RN, Chapter 14

Bayesian Belief Networks

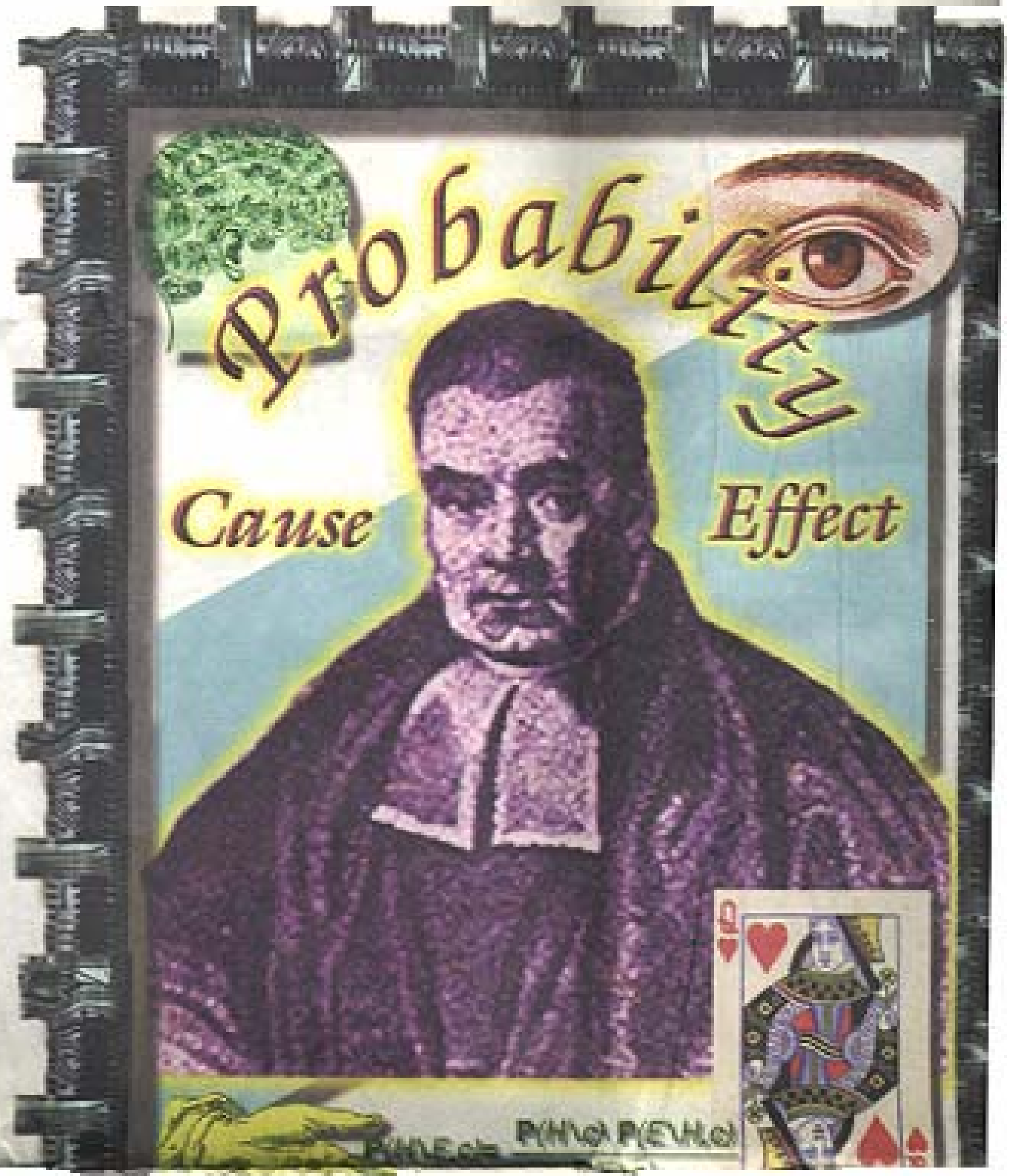
Decision Theoretic Agents

- Introduction to Probability [Ch13]
- Belief networks [Ch14]
  - Introduction [Ch14.1-14.2]
  - Bayesian Net Inference [Ch14.4] (Bucket Elimination)
- Dynamic Belief Networks [Ch15]
- Single Decision [Ch16]
- Sequential Decisions [Ch17]

BUSINESS

MONDAY TECHNOLOGY SPECIAL

The future of software may lie in the obscure theories of an 18th century cleric named Thomas Bayes.



