On the Power of Belief Propagation: A Constraint Propagation Perspective

Rina Dechter

Bozhena Bidyuk

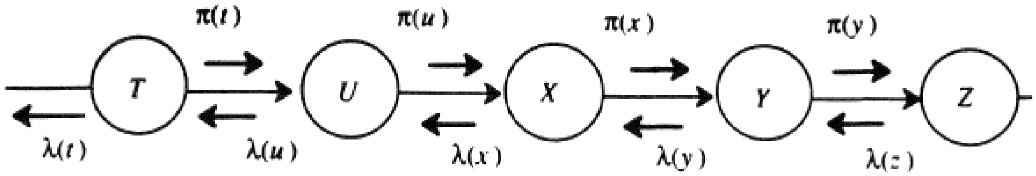
Robert Mateescu

Emma Rollon

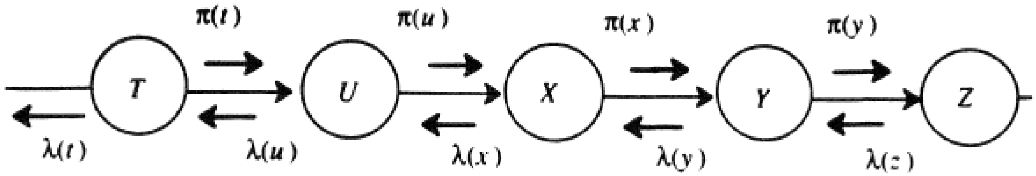
UCIrvine

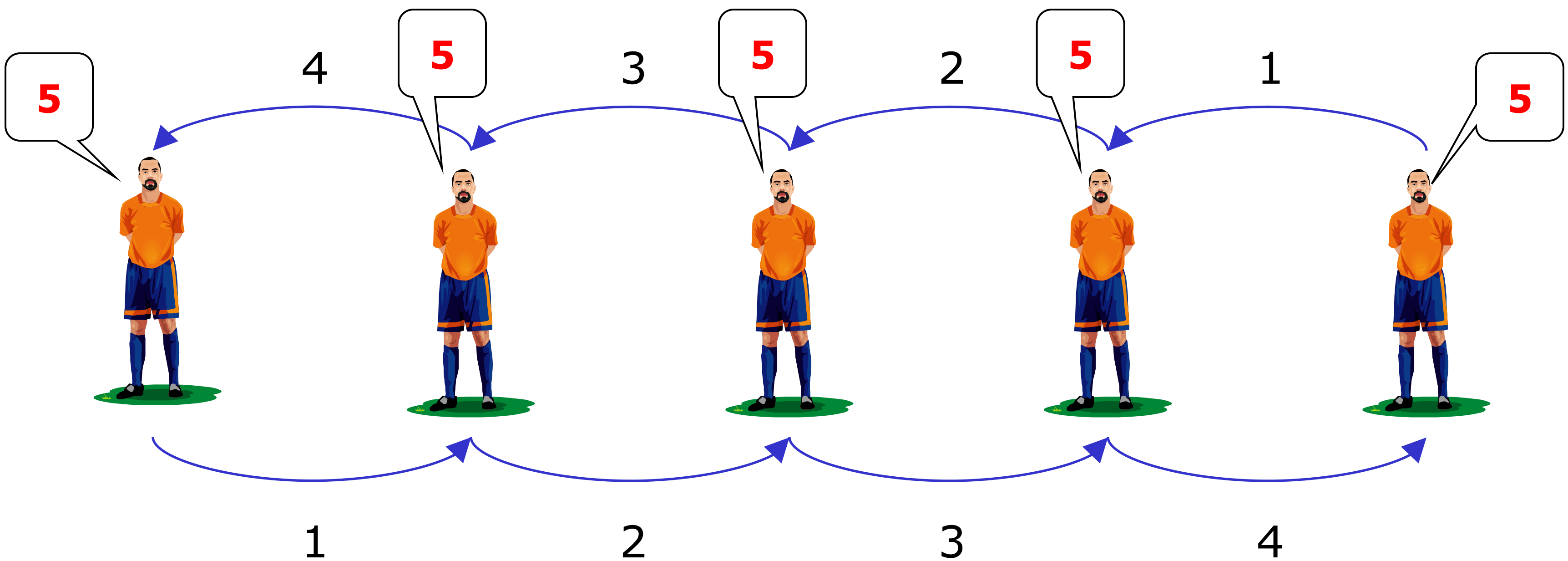
University of California, Irvine

Distributed Belief Propagation



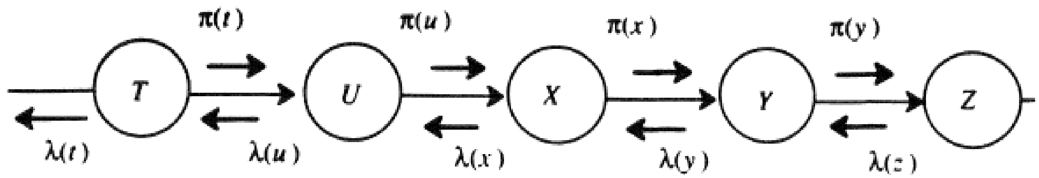
Distributed Belief Propagation



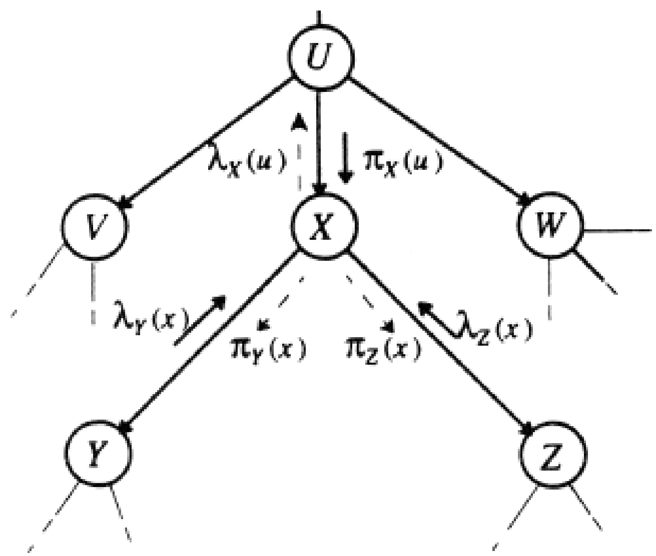


How many people?

