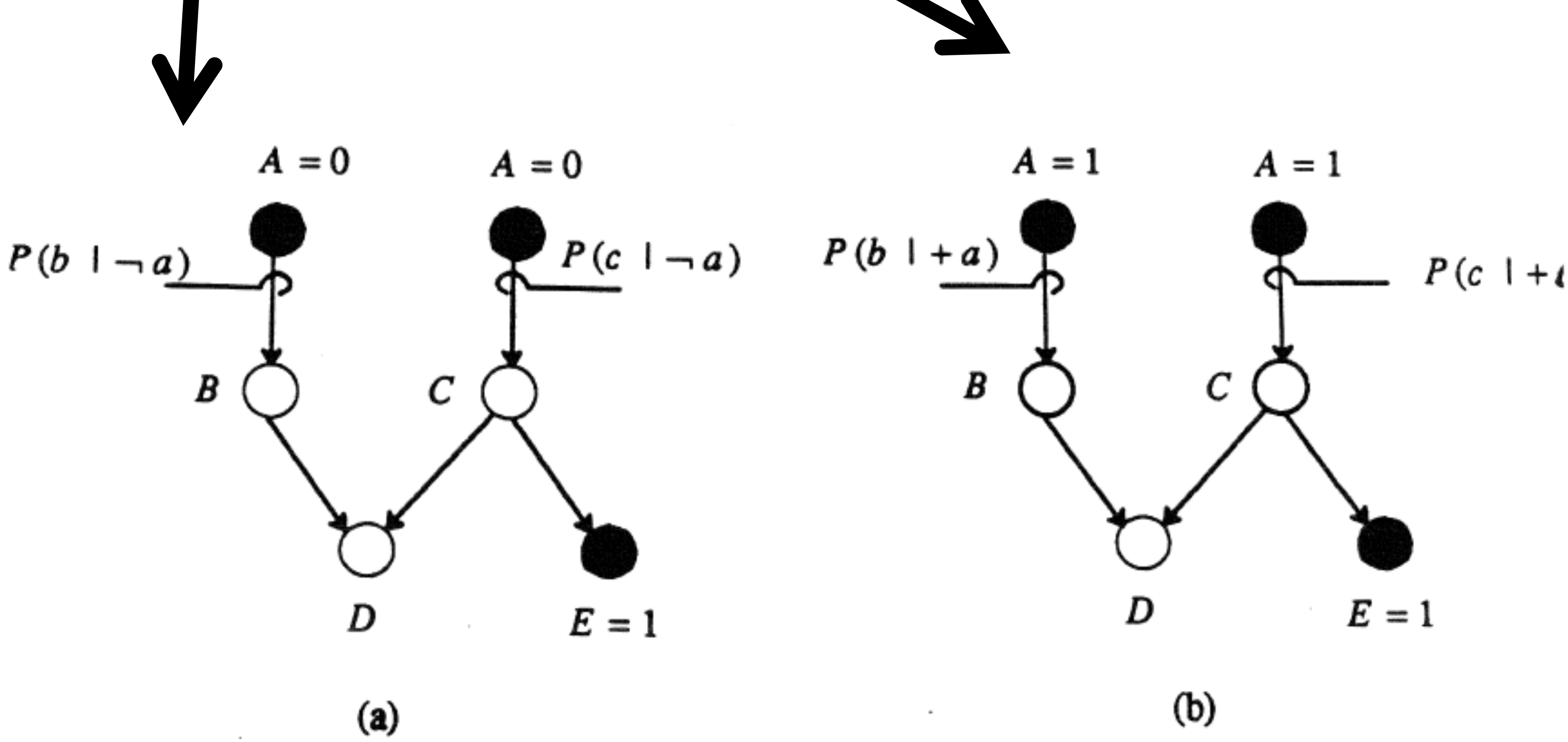
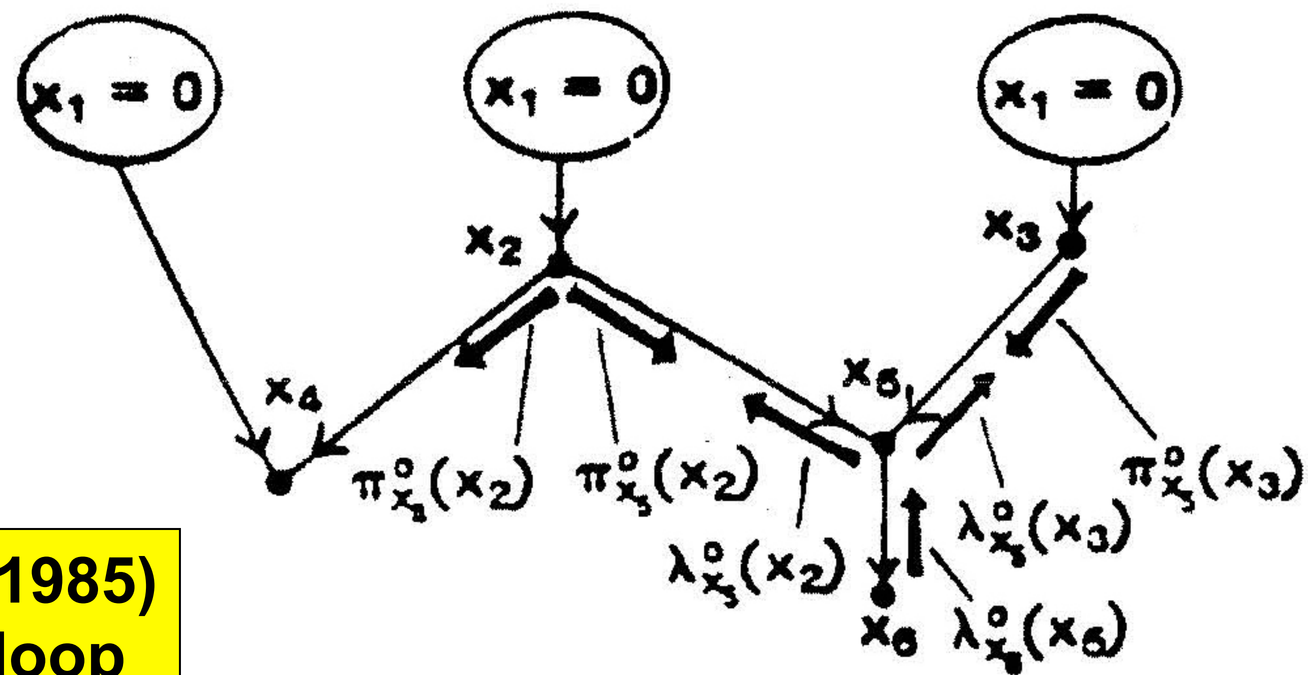


Figure 4.28. The multiply connected network of Figure 4.23 is decomposed into two polytrees corresponding to the two instantiations of A.

The Cutset Conditioning

$$\pi^0_{x_6}(x_5) = \sum_{x_2,x_3=0,1} P(x_5|x_2,x_3)\pi^0_{x_5}(x_2)\pi^0_{x_5}(x_3)$$

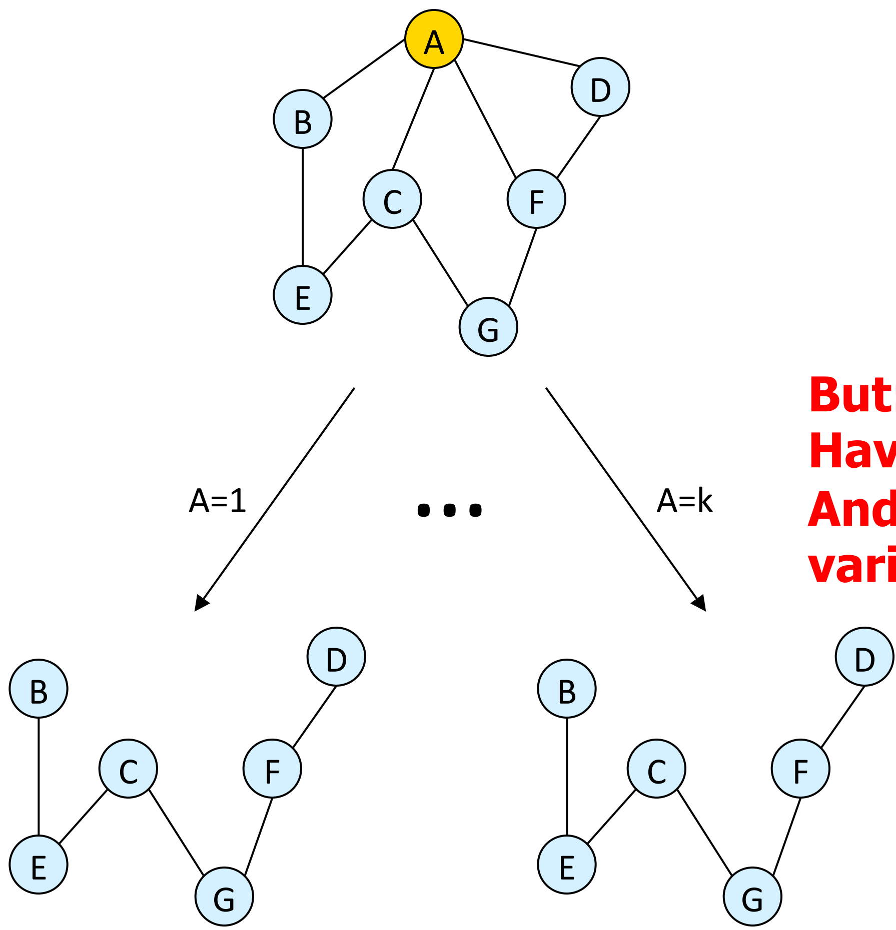
$$\pi^1_{x_6}(x_5) = \sum_{x_2,x_3=0,1} P(x_5|x_2,x_3)\pi^1_{x_5}(x_2)\pi^1_{x_5}(x_3)$$