Motivation

- Gates says *[LATimes, 28/Oct/96]:*

  Microsoft's competitive advantages is its expertise in "Bayesian networks"

- *Current Products*
  - *Microsoft Pregnancy and Child Care (MSN)*
  - *Answer Wizard (Office, …)*
  - *Print Troubleshooter*

    Excel Workbook Troubleshooter

    Office 95 Setup Media Troubleshooter

    Windows NT 4.0 Video Troubleshooter

    Word Mail Merge Troubleshooter

Motivation (II)

- US Army: **SAIP** (Battalion Detection from SAR, IR… GulfWar)
- NASA: **Vista** (DSS for Space Shuttle)
- GE: **Gems** (real-time monitor for utility generators)
- Intel: (infer possible processing problems from end-of-line tests on semiconductor chips)
- KIC:
  - medical: sleep disorders, pathology, trauma care, hand and wrist evaluations, dermatology, home-based health evaluations
  - DSS for capital equipment: locomotives, gas-turbine engines, office equipment

Motivation (III)

- Lymph-node pathology diagnosis
- Manufacturing control
- Software diagnosis
- Information retrieval
- *Types of tasks*
  - *Classification/Regression*
  - *Sensor Fusion*
  - *Prediction/Forecasting*
  - *Modeling*

Motivation

- Challenge: To decide on proper action
  - Which treatment, given symptoms?
  - Where to move?
  - Where to search for info?
  - . . .
- Need to know dependencies in world
  - between symptom and disease
  - between symptom_1 and symptom_2
  - between disease_1 and disease_2
  - . . .
- Q: Full joint?
  - A: Too big ($\geq 2^n$)
  - Too slow (inference requires adding 2^k . . .)
- Better:
  - Encode dependencies
  - Encode only *relevant* dependencies

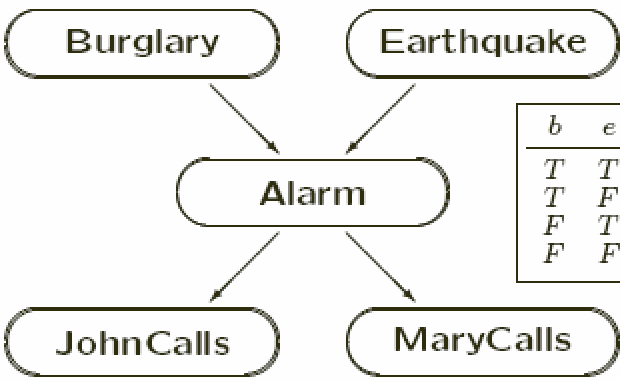
Components of a Bayesian Net

Directed Acyclic Graph:

$$\mathcal{BN} = \left\{ \begin{array}{lll} \mathcal{N} & \text{Nodes} & \equiv \text{Variables} \\ \mathcal{A} & \text{Arcs} & \equiv \text{Dependencies} \\ \mathcal{C} & \text{CPTables} & \equiv \text{“weights”} \end{array} \right\}$$

$P(B)$
0.001

$P(E)$
0.002



b	e	$P(A \mid B=b, E=e)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

a	$P(J \mid A=a)$
T	0.90
F	0.05

a	$P(M \mid A=a)$
T	0.70
F	0.01

- **Nodes**: one for each random variable
- **Arcs**: one for each direct influence between two random variables
- **CPT**: each node stores a conditional probability table

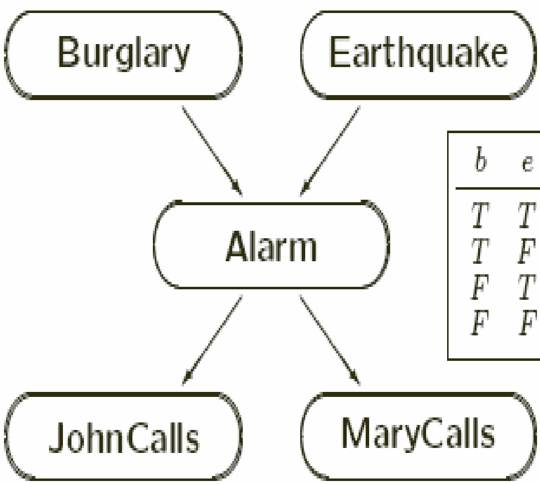
  P(Node | Parents(Node))

  to quantify effects of “parents" on child

Causes, and Bayesian Net

$P(B)$
0.001

$P(E)$
0.002



b	e	$P(A \mid B=b, E=e)$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

a	$P(J \mid A=a)$
T	0.90
F	0.05

a	$P(M \mid A=a)$
T	0.70
F	0.01

- What "causes" Alarm?
  A: Burglary, Earthquake
- What "causes" JohnCall?
  A: Alarm
  N.b., NOT Burglary, ...
- Why not Alarm $\Rightarrow$ MaryCalls?