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 in Polytrees

EVIDENCE DECOMPOSITION

$e^-_{XY_j}$ stands for evidence contained in the subnetwork on the *head* side of the link $X \rightarrow Y_j$,

$e^+_{U_iX}$ stands for evidence contained in the subnetwork on the *tail* side of the link $U_i \rightarrow X$.

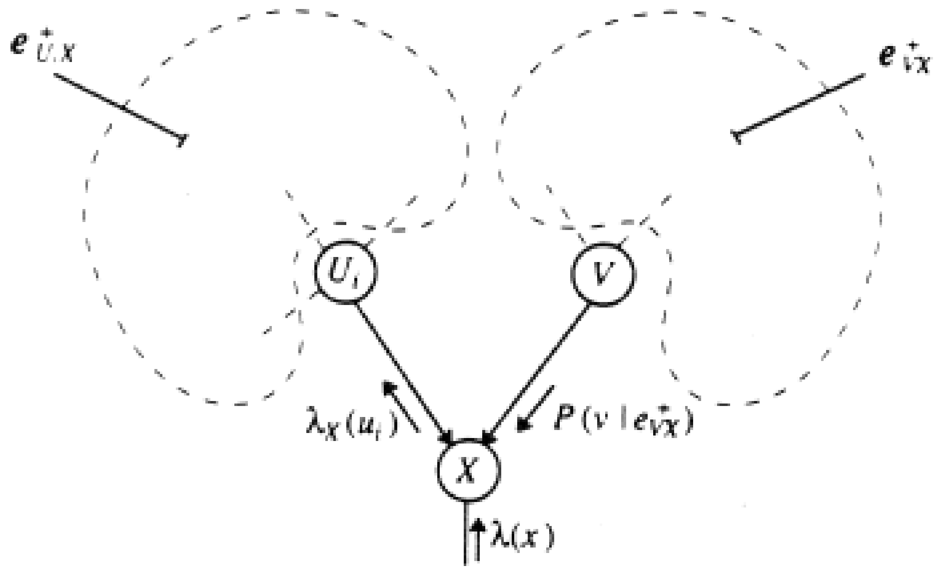
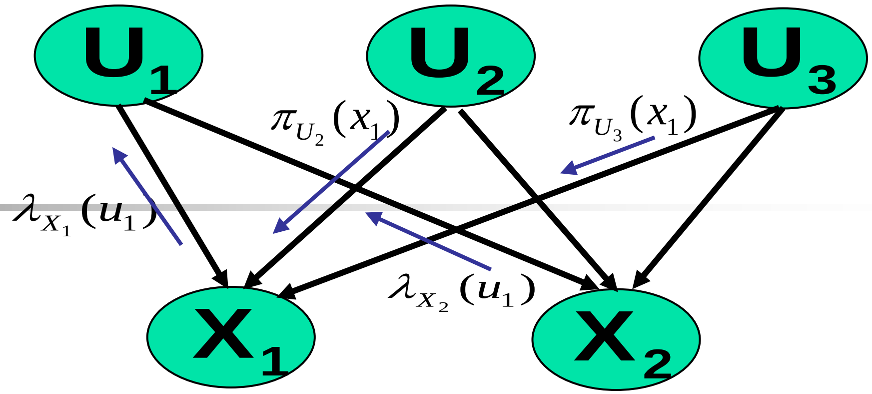


Figure 4.19. Variables, messages, and evidence sets used in the derivation of $\lambda_X(u_i)$.