Bayes Net (1985)
Breaking a loop

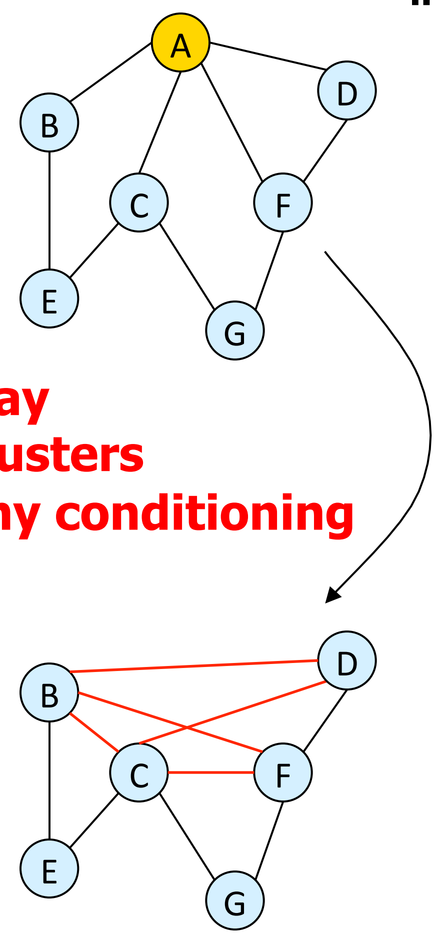
Search vs. clustering

Search (conditioning) **Exponential in cutset**



k "sparser" problems

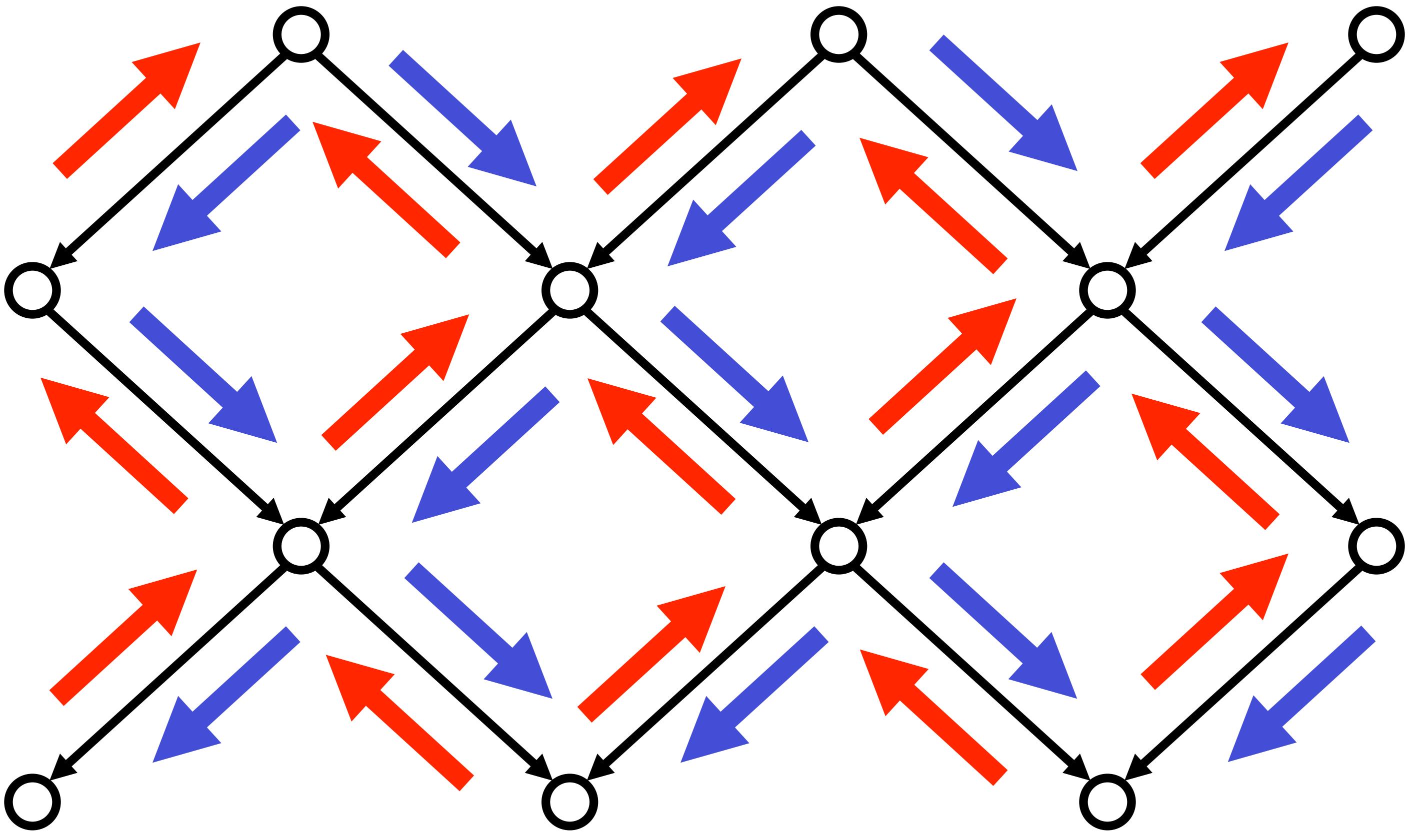
Inference (elimination) **Expponential in w**



1 "denser" problem

But: The graph may Have too large clusters And need too many conditioning variables

Belief Propagation when there are Loops



Some Applications

Human thought processes

Enable learning of BN
From data

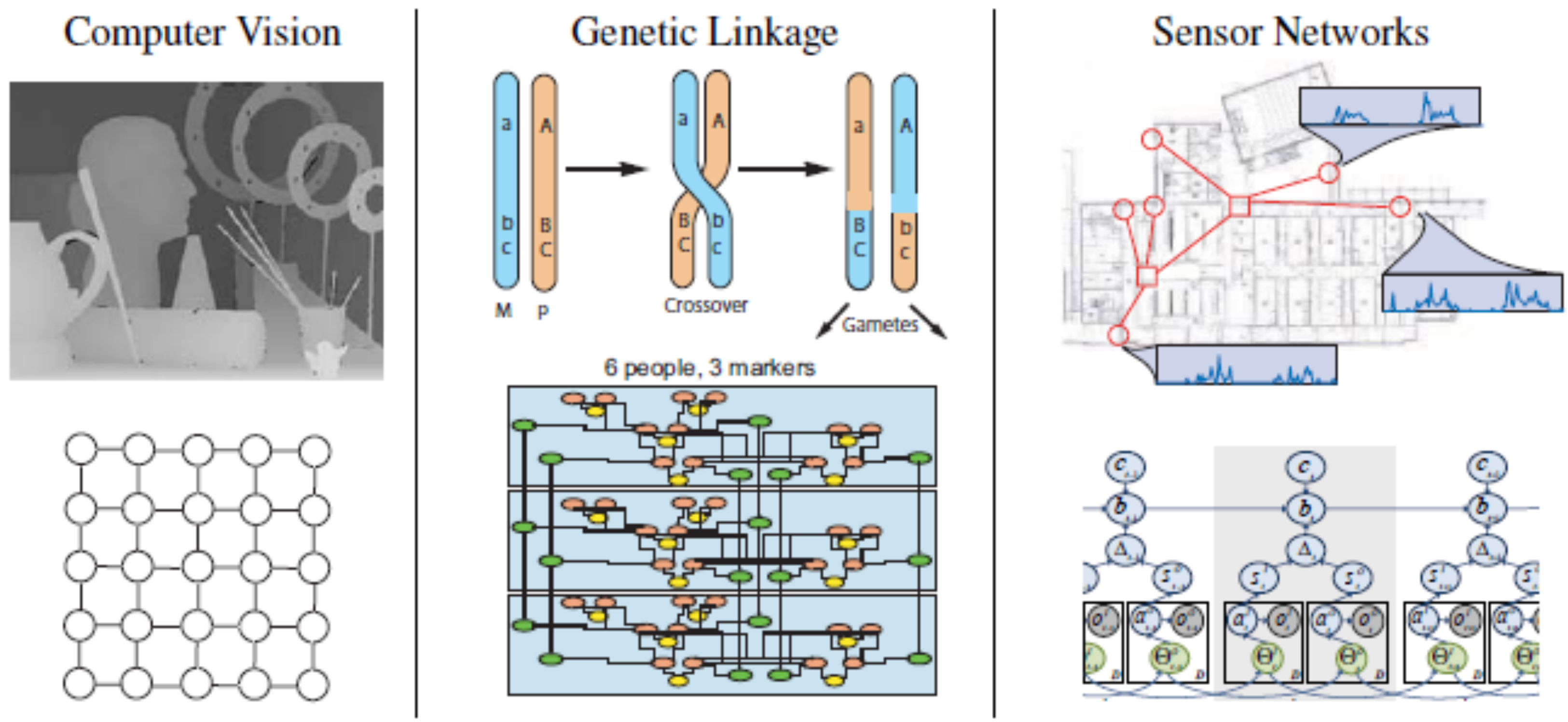


Figure 1: Application areas and graphical models used to represent their respective systems: (a) Finding correspondences between images, including depth estimation from stereo; (b) Genetic linkage analysis and pedigree data; (c) Understanding patterns of behavior in sensor measurements using spatio-temporal models.

Outline

- Bayesian networks from historical perspective
- Bayesian Networks
- Belief propagation on trees
- From trees to graphs
- From Bayesian network to graphical models; general exact and approximate algorithms
- Some observation on loopy belief propagation

Constraint Networks

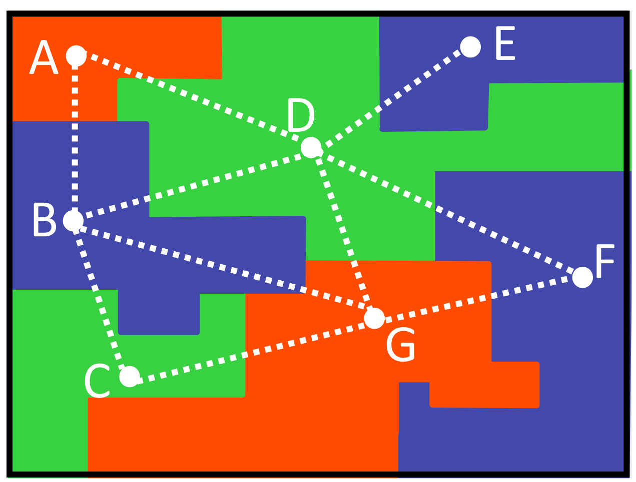
Map coloring

Variables: countries (A B C etc.)

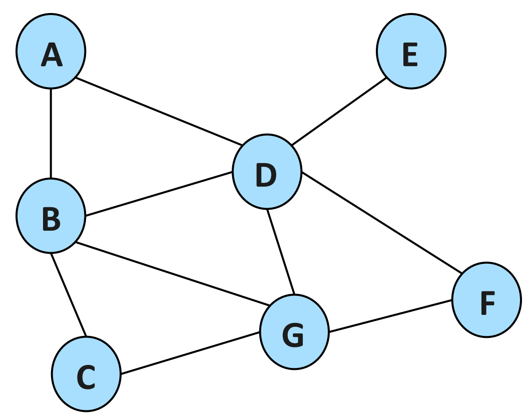
Values: colors (red green blue)

Constraints: $A \neq B$, $A \neq D$, $D \neq E$, …

A	B
red	green
red	yellow
green	red
green	yellow
yellow	green
yellow	red



Constraint graph



Constraint Satisfaction Tasks

Example: map coloring

Variables - countries (A,B,C,etc.)

Values - colors (e.g., red, green, yellow)

Constraints:

$A \neq B, \; A \neq D, \; D \neq E,$ *etc.*

A	B	C	D	E…
red	green	red	green	blue
red	blu	gre	green	blue
…	…	…	…	gree n
…	…	…	…	red
red	blu e	red	green	red

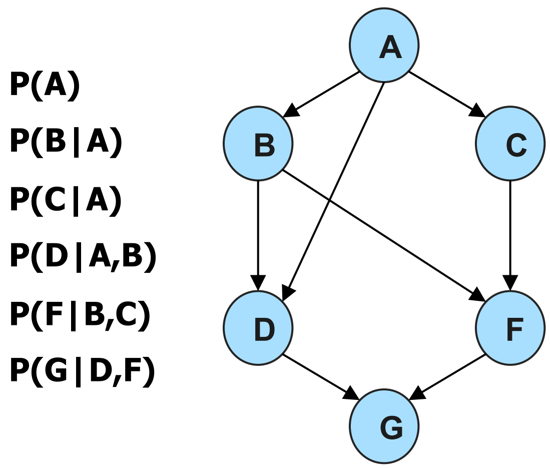
Are the constraints consistent?