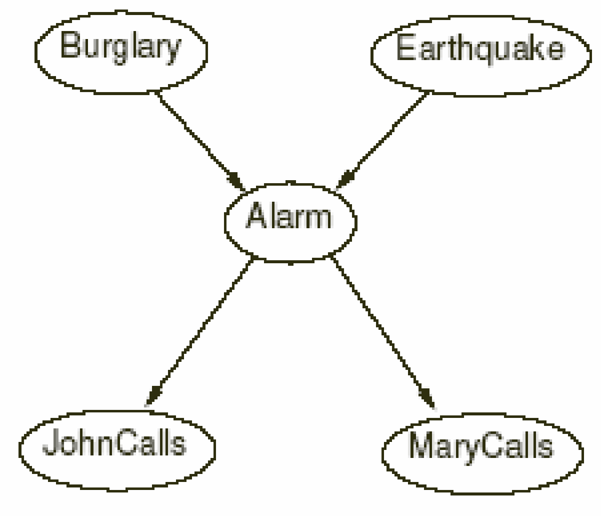
  (CPTable =

Alarm	$P(\text{MC} \mid \text{A})$
T	1.0
F	0.0

)

A: Mary not always home
... phone may be broken
...

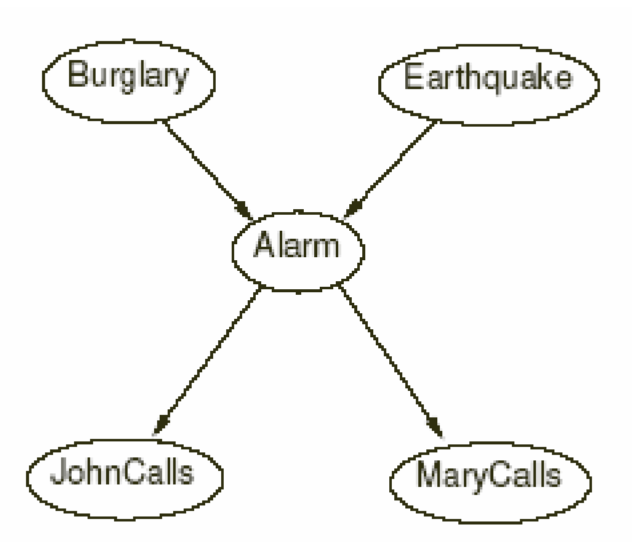
Independence in a Belief Net



- Burglary, Earthquake independent
  - $B \perp E$
- Given Alarm, JohnCalls and MaryCalls independent
  - $J \perp M \mid A$

  - JohnCalls is correlated with MaryCalls $\neg(J \perp M)$ as suggest Alarm
  - But given Alarm, JohnCalls gives no NEW evidence wrt MaryCalls

Conditional Independence



Local Markov Assumption:
A variable X is independent of its non-descendants given its parents
$(X_i \perp NonDescendants_{X_i} \mid Pa_{X_i})$

- $B \perp E \mid \{\}$ $(B \perp E)$
- $M \perp \{B,E,J\} \mid A$
- Given graph G,
  $I_{LM}(G) = \{ (X_i \perp NonDescendants_{X_i} \mid Pa_{X_i}) \}$

Factoid: Chain Rule

- $P(A,B,C) = P(A \mid B,C)\, P(B,C)$

  $= P(A \mid B,C)\, P(B|C)\, P(C)$

- In general:

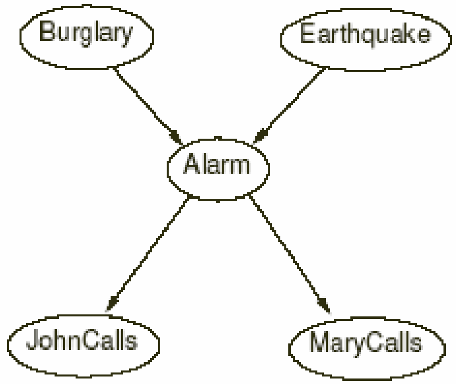
  $P(X_1, X_2, \ldots, X_m) =$

  $P(X_1 \mid X_2, \ldots, X_m)\, P(X_2, \ldots, X_m) =$

  $P(X_1 \mid X_2, \ldots, X_m)\, P(X_2 \mid X_3, \ldots, X_m)\, P(X_3, \ldots, X_m) =$