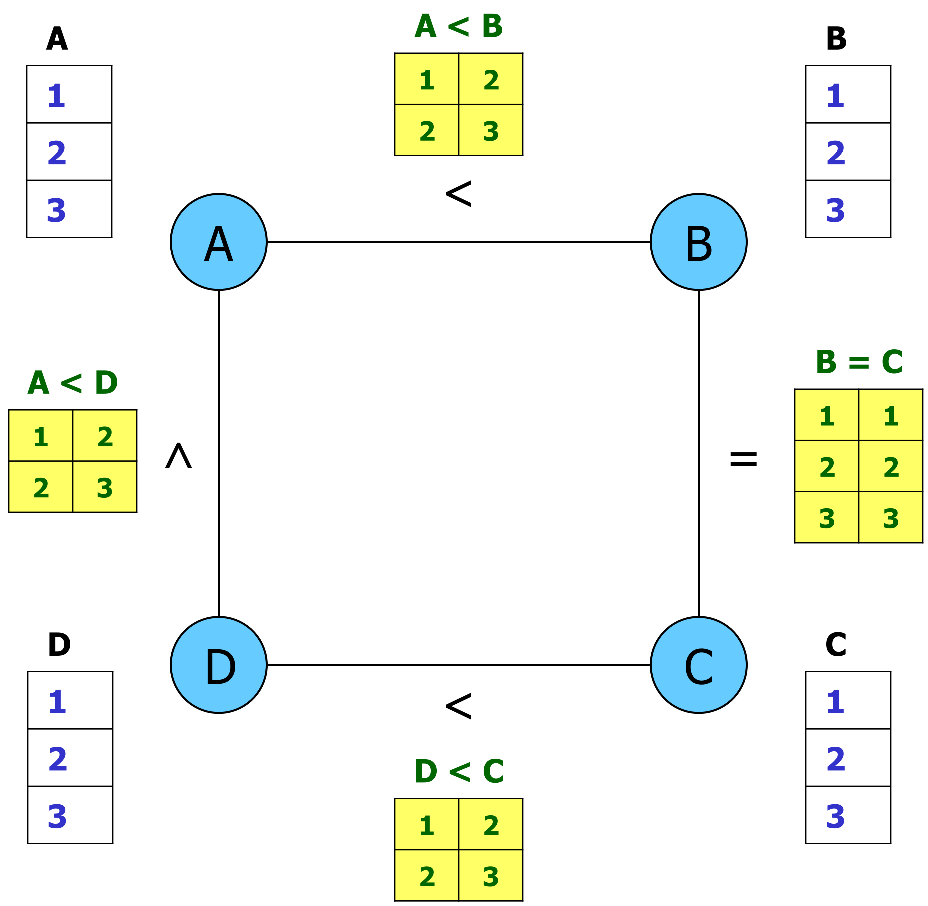
BP on Loopy Graphs



- Pearl (1988): use of BP to loopy networks
- McEliece, et. Al 1988: IBP's success on coding networks
- Lots of research into convergence … and accuracy (?), but:
  - Why IBP works well for coding networks
  - Can we characterize other good problem classes
  - Can we have any guarantees on accuracy (even if converges)

Arc-consistency

- Sound
- Incomplete
- Always converges (polynomial)