Find a solution, find all solutions

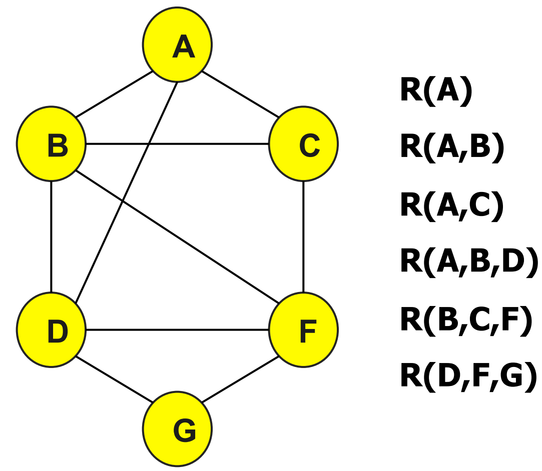
Count all solutions

Find a good solution

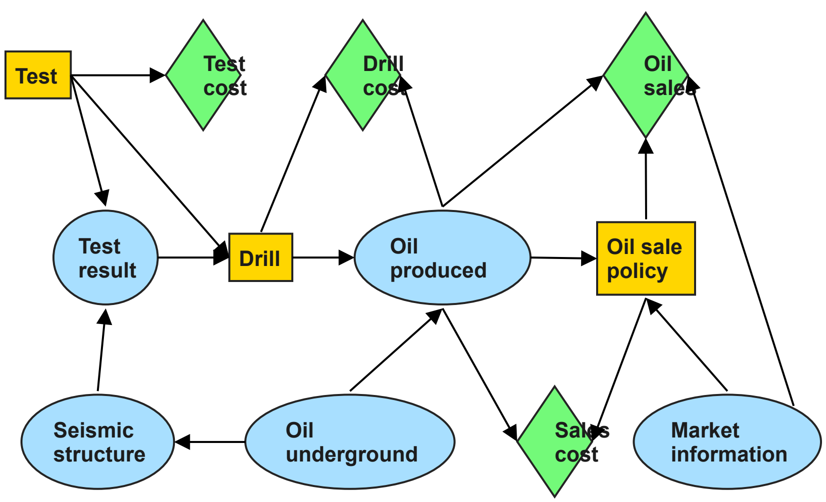
Graphical Models



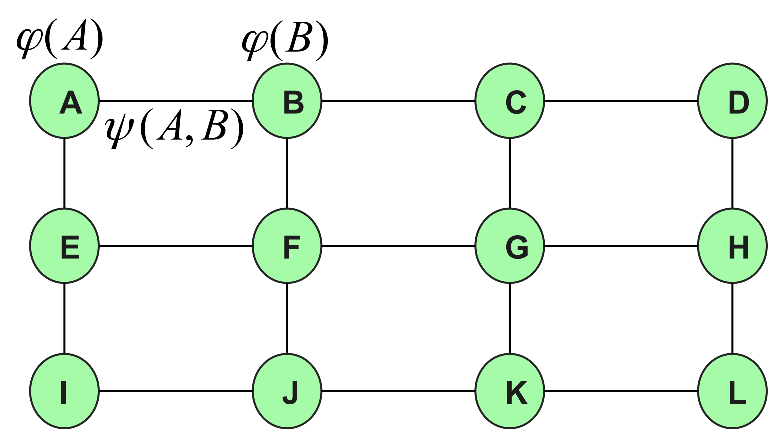
a) Belief network



b) Constraint network



c) Influence diagram

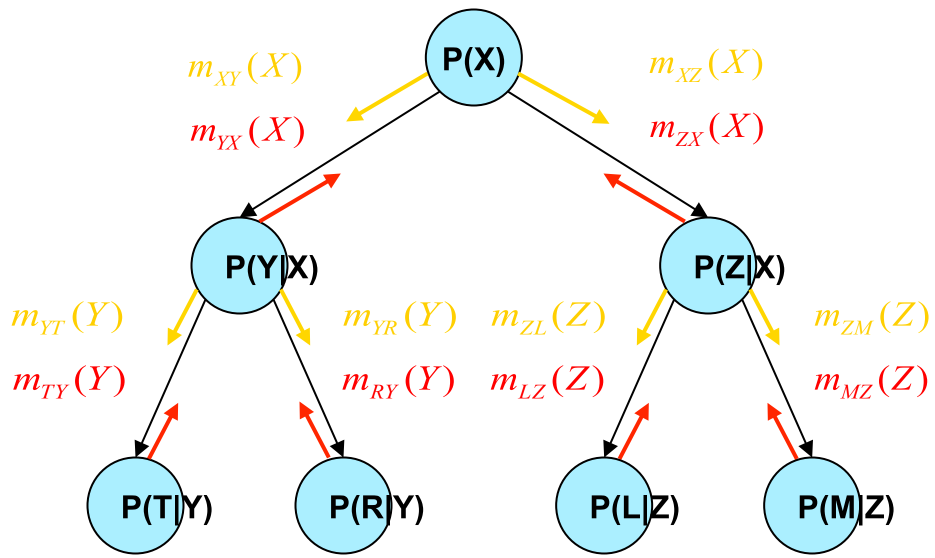


d) Markov network

Tree Solving is Easy

Belief updating (sum-prod)

CSP – consistency (projection-join)



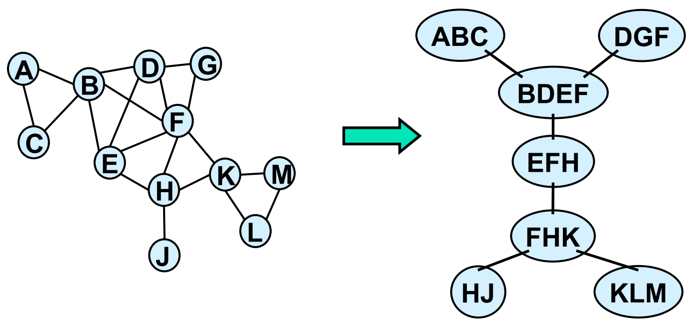
MPE (max-prod)

#CSP (sum-prod)

Trees are processed in linear time and memory

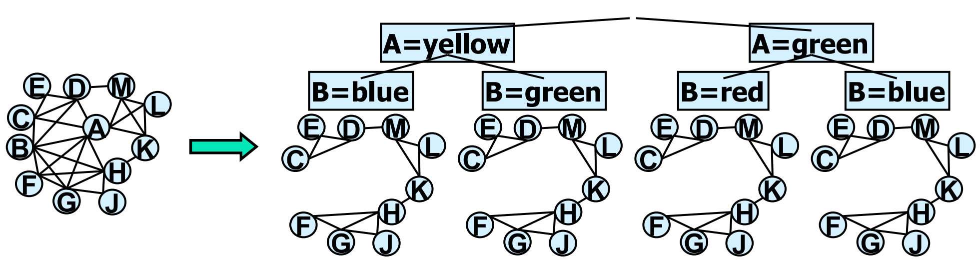
Inference vs. Conditioning

- **By Inference (thinking)**



Exponential in treewidth
Time and memory

- **By Conditioning (guessing)**

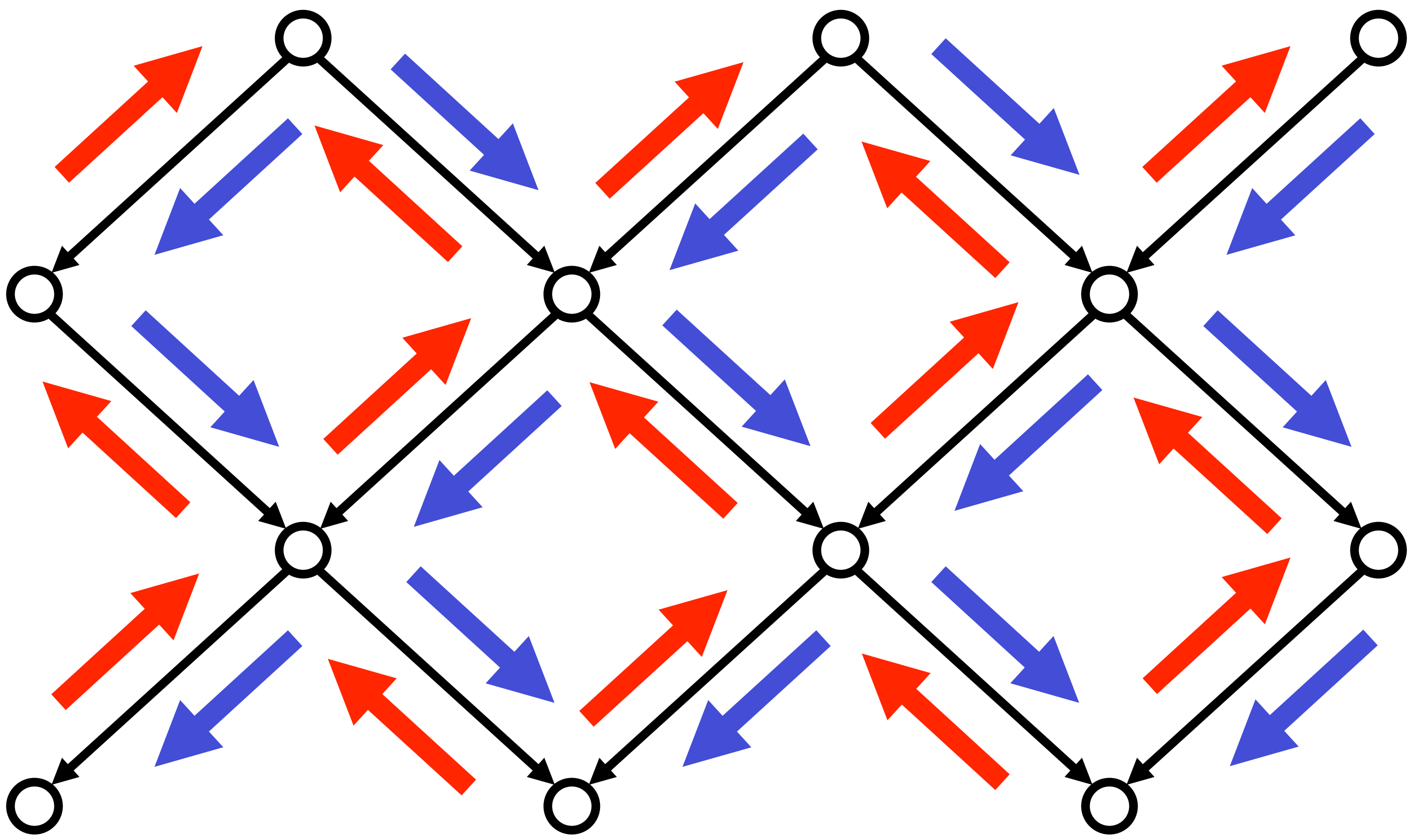


Exponential in cycle-cutset
Time-wise, linear memory

Outline

- Bayesian networks from historical perspective
- Bayesian Networks
- Belief propagation on trees
- From trees to graphs
- From Bayesian network to graphical models
- Some observations on loopy belief propagation