  $\prod_i P(X_i \mid X_{i+1}, \ldots, X_m)$

Joint Distribution



P(+j, +m, +a, -b, -e)

J ⊥ {M,B,E} | A

= ~~P(+j | +m, +a, -b, -e)~~ → P(+j | +a)

M ⊥ {B,E} | A

~~P(+m | +a, -b, -e)~~ → P(+m | +a)

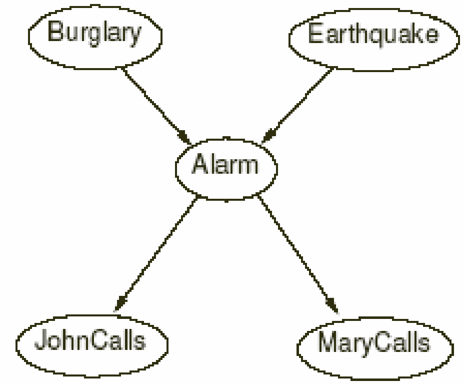
~~P(+a| -b, -e)~~ → P(+a | -b,-e)

B ⊥ E

~~P(-b | -e)~~ → P(-b)

~~P(-e)~~ → P(-e)

Joint Distribution



$P(+j, +m, +a, -b, -e)$

$= P(+j \mid +a)$

$P(+m \mid +a)$

$P(+a \mid -b, -e)$

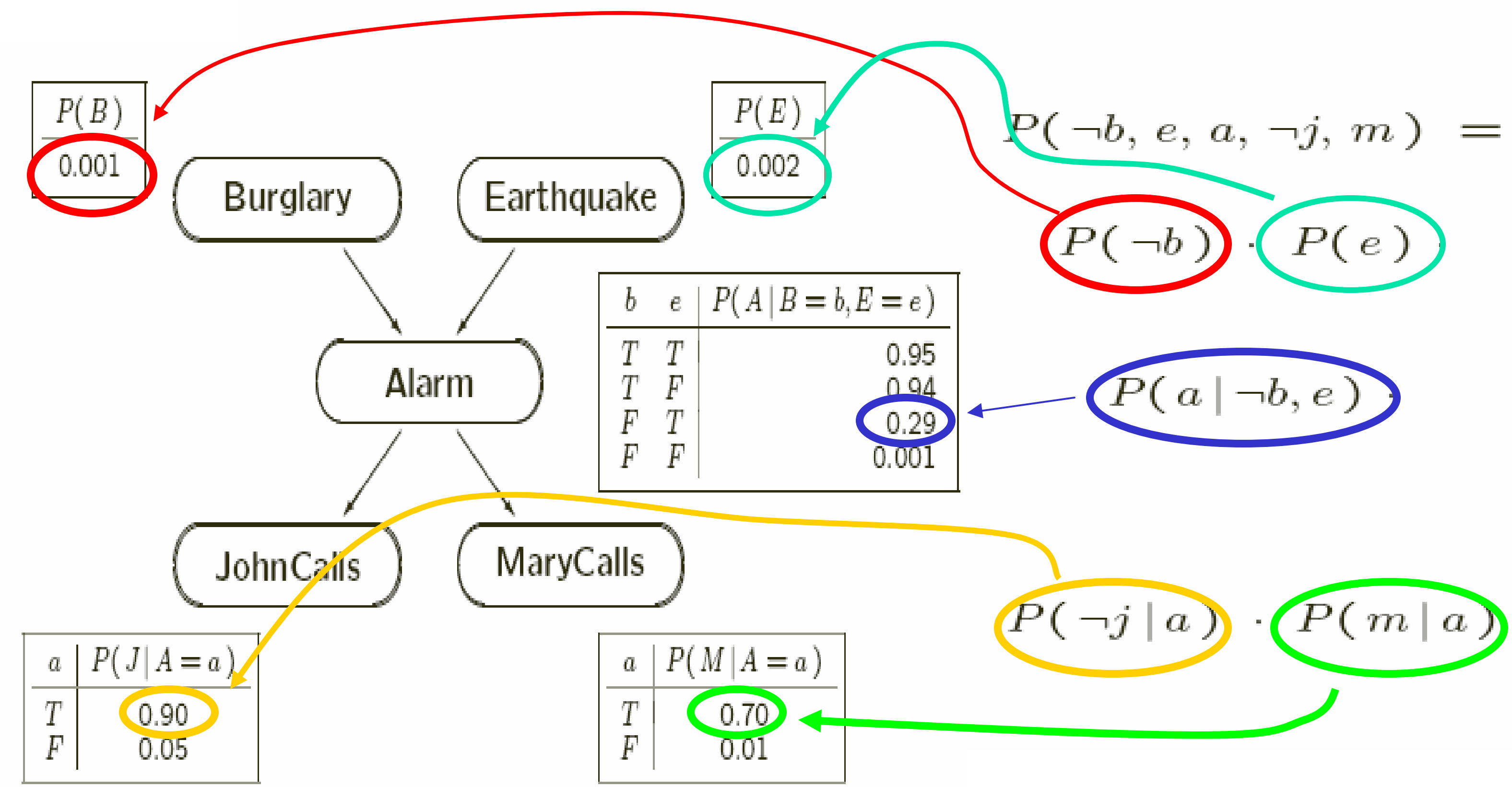
$P(-b)$

$P(-e)$

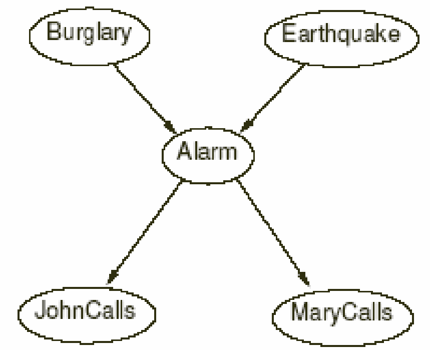
Recovering Joint

$$
\begin{aligned}
P(\neg b, e, a, \neg j, m) = \; & P(\neg b)\; P(e \mid \neg b)\; P(a \mid e, \neg b)\; P(\neg j \mid a, e, \neg b)\; P(m \mid \neg j, a, e, \neg b) \\
& P(\neg b)\; P(e)\; P(a \mid e, \neg b)\; P(\neg j \mid a)\; P(m \mid a) \\
& 0.99 \times 0.02 \times 0.29 \times 0.1 \times 0.70
\end{aligned}
$$