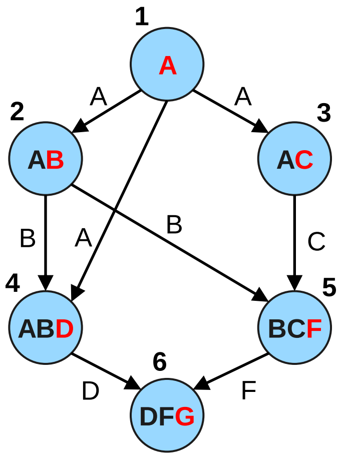


Flattening the Bayesian Network

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

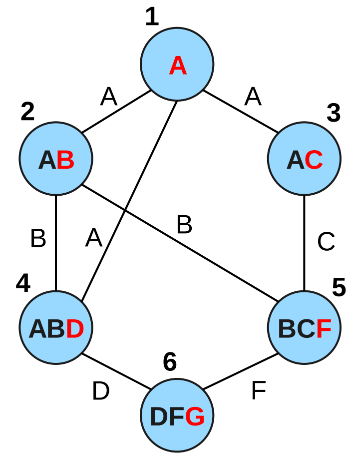
D	F	G	P(G|D,F)
1	2	3	1
2	1	3	1
…	…	…	0

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

Belief Zero Propagation = Arc-Consistency

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

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

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

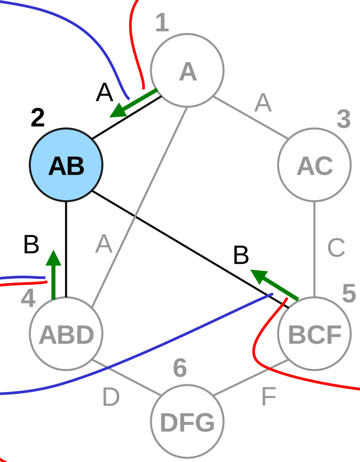
A	$h_1^2(A)$
1	>0
2	>0
3	>0
…	0

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

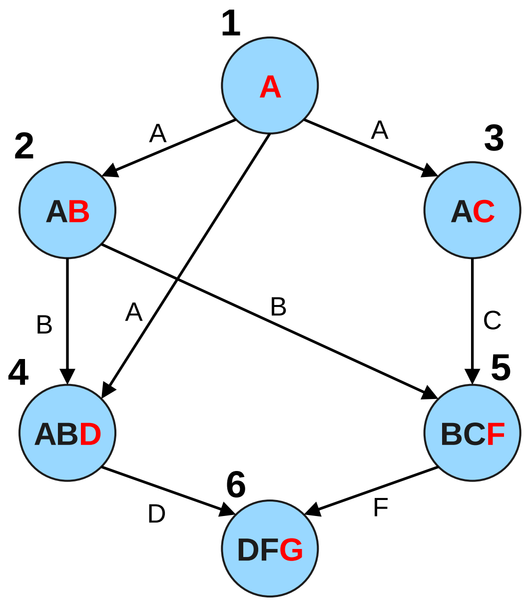
Flat Network - Example

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

IBP Example – Iteration 1

R_1

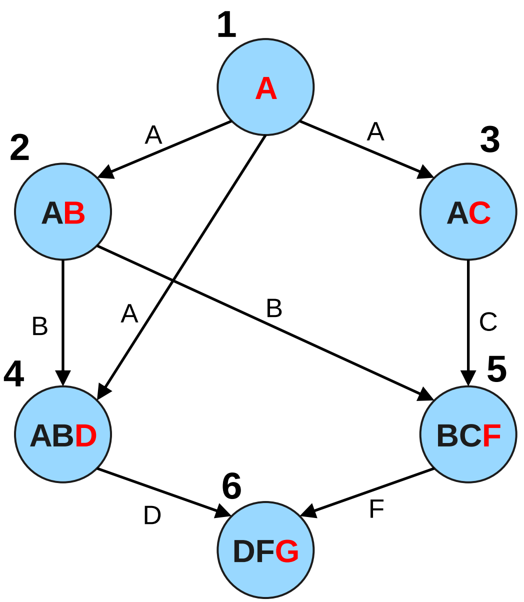
A	P(A)
1	>0
3	>0
…	0

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

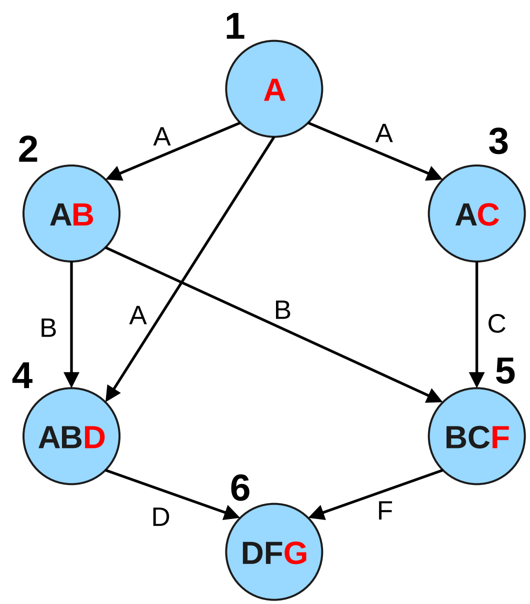
IBP Example – Iteration 2

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