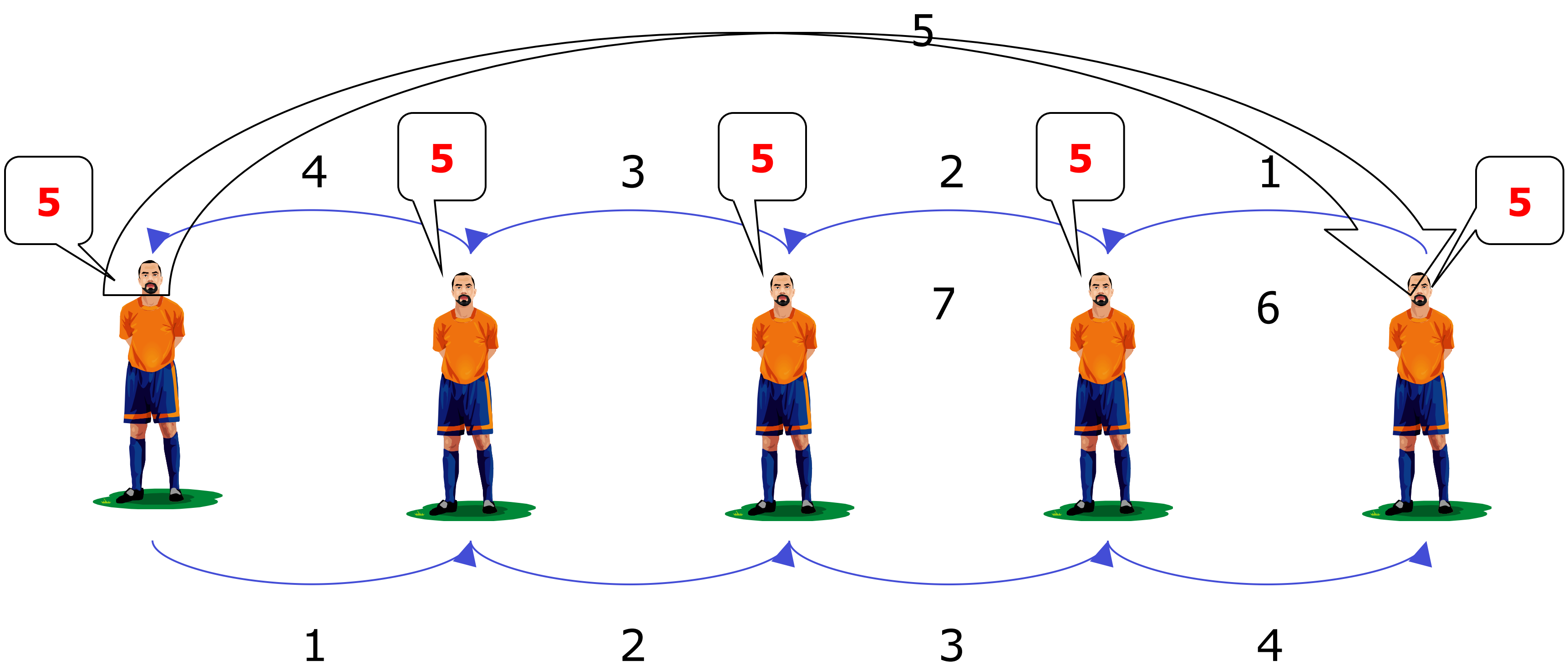
Belief Propagation when there are Loops

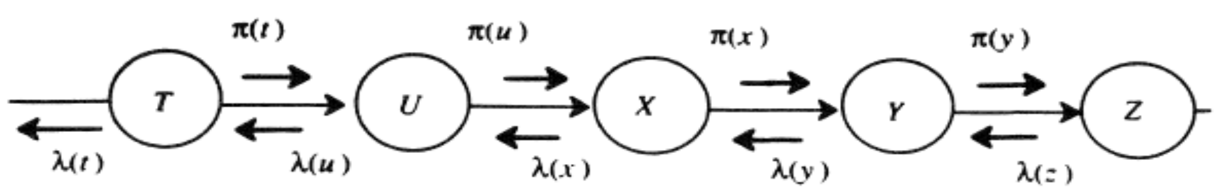


Distributed Belief Propagation

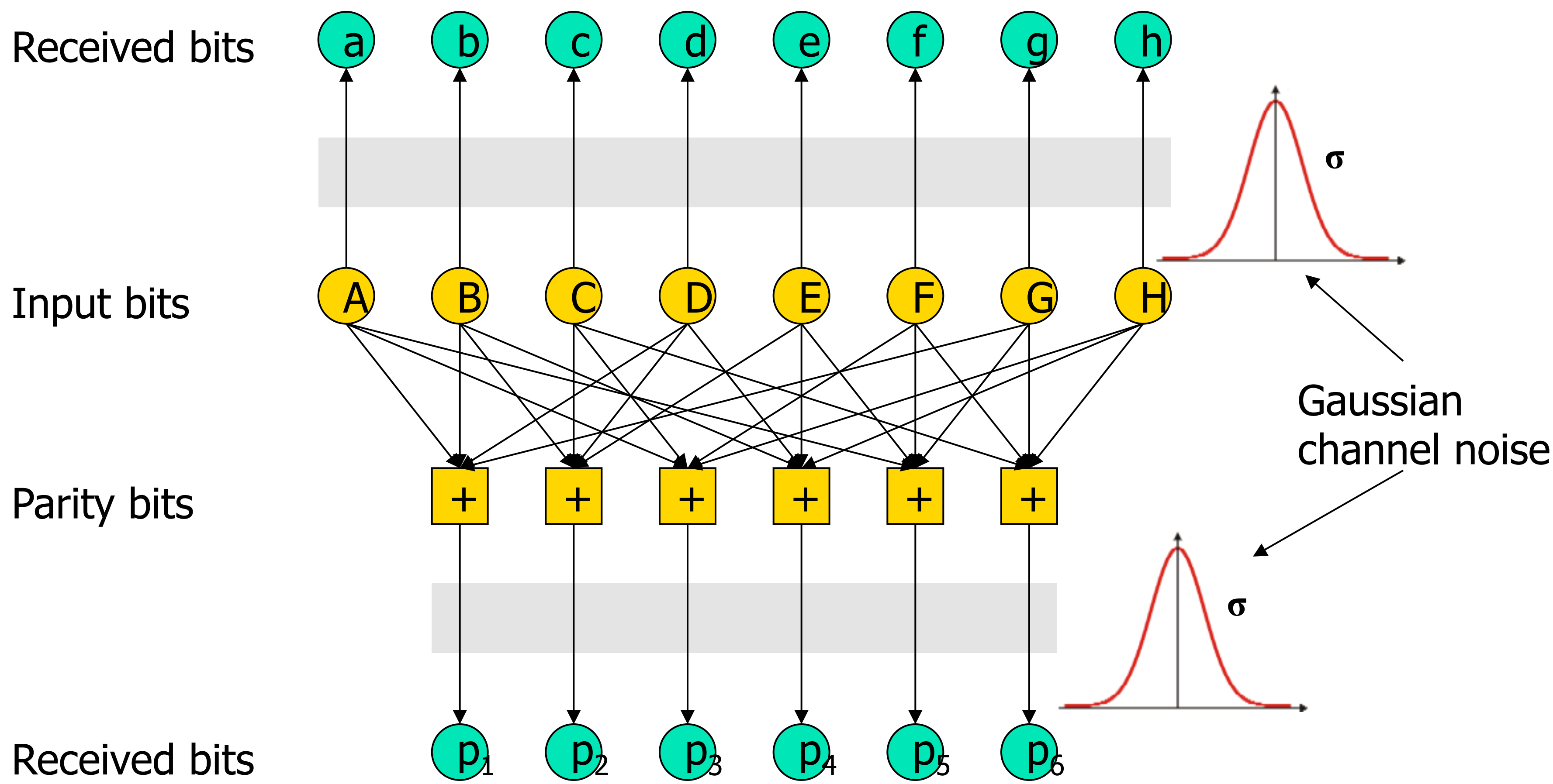
The essence of belief propagation is to make global information be shared locally by every entity

How many people?

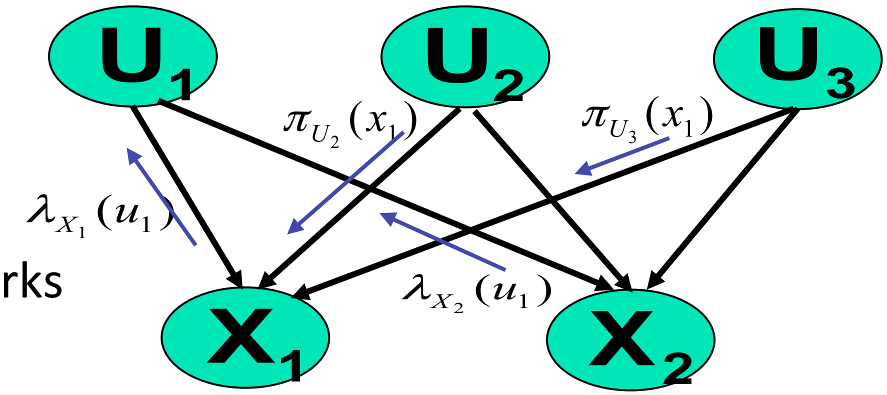




Linear Block Codes



Belief Propagation on Loopy Graphs



- Pearl (1988): use of BP to loopy networks
- McEliece, et. Al 1988: BP's success on coding networks
- Lots of research into convergence … and accuracy if convergence, still:
  - Why BP works well for coding networks
  - When does BP converge?
  - How accurate is it when converged?
  - Can we characterize other good problem classes?
  - Can we have any guarantees on accuracy (if we have converges)

Constraint Propagation

Arcs-consistency

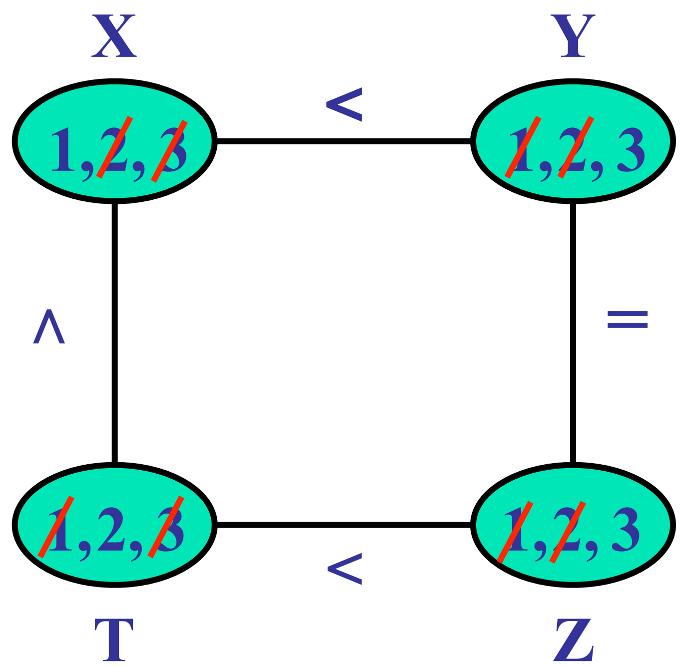
1 ≤ X, Y, Z, T ≤ 3

X < Y

Y = Z

T < Z

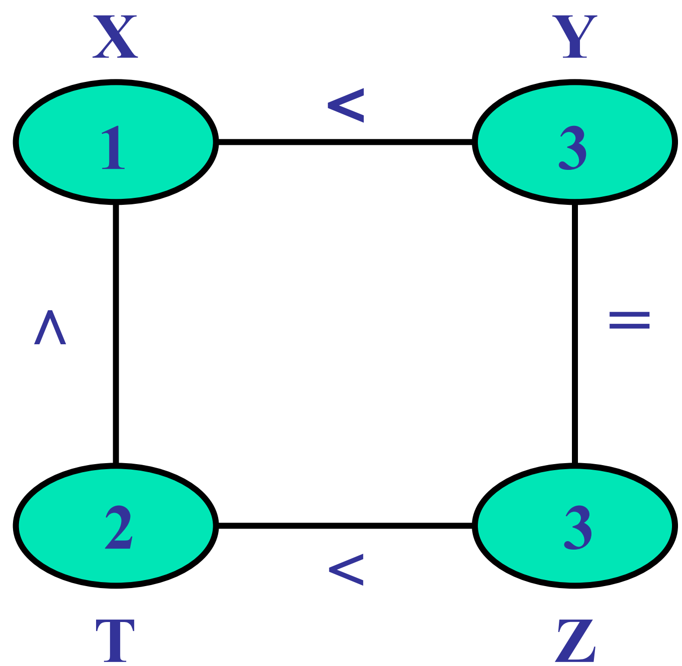
X ≤ T



Arc-consistency

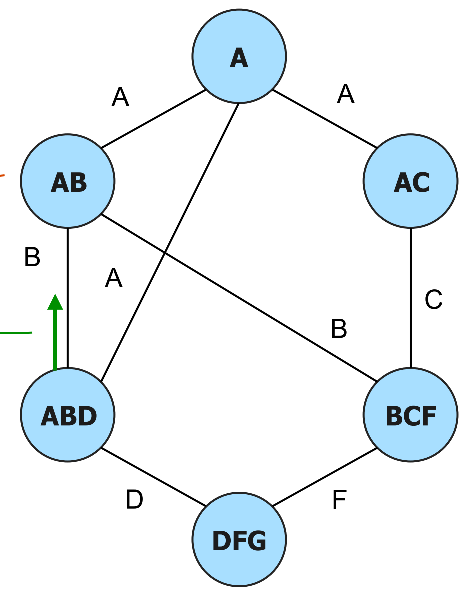
$$R_X \leftarrow \prod_X R_{XY} \bowtie D_Y$$

$1 \le X, Y, Z, T \le 3$
$X < Y$
$Y = Z$
$T < Z$
$X \le T$



Distributed Relational Arc-consistency

- Can be applied to nodes being the constraints (dual network):



$$h_i^j \leftarrow \pi_{l_{ij}}(R_i \bowtie (\bowtie_{k\in ne(i)} h_k^i)) \qquad (1)$$

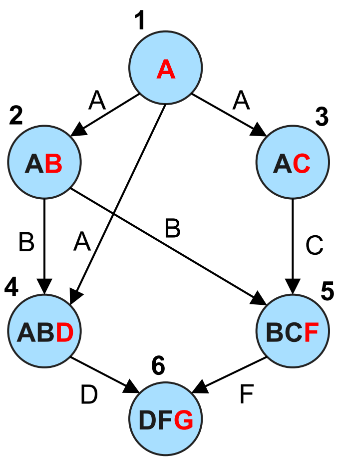
$$R_i \leftarrow R_i \cap (\bowtie_{k\in ne(i)} h_k^i) \qquad (2)$$