Node independent of predecessors, given parents

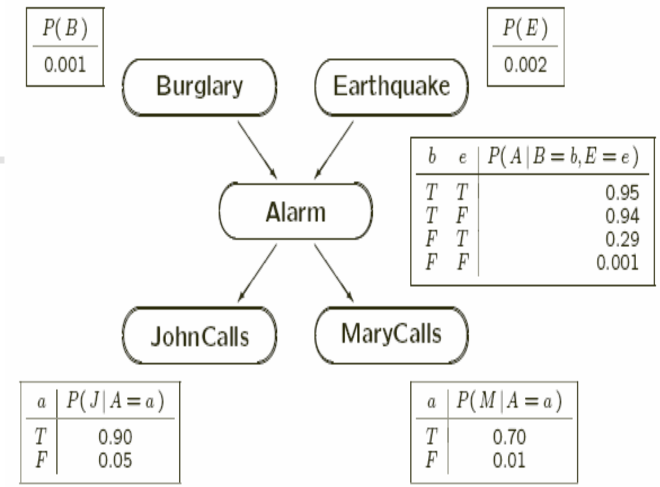


Meaning of Belief Net



- A BN represents
  - joint distribution
  - condition independence statements
- $P(J, M, A, \neg B, \neg E)$
  $= P(\neg B) P(\neg E) P(A|\neg B, \neg E) P(J | A) P(M |A)$
  $= 0.999 \times 0.998 \times 0.001 \times 0.90 \times 0.70 = 0.00062$
- In gen'l, $P(X_1,X_2, . . . ,X_m) = \prod_i P(X_i |X_{i+1}, . . . ,X_m)$
- Independence means
  $P(X_i |X_{i+1}, . . . ,X_m) = P(X_i | Parents(X_i))$
  Node independent of predecessors, given parents
- So... $P(X_1,X_2, . . . ,X_m) = \prod_i P(X_i | Parents(X_i))$

Comments



- BN used 10 entries
  ... can recover full joint (2^5 entries)
  (Given structure,
  other $2^5 - 10$ entries are REDUNDANT)

$\Rightarrow$ Can compute
P(Burglary | JohnCalls, ¬MaryCalls) :
Get joint, then marginalize, conditionalize, ...
∃ better ways. . .