IBP Example – Iteration 3

R_1

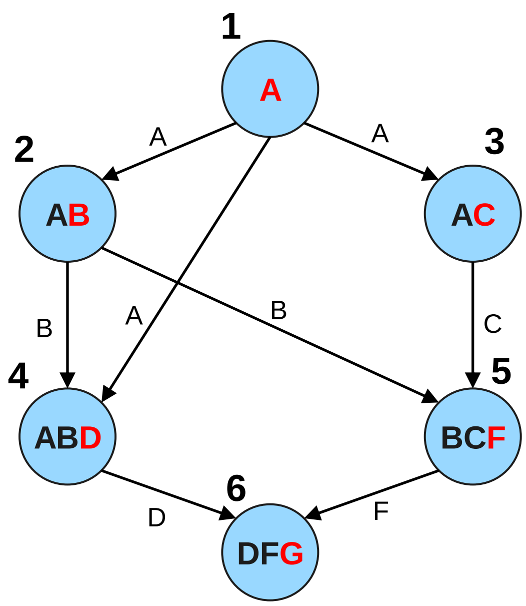
A	P(A)
1	>0
3	>0
…	0

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

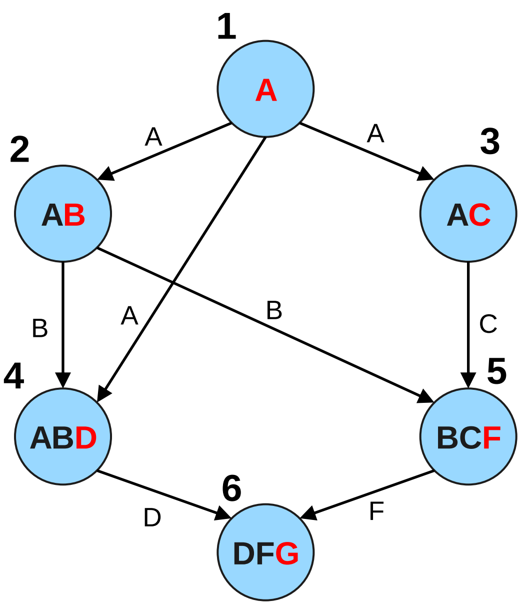
IBP Example – Iteration 4

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

IBP Example – Iteration 5

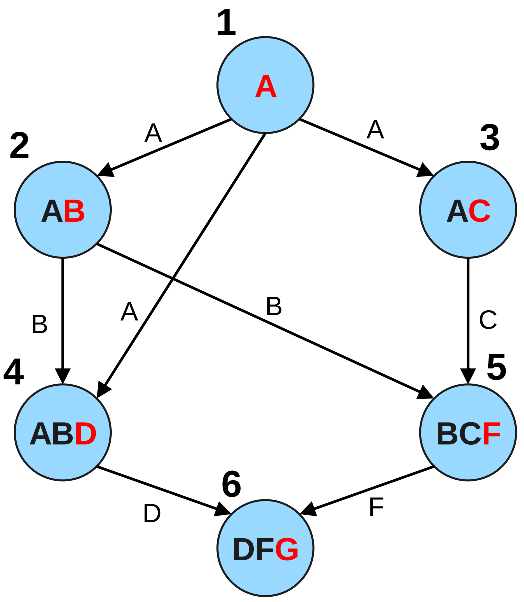
R_1

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

IBP – Inference Power for Zero Beliefs

- **Theorem:**
  Iterative BP performs arc-consistency on the flat network.
- **Soundness:**
  - Inference of zero beliefs by IBP converges
  - All the inferred zero beliefs are correct
- **Incompleteness:**
  - Iterative BP is as weak and as strong as arc-consistency
- **Continuity Hypothesis:** IBP is sound for zero -→ IBP is accurate for extreme beliefs? Tested empirically

Experimental Results

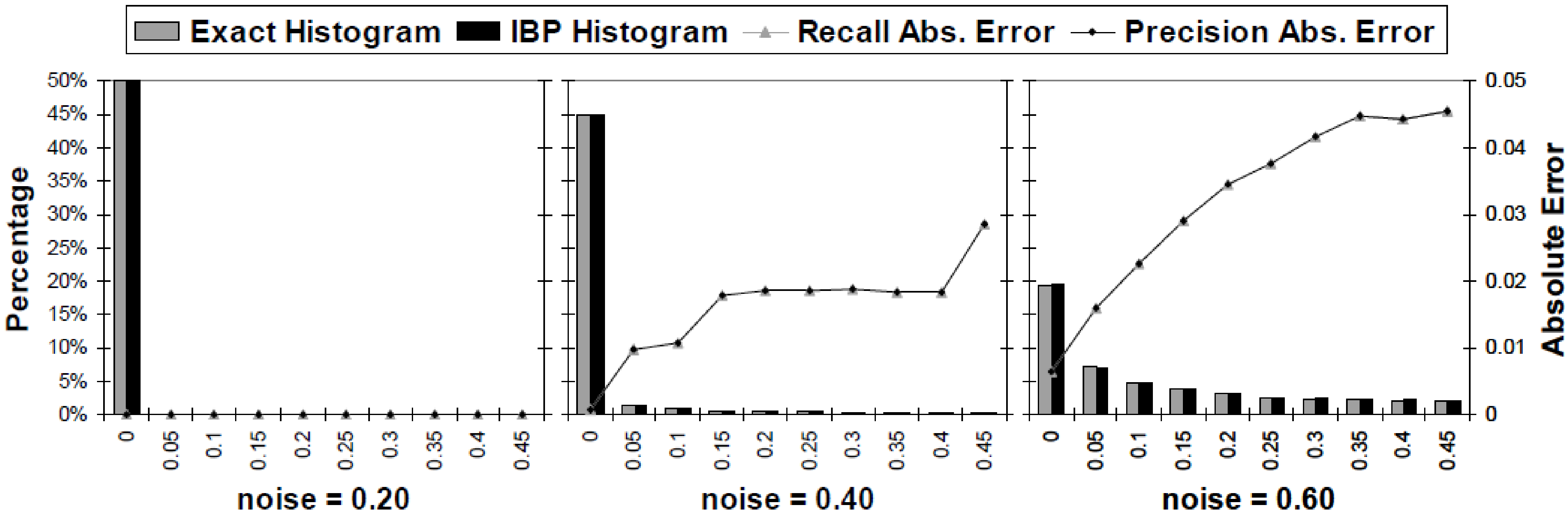
We investigated empirically if the results for zero beliefs extend to ε-small beliefs ($\varepsilon > 0$)