Flattening the Bayesian Network

A	P(A)
1	.2
2	.5
3	.3
…	0

A	B	P(B|A)
1	2	.3
1	3	.7
2	1	.4
2	3	.6
3	1	.1
3	2	.9
…	…	0

A	C	P(C|A)
1	2	1
3	2	1
…	…	0



A	B	D	P(D|A,B)
1	2	3	1
1	3	2	1
2	1	3	1
2	3	1	1
3	1	2	1
3	2	1	1
…	…	…	0

B	C	F	P(F|B,C)
1	2	3	1
3	2	1	1
…	…	…	0

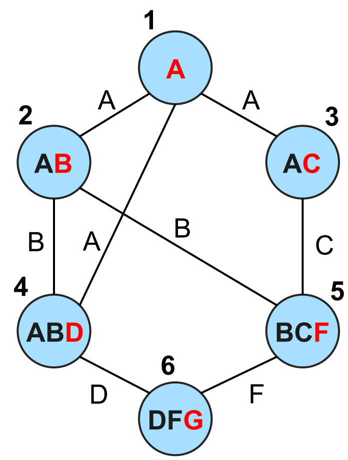
D	F	G	P(G|D,F)
1	2	3	1
2	1	3	1
…	…	…	0

Belief network

A
1
2
3

A	B
1	2
1	3
2	1
2	3
3	1
3	2

A	C
1	2
3	2



A	B	D
1	2	3
1	3	2
2	1	3
2	3	1
3	1	2
3	2	1

B	C	F
1	2	3
3	2	1

D	F	G
1	2	3
2	1	3

Flat constraint network

THEOREM 11. *Given a belief network* $\mathcal{B} = \langle X, D, G, P \rangle$, *where* $X = \{X_1, \ldots, X_n\}$, *for any tuple* $x = (x_1, \ldots, x_n)$: $P_{\mathcal{B}}(x) > 0 \Leftrightarrow x \in sol(flat(\mathcal{B}))$, *where* $sol(flat(\mathcal{B}))$ *is the set of solutions of* $flat(B)$.

Belief Zero Propagation = Arc-consistency

IBP over belief propagation applies relational arc-consistency on the flat network

$$h_i^j = \sum_{elim(i,j)} (p_i \cdot (\prod_{\{k \in ne_j(i)\}} h_k^i))$$

$$h_i^j = \pi_{l_{ij}} (R_i \bowtie (\bowtie_{k \in ne(i)} h_k^i))$$

A	B	P(B|A)
1	2	>0
1	3	>0
2	1	>0
2	3	>0
3	1	>0
3	2	>0
…	…	0

h_1^2

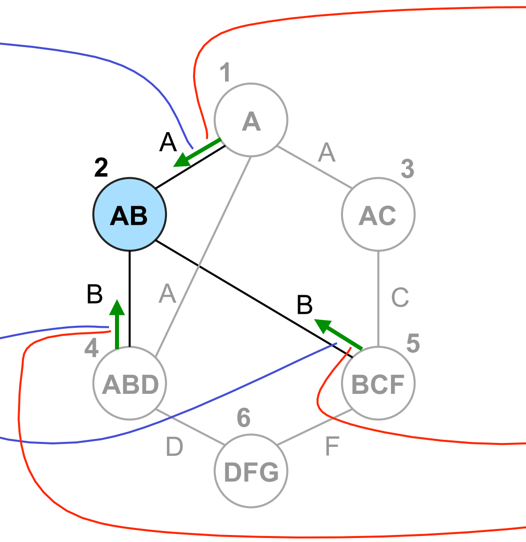
A	$h_1^2(A)$
1	>0
2	>0
3	>0
…	0

h_4^2

B	$h_1^2(B)$
1	>0
2	>0
3	>0
…	0

h_5^2

B	$h_1^2(B)$
1	>0
3	>0
…	0



A	B
1	2
1	3
2	1
2	3
3	1
3	2

h_1^2

A
1
2
3

h_4^2

B
1
2
3

h_5^2

B
1
3

Updated belief:

$$Bel(A,B) = P(B \mid A) \cdot h_1^2 \cdot h_4^2 \cdot h_5^2 =$$

=

A	B	Bel (A,B)
1	3	>0
2	1	>0
2	3	>0
3	1	>0
…	…	0

Updated relation:

$$R(A,B) = R(A,B) \bowtie h_1^2 \bowtie h_4^2 \bowtie h_5^2 =$$

=

A	B
1	3
2	1
2	3
3	1

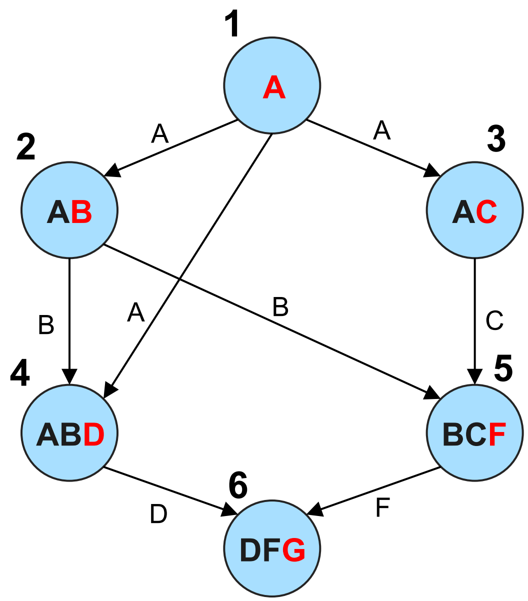
An Example

R_1

A	P(A)
1	.2
2	.5
3	.3
…	0

R_2

A	B	P(B|A)
1	2	.3
1	3	.7
2	1	.4
2	3	.6
3	1	.1
3	2	.9
…	…	0



R_3

A	C	P(C|A)
1	2	1
3	2	1
…	…	0

R_4

A	B	D	P(D|A,B)
1	2	3	1
1	3	2	1
2	1	3	1
2	3	1	1
3	1	2	1
3	2	1	1
…	…	…	0

R_5

B	C	F	P(F|B,C)
1	2	3	1
3	2	1	1
…	…	…	0

R_6

D	F	G	P(G|D,F)
1	2	3	1
2	1	3	1
…	…	…	0

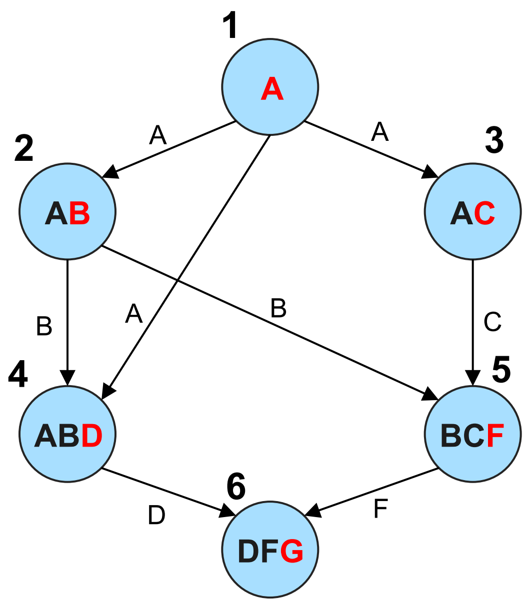
Belief Propagation Example: Iteration 1

R_1

A	P(A)
1	>0
3	>0
…	0

R_2

A	B	P(B|A)
1	3	1
2	1	>0
2	3	>0
3	1	1
…	…	0



R_3

A	C	P(C|A)
1	2	1
3	2	1
…	…	0

R_4

A	B	D	P(D|A,B)
1	3	2	1
2	3	1	1
3	1	2	1
3	2	1	1
…	…	…	0

R_5

B	C	F	P(F|B,C)
1	2	3	1
3	2	1	1
…	…	…	0

R_6

D	F	G	P(G|D,F)
2	1	3	1
…	…	…	0

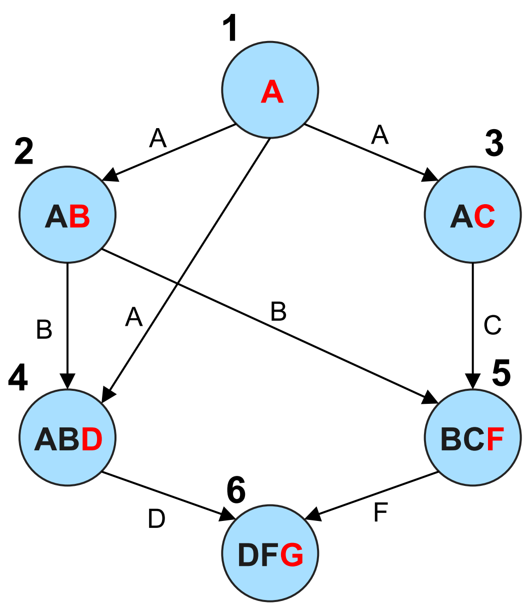
Belief Propagation Example: Iteration 2

R_1

A	P(A)
1	>0
3	>0
…	0

R_2

A	B	P(B|A)
1	3	1
3	1	1
…	…	0



R_3

A	C	P(C|A)
1	2	1
3	2	1
…	…	0

R_4

A	B	D	P(D|A,B)
1	3	2	1
3	1	2	1
…	…	…	0

R_5

B	C	F	P(F|B,C)
3	2	1	1
…	…	…	0

R_6

D	F	G	P(G|D,F)
2	1	3	1
…	…	…	0

Belief Propagation Example: Iteration 3

R_1

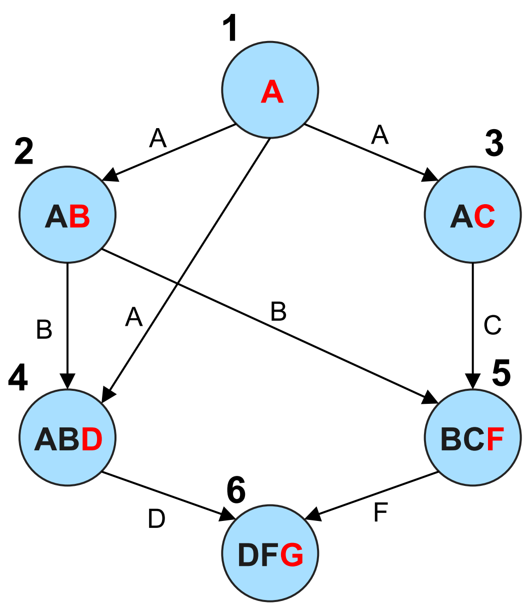
A	P(A)
1	>0
3	>0
…	0

R_2

A	B	P(B|A)
1	3	1
…	…	0

R_3

A	C	P(C|A)
1	2	1
3	2	1
…	…	0



R_4

A	B	D	P(D|A,B)
1	3	2	1
3	1	2	1
…	…	…	0

R_5

B	C	F	P(F|B,C)
3	2	1	1
…	…	…	0

R_6

D	F	G	P(G|D,F)
2	1	3	1
…	…	…	0

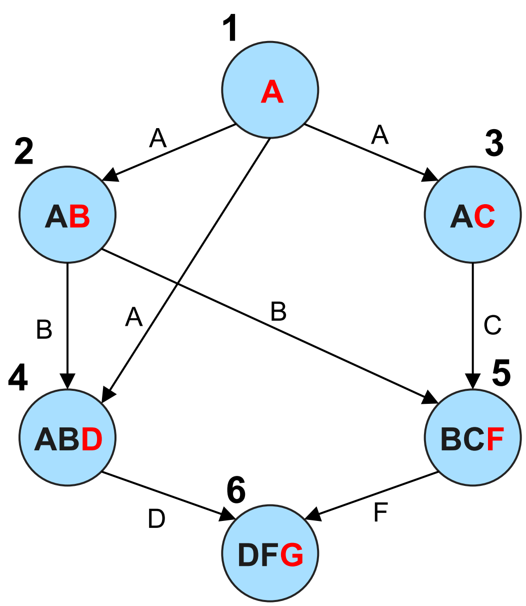
Belief Propagation Example: Iteration 4

R_1

A	P(A)
1	1
…	0

R_2

A	B	P(B|A)
1	3	1
…	…	0



R_3

A	C	P(C|A)
1	2	1
3	2	1
…	…	0

R_4

A	B	D	P(D|A,B)
1	3	2	1
…	…	…	0

R_5

B	C	F	P(F|B,C)
3	2	1	1
…	…	…	0

R_6

D	F	G	P(G|D,F)
2	1	3	1
…	…	…	0

Belief Propagation Example: Iteration 5

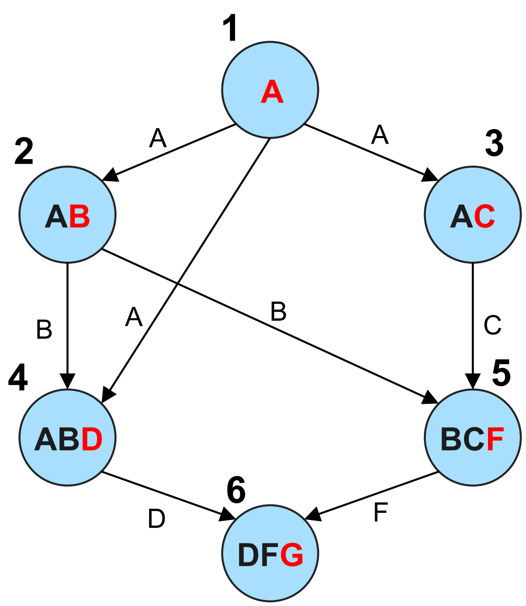
R_1

A	P(A)
1	1
…	0

A	B	C	D	F	G	Belief
1	3	2	2	1	3	1
…	…	…	…	…	…	0

R_2

A	B	P(B|A)
1	3	1
…	…	0



R_3

A	C	P(C|A)
1	2	1
…	…	0

R_4

A	B	D	P(D|A,B)
1	3	2	1
…	…	…	0

R_5

B	C	F	P(F|B,C)
3	2	1	1
…	…	…	0

R_6

D	F	G	P(G|D,F)
2	1	3	1
…	…	…	0

The Inference Power of BP for Zero Beliefs

- **Theorem:**

  Iterative BP performs arc-consistency on the flat network.

  (no more, no less)

- **Soundness:**
  - Inference of zero beliefs by IBP converges (in nk iterations, n variables, k= |domain|)
  - All the inferred zero beliefs are correct

- **Incompleteness:**
  - BP is as weak and as strong as arc-consistency (weak for graph coloring, strong for implicational constraints.)

- **Continuity Hypothesis:** IBP is sound for zero -→ IBP is accurate for extreme beliefs? Tested empirically

Experimental Results

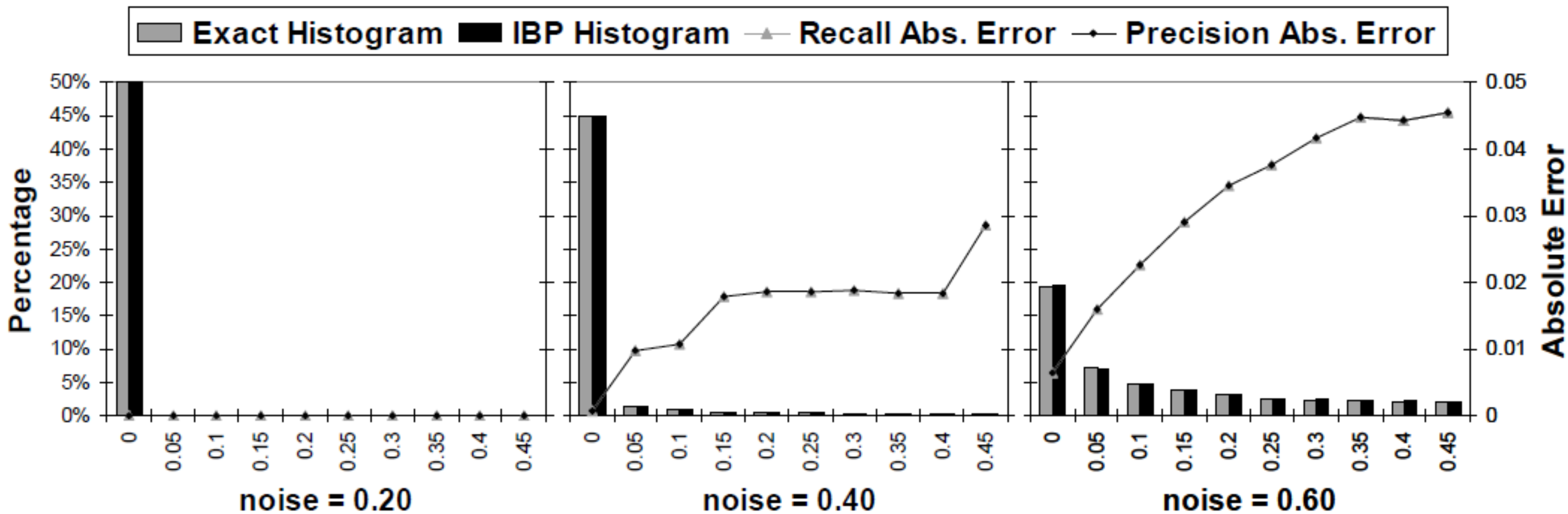
We investigated empirically if the results for zero beliefs extend to ε-small beliefs ($\varepsilon > 0$)

Have determinism?

- Network types:
  - YES
    - Coding
    - Linkage analysis*
    - Grids*
  - NO
    - Two-layer noisy-OR*
    - CPCS54, CPCS360
- Measures:
  - Exact/IJGP histogram
  - Recall absolute error
  - Precision absolute error
- Algorithms:
  - IBP
  - IJGP