- Note: Given structure, ANY CPT is consistent.
  ∄ redundancies in BN. . .

Conditional Independence

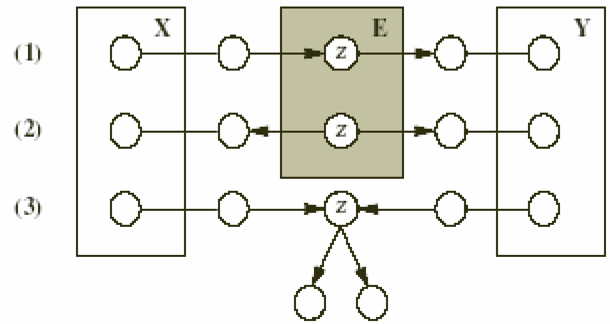
- Node X is independent of its non-descendants given assignment to immediate parents parents(X)

- **General** question: "X ⊥ Y | **E**"
  - Are nodes **X** independent of nodes **Y**, given assignments to (evidence) nodes **E**?

- **Answer**: If every undirected path from X to Y is d-separated by **E**, then X ⊥ Y | **E**

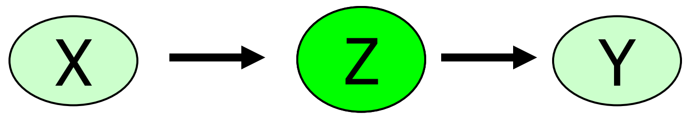
- *d-separated* if every path from X to Y is blocked by **E**

  . . . if ∃ node Z on path s.t.
  1. Z ∈ **E**, and Z has 1 out-link (on path)
  2. Z ∈ **E**, and Z has 2 out-link, *or*
  3. Z has 2 in-links, Z ∉ **E**, no child of Z in **E**



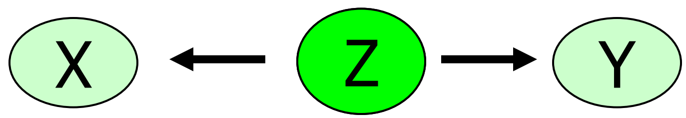
d-separation Conditions

$\neg(X \perp Y)$



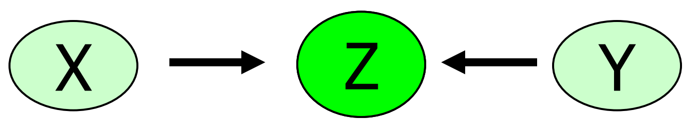
$X \perp Y \mid Z$

$\neg(X \perp Y)$



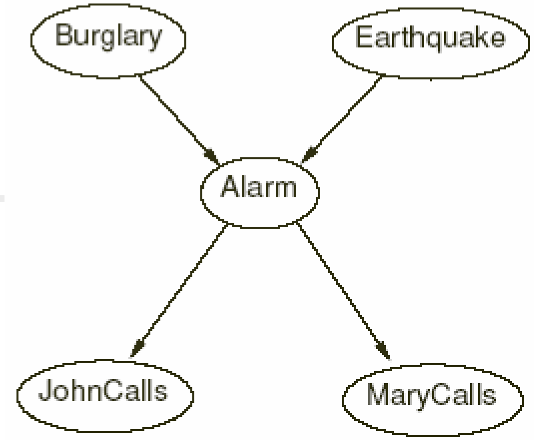
$X \perp Y \mid Z$

$X \perp Y$



$\neg(X \perp Y \mid Z)$

d-Separation



- Burglary and JohnCalls are conditionally independent given Alarm
- JohnCalls and MaryCalls are conditionally independent given Alarm
- Burglary and Earthquake are independent given no other information
- But. . .
  - Burglary and Earthquake are dependent given Alarm
  - Ie, Earthquake may "explain away" Alarm … decreasing prob of Burglary

"V"-Connections

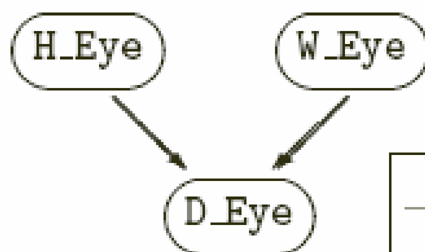
- What colour are my wife's eyes?

- Would it help to know MY eye color?

NO! H_Eye and W_Eye are independent!

- We have a DAUGHTER, who has BLUE eyes

Now do you want to know my eye-color?

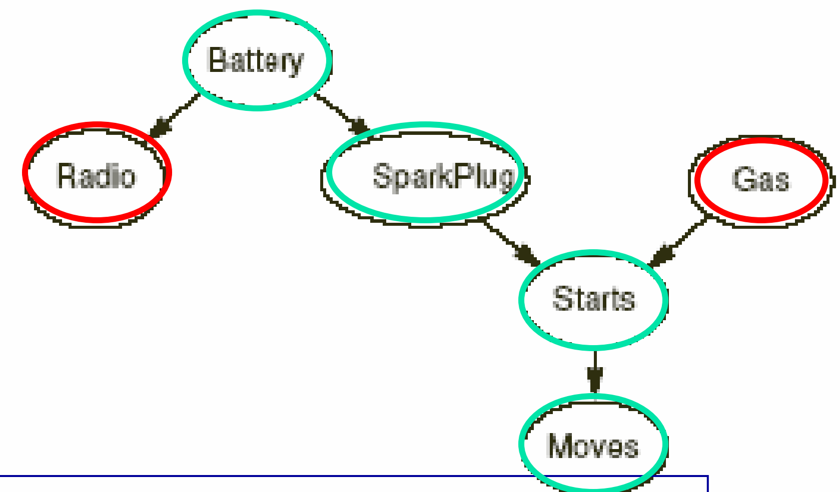


h	w	$P(D\_Eye = Blue \mid h, w)$
bl	bl	1.00
bl	br	0.50
br	bl	0.50
br	br	0.25

- H_Eye and W_Eye became dependent!

Example of *d*-separation, II

d-separated if
every path from X to Y is blocked by **E**



Is Radio *d*-separated from Gas given . . .

1. **E** = {} ?
   YES: $P(R \mid G) = P(R)$
   Starts $\notin$ **E**, and Starts has 2 in-links
2. **E** = Starts ?
   NO!! $P(R \mid G, S) \neq P(R \mid S)$
   Starts $\in$ **E**, and Starts has 2 in-links
3. **E** = Moves ?
   NO!! $P(R \mid G, M) \neq P(R \mid M)$
   Moves $\in$ **E**, Moves child-of Starts, and Starts has 2 in-links (on path)
4. **E** = SparkPlug ?
   YES: $P(R \mid G, Sp) = P(R \mid Sp)$
   SparkPlug $\in$ **E**, and SparkPlug has 1 out-link
5. **E** = Battery ?
   YES: $P(R \mid G, B) = P(R \mid B)$
   Battery $\in$ **E**, and Battery has 2 out-links

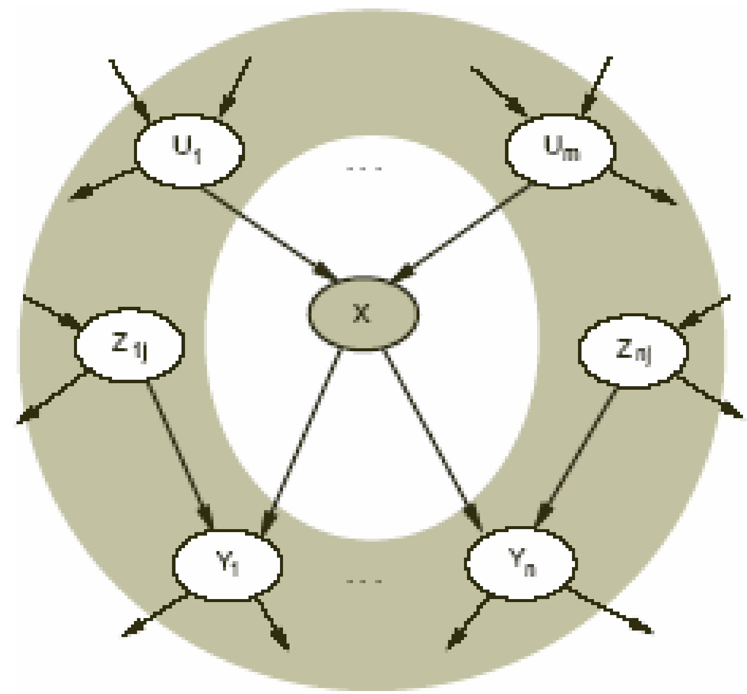
If car does not start

If car does not MOVE,
expect radio to NOT work.
Unless you see it is out of gas!

Markov Blanket

Each node is conditionally independent of all others given its *Markov blanket:*

- parents
- children
- children's parents



Simple Forms of CPTable

- In gen'l: CPTable is function mapping *values of parents* to *distribution over child*

$$f: \left[\prod_{U \in Parents(X)} Dom(U) \right] \times Dom(X) \mapsto [0.1]$$

(Actually, $f' : \prod_{U \in Parents(X)} Dom(U) \mapsto$ dist over X)

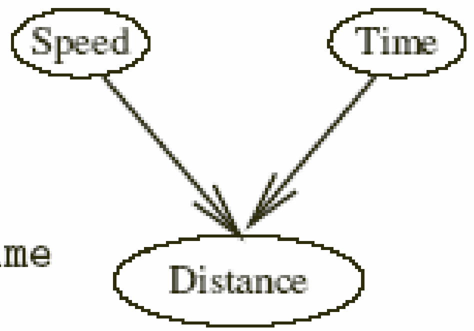
Cold	Flu	Malaria	P(Fever | C,F,M)	P(¬Fever | C,F,M)
F	F	F	0.0	1.0
F	F	T	0.9	0.1
F	T	F	0.8	0.2
F	T	T	0.98	0.02
T	F	F	0.4	0.6
T	F	T	0.94	0.06
T	T	F	0.88	0.12
T	T	T	0.988	0.012

f(+Col, -Flu, +Mal) = ⟨0.94 0.06⟩

- Standard: Include $\Pi_{U \in Parents(X)} |Dom(U)|$ rows, each with *|Dom(X)| - 1* entries
- But... can be structure within CPTable: Deterministic, Noisy-Or, Decision Tree, . . .

Deterministic Node

- Given value of parent(s), specify unique value for child (logical, functional)

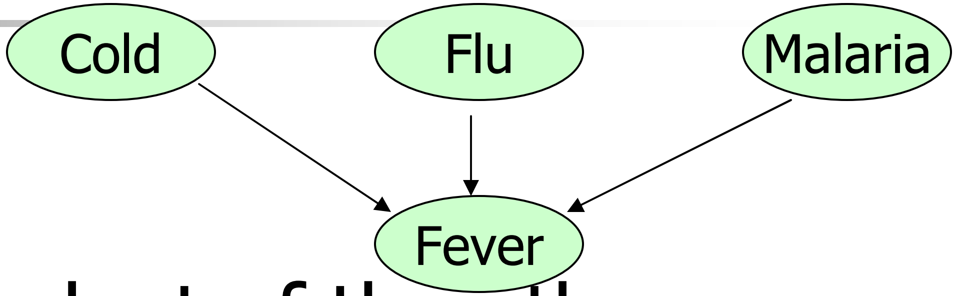


$$P(\texttt{Distance} \mid \texttt{Rate},\ \texttt{Time}) = \begin{cases} 1.0 & \text{if } \texttt{Distance} = \texttt{Rate} \cdot \texttt{Time} \\ 0.0 & \text{otherwise} \end{cases}$$

As if each row has just one 1, rest 0s:

Rate	Time	P(Dist=0 | R,T)	P(Dist=1 | R,T)	P(Dist=2 | R,T)
0	1	1.0	0.0	0.0
1	0	1.0	0.0	0.0
1	1	1.0	1.0	0.0
1	2	0.0	0.0	1.0
2	1	0.0	0.0	1.0
⋮		⋮		

Noisy-OR CPTable



- Each cause is independent of the others
- All possible causes are listed

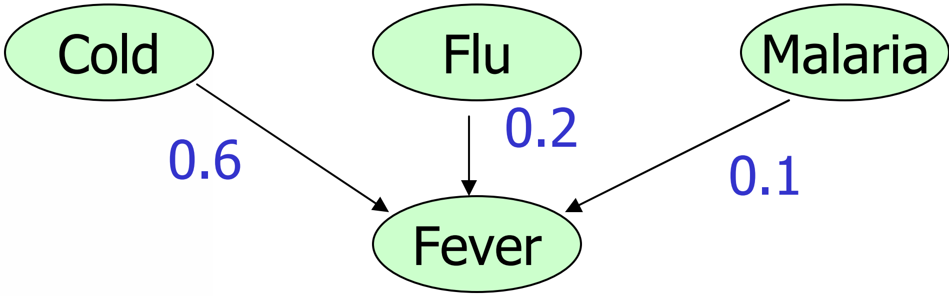
Want: No Fever if none of Cold, Flu or Malaria

$P(\neg Fev \mid \neg Col, \neg Flu, \neg Mal) = 1.0$

+ Whatever inhibits cold from causing fever
  is independent of
  whatever inhibits flu from