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

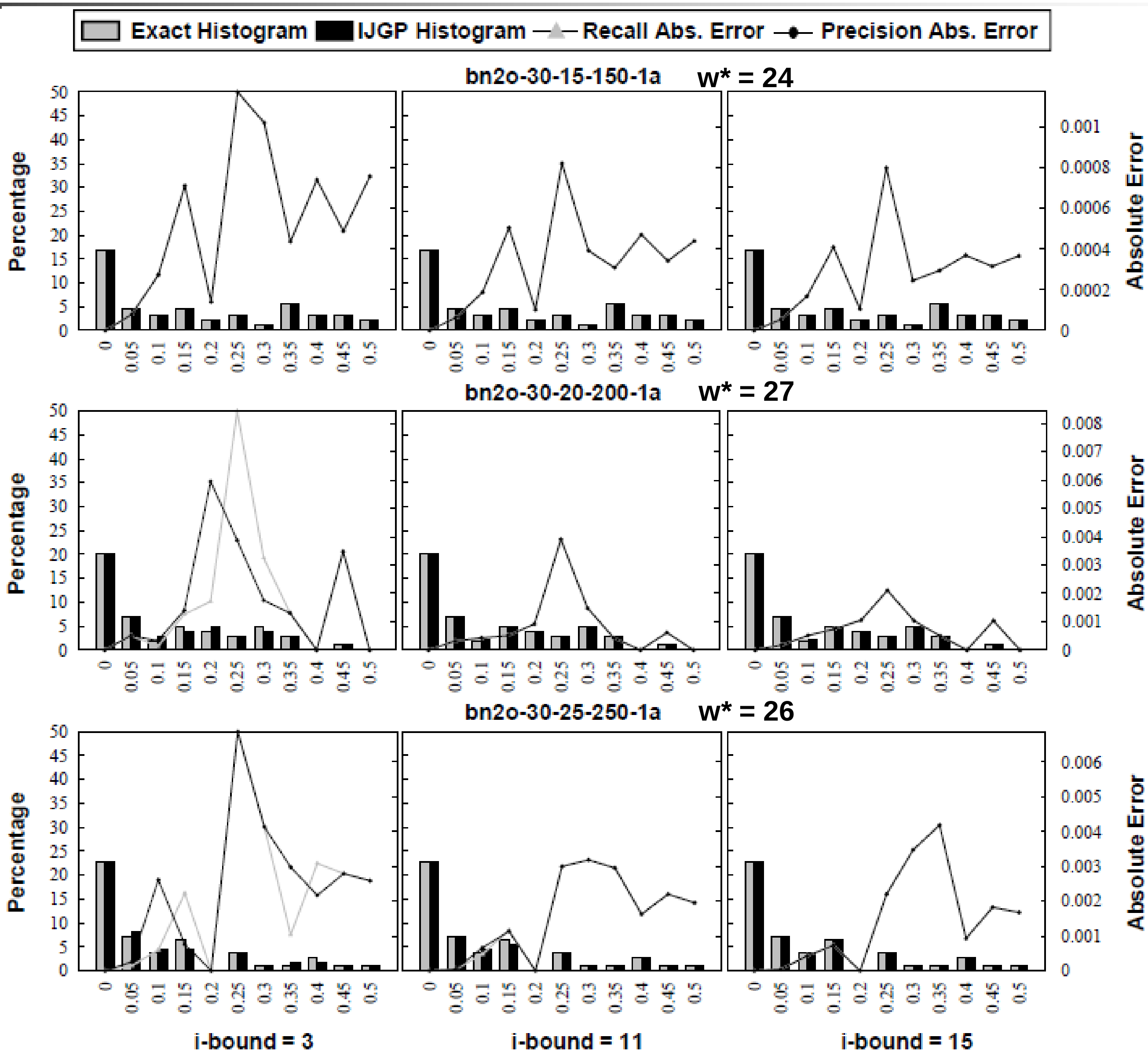
* Instances from the UAI08 competition

Networks with Determinism: Coding

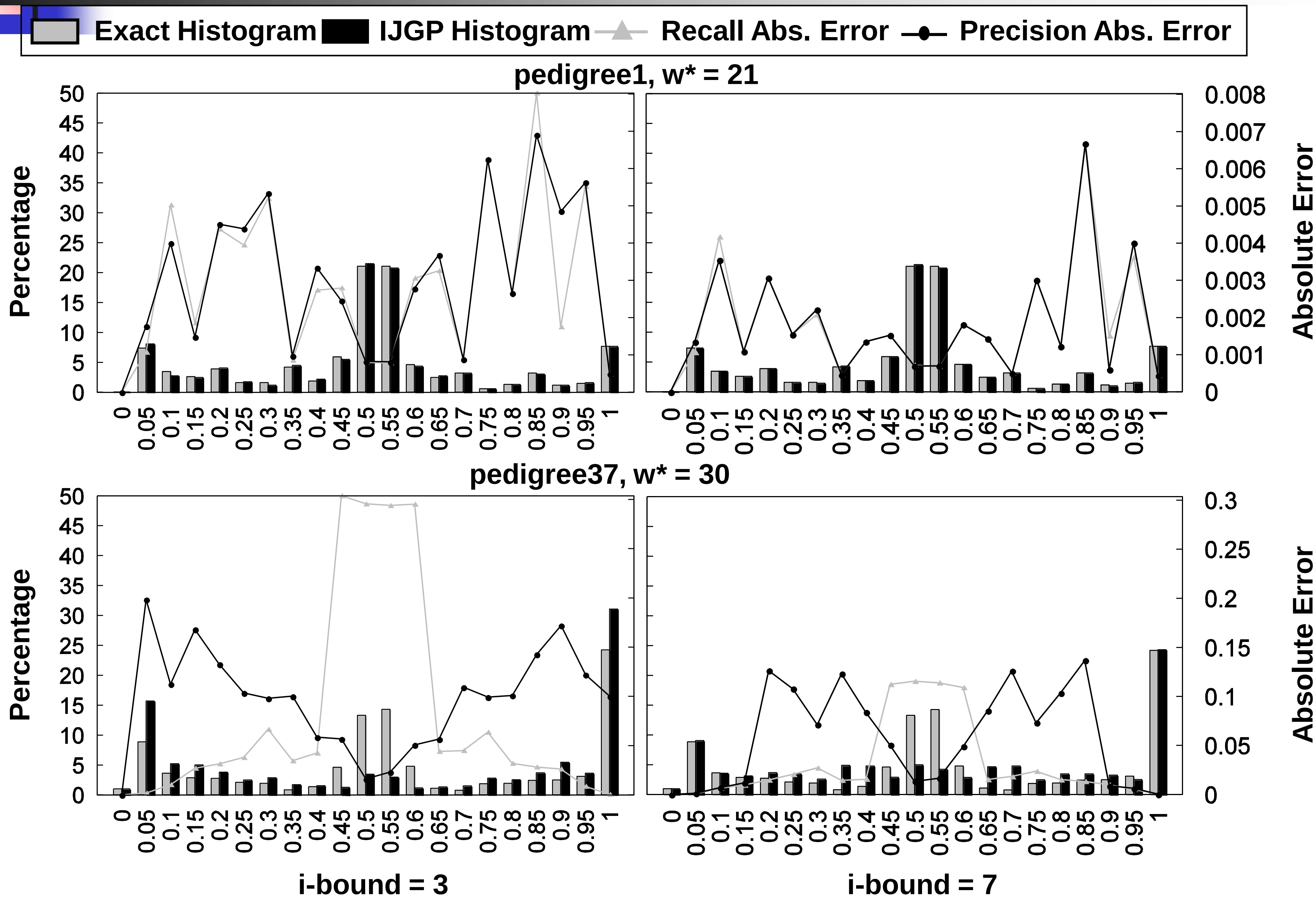


N=200, 1000 instances, w*=15

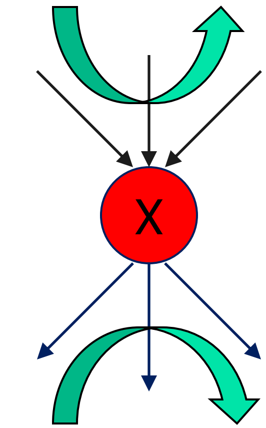
Nets w/o Determinism: bn2o

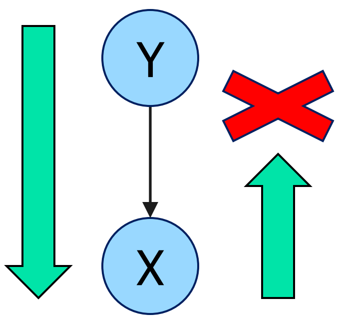