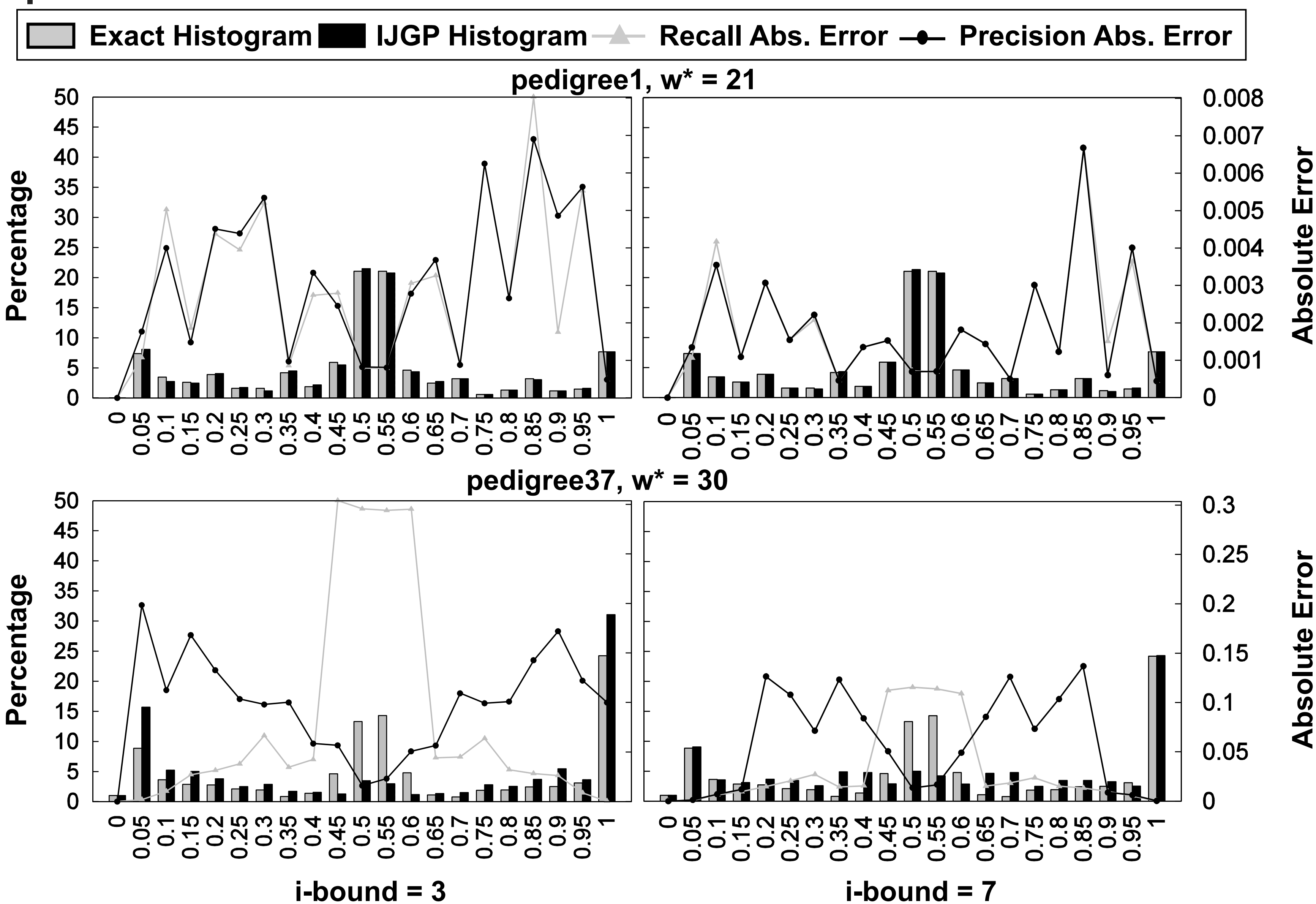
* Instances from the UAI08 competition

Networks with Determinism: Coding

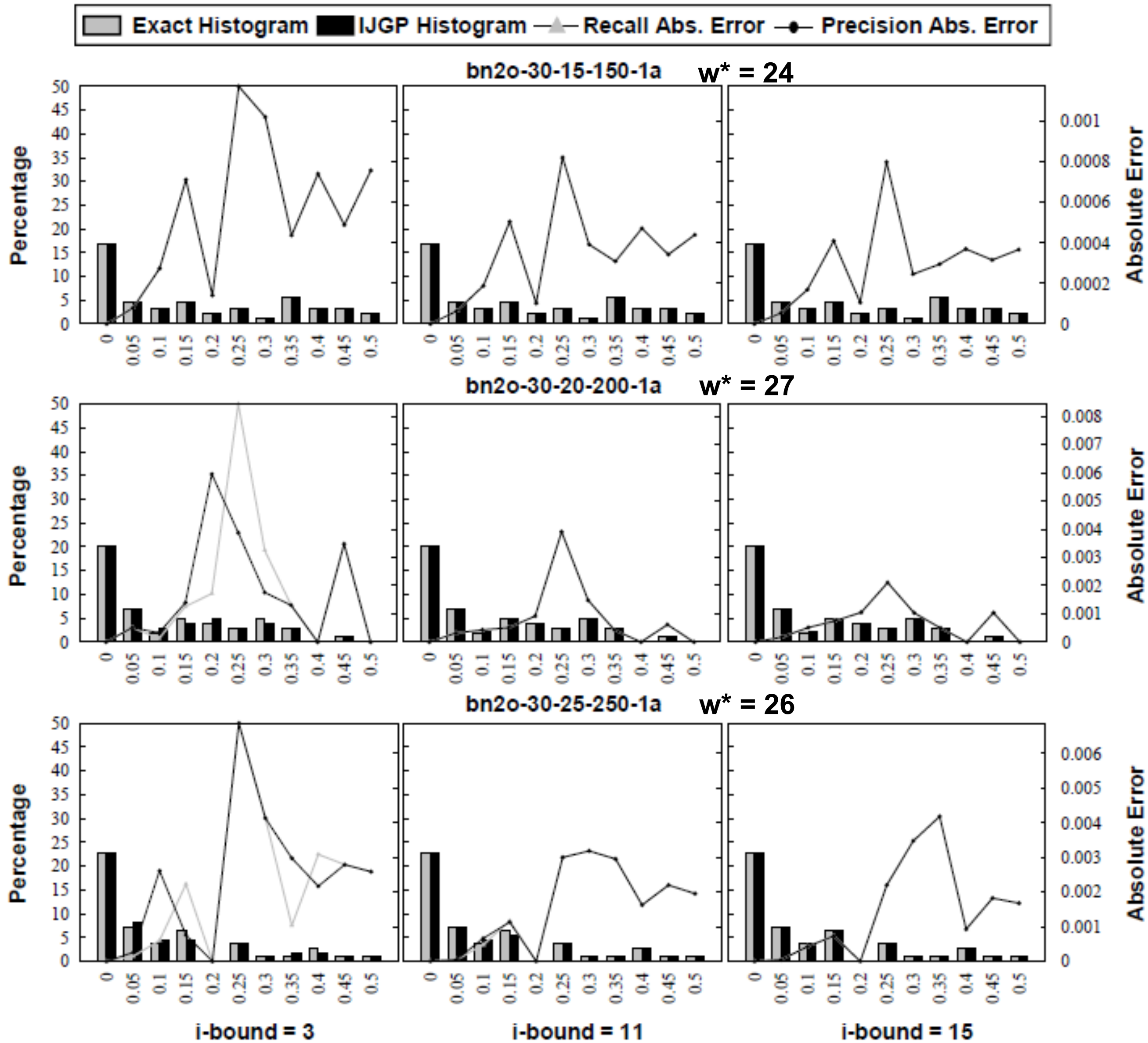


N=200, 1000 instances, w*=15

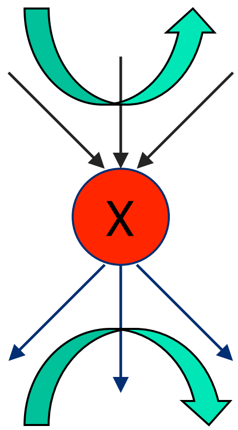
Nets with Determinism: Linkage

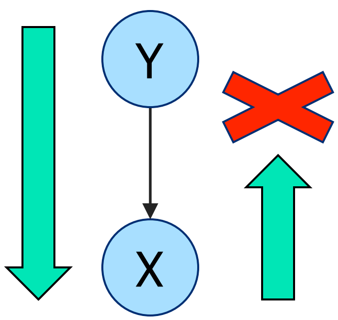


Nets w/o Determinism: bn2o



Cutset Phenomena & Irrelevant Nodes

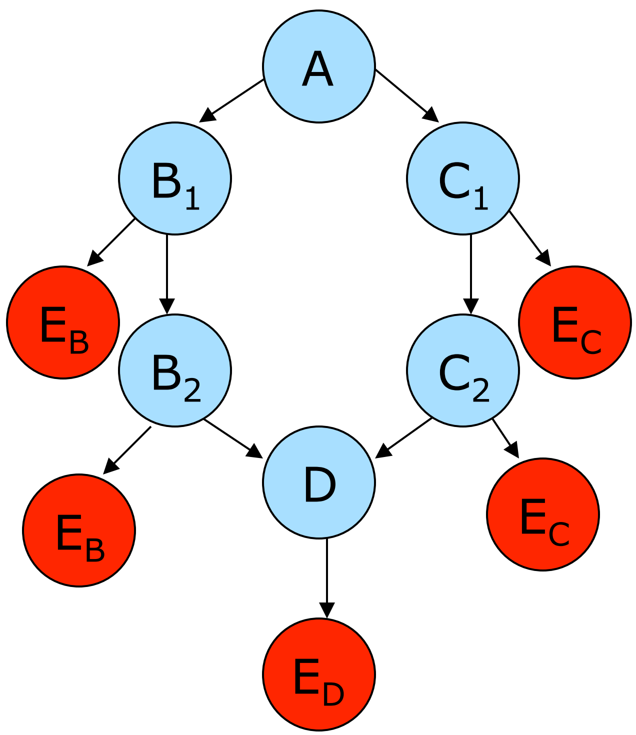


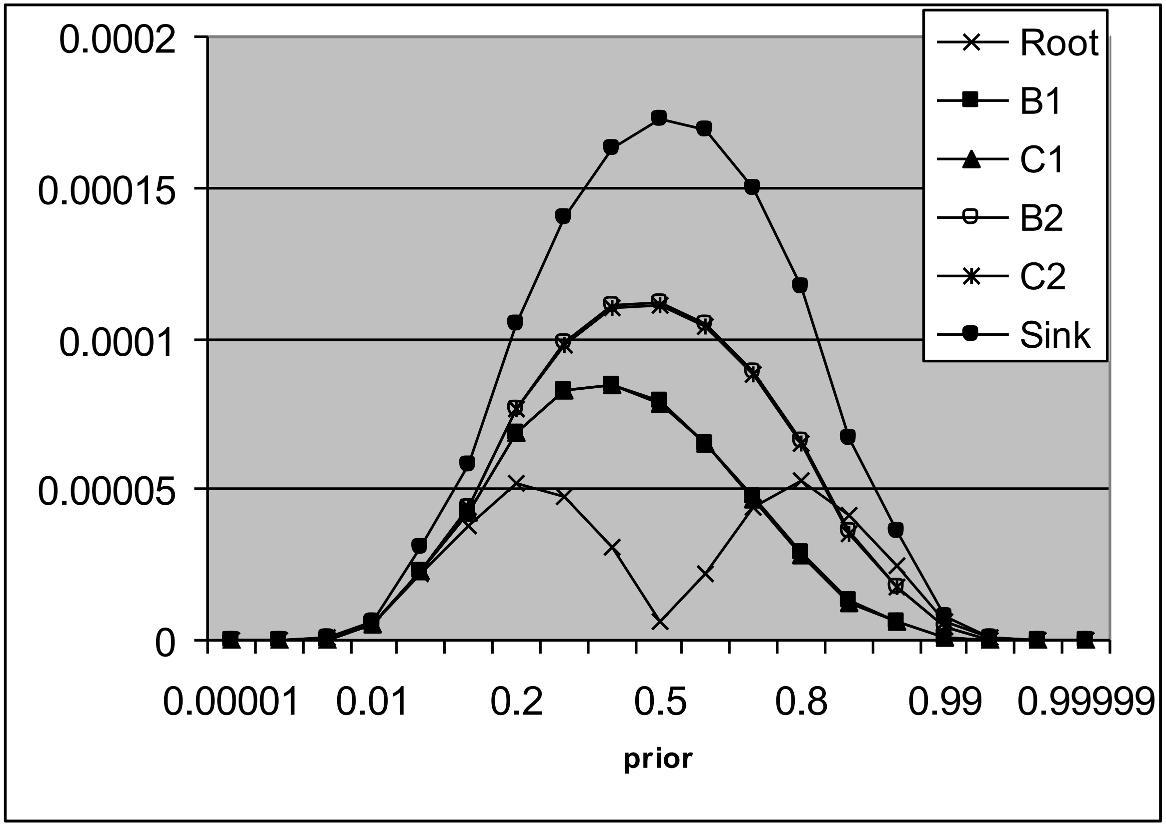


- **Observed variables** break the flow of inference
  - **BP is exact when evidence variables form a loop-cutset**
- **Unobserved variables** without observed descendents send zero-information to the parent variables – it is irrelevant
  - **In a network without evidence, BP converges in one iteration top-down**

Nodes with Extreme Support

Observed variables with xtreme priors or xtreme support can nearly-cut information flow:





Average Error vs. Priors

Conclusion: Networks with Determinism

BP converges & sound for zero beliefs

- IBP's power to infer zeros is as weak or as strong as arc-consistency
- However: inference of extreme beliefs can be wrong.
- Cutset property (Bidyuk and Dechter, 2000):
  - Evidence and inferred singleton act like cutset
  - If zeros are cycle-cutset, all beliefs are exact
  - Extensions to epsilon-cutset were supported empirically.
- IJGP is an anytime good tradeoff propagation scheme

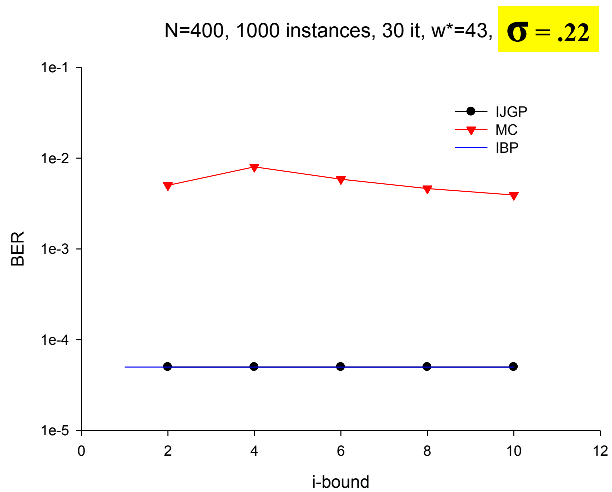
Thank You

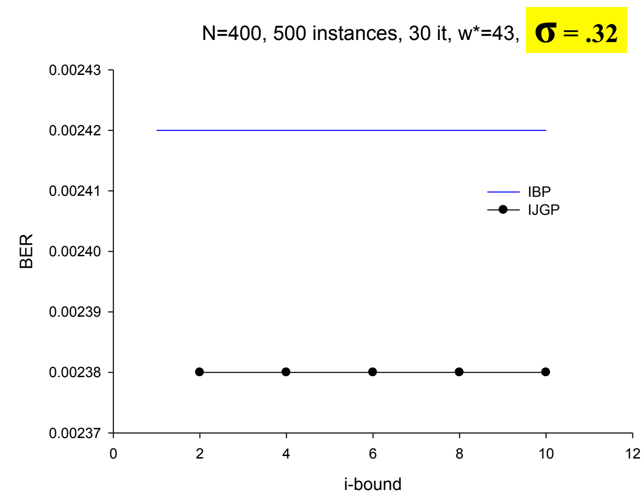
Rina Dechter, Bozhena Bidyuk, Robert Mateescu and Emma Rollon.
"On the Power of Belief Propagation: A Constraint Propagation Perspective" in
Festschrift book in honor of Judea Pearl, 2010

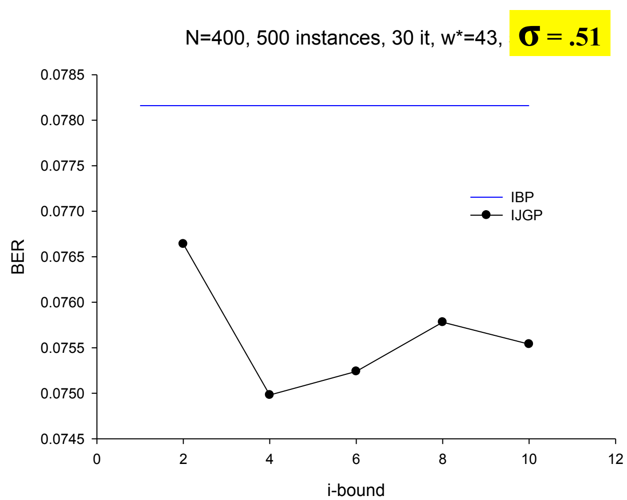
Heuristics, Probability and Causality, A Tribute to Judea Pearl
Editors Rina Dechter, Hector Geffner and Joseph Y. Halpern
http://bayes.cs.ucla.edu/TRIBUTE/pearl-tribute2010.htm

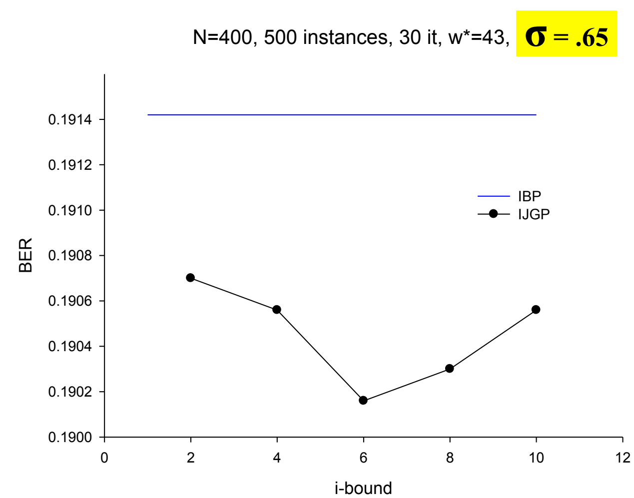
http://www.ics.uci.edu/~dechter/publications.html