Nets with Determinism: Linkage



The Cutset Phenomena & irrelevant nodes

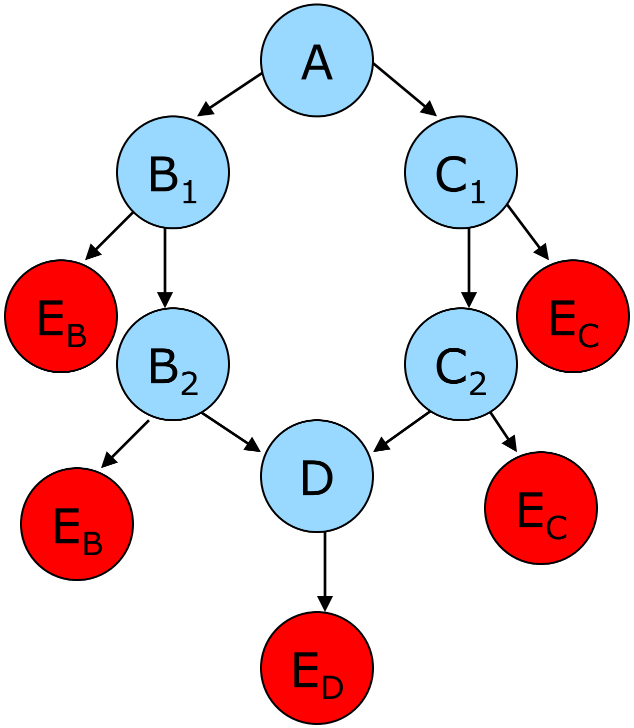


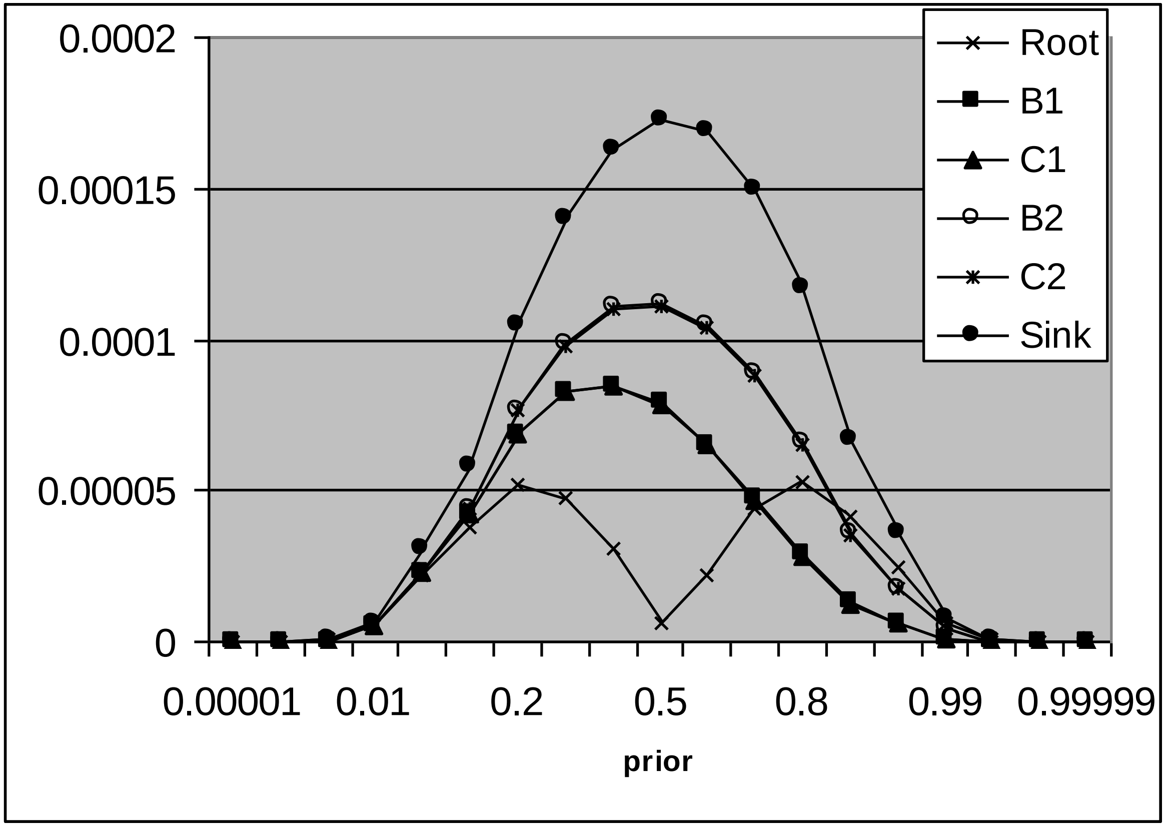


- **Observed variables** break the flow of inference
  - **IBP is exact when evidence variables form a cycle-cutset**
- **Unobserved variables** without observed descendents send zero-information to the parent variables – it is irrelevant
  - **In a network without evidence, IBP converges in one iteration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

Nodes with extreme support

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





Average Error vs. Priors

Conclusion: For Networks with Determinism

- IBP 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.