Coding Networks – Bit Error Rate

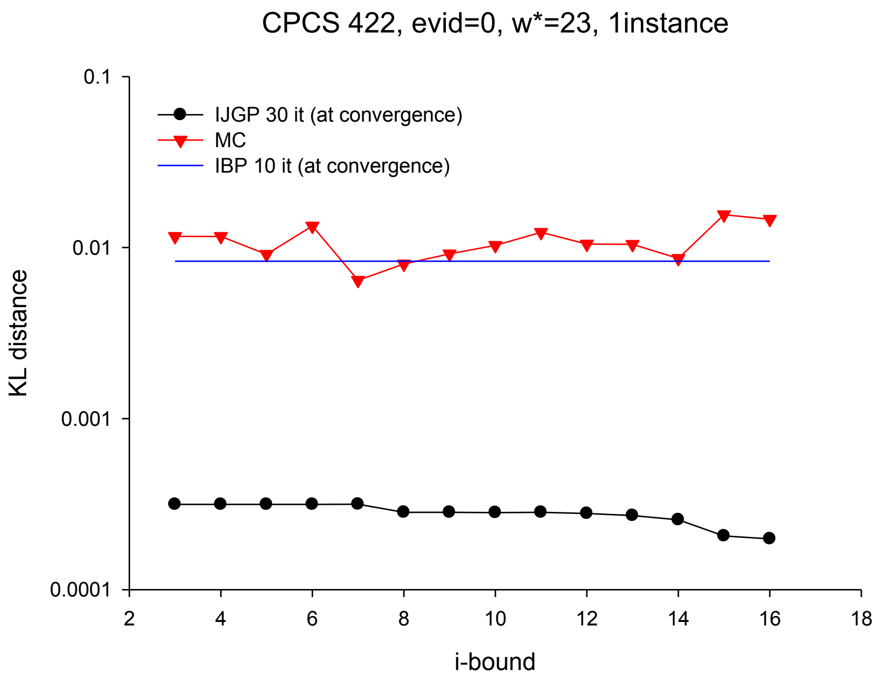


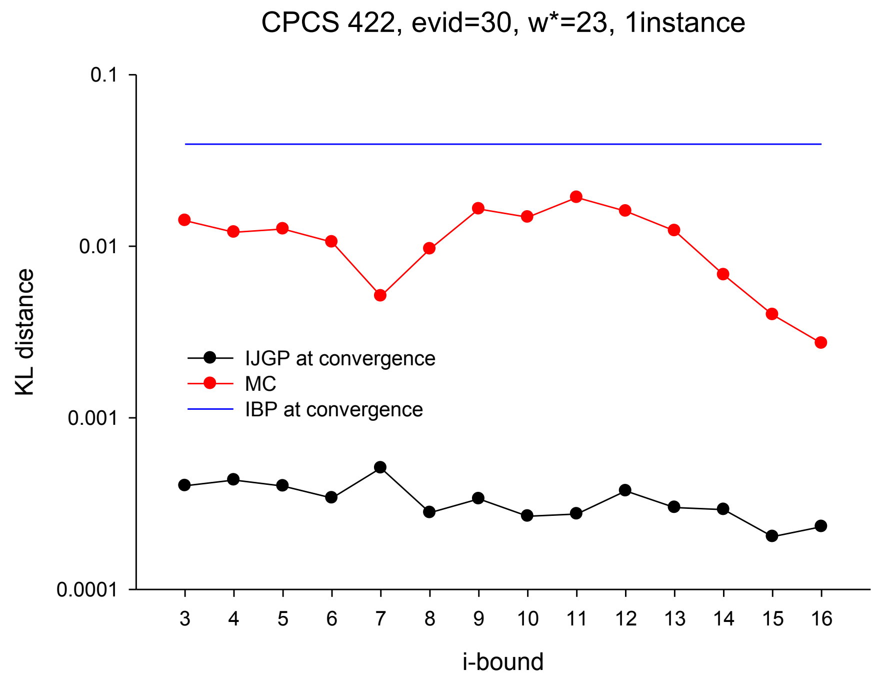






CPCS 422 – KL Distance

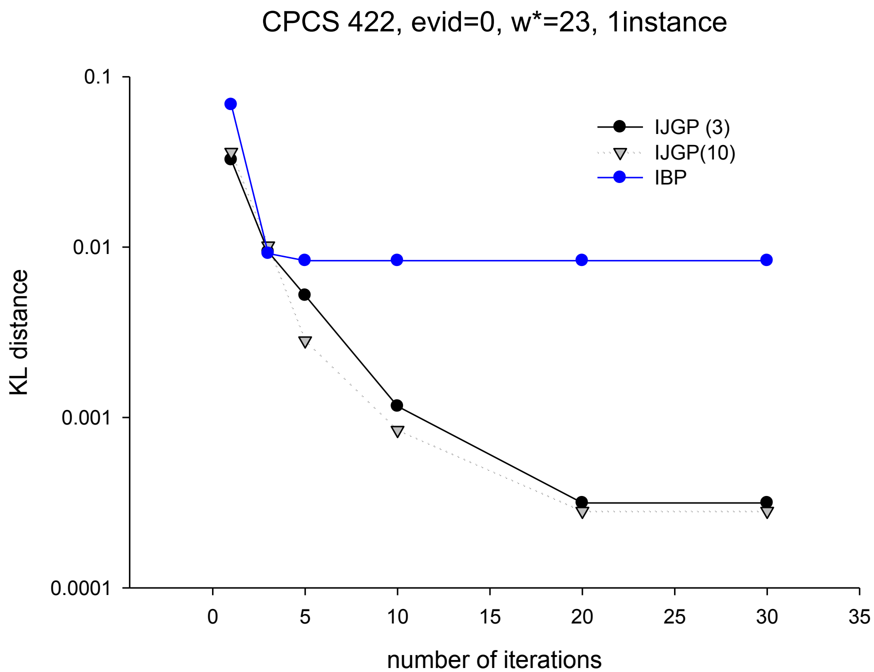




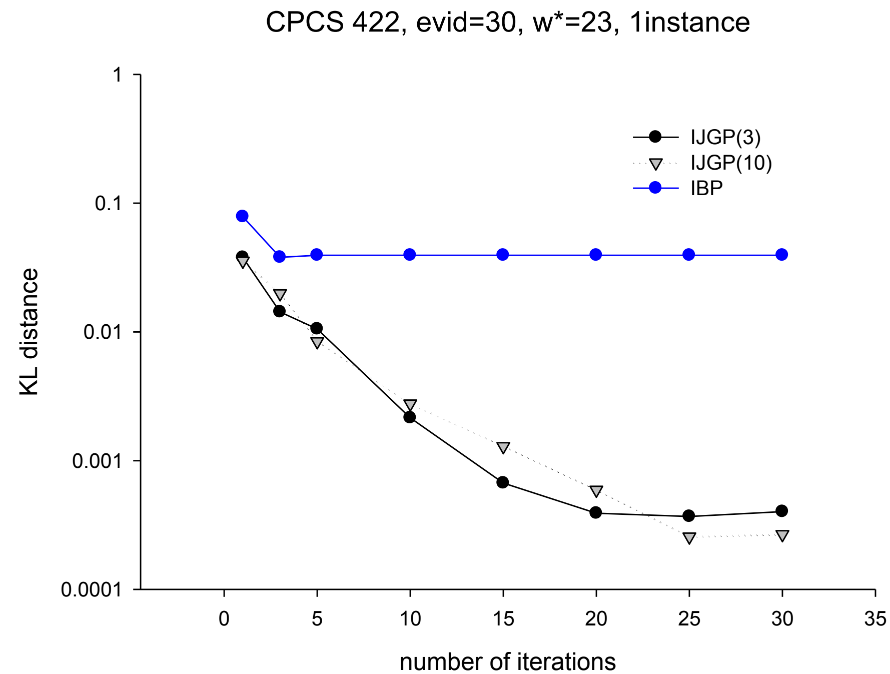
evidence=0

evidence=30

CPCS 422 – KL vs. Iterations



evidence=0

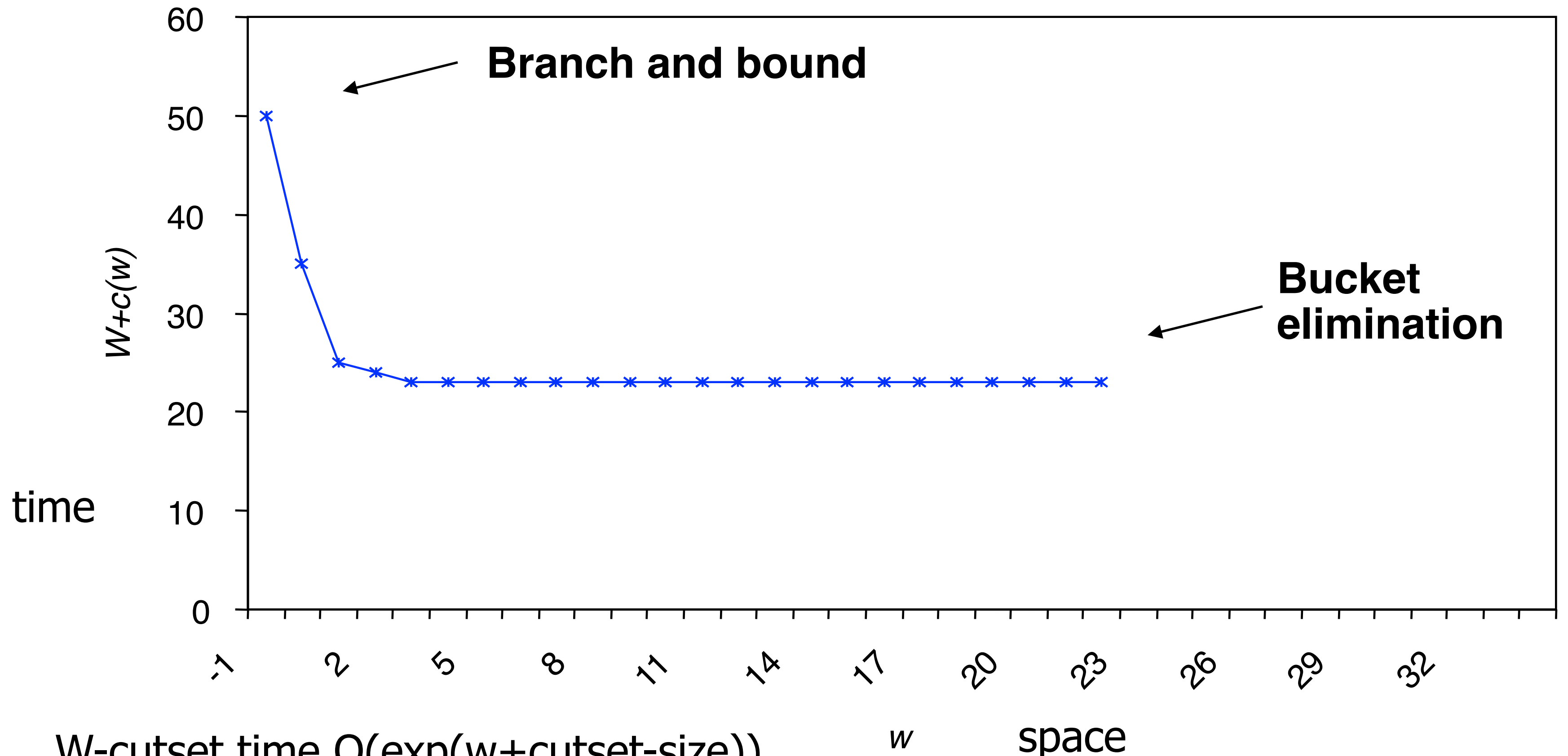


evidence=30

Time vs Space for w-cutset

(Dechter and El-Fatah, 2000)
(Larrosa and Dechter, 2001)
(Rish and Dechter 2000)

- **Random Graphs (50 nodes, 200 edges, average degree 8, $w^* \approx 23$)**



W-cutset time O(exp(w+cutset-size))
Space O(exp(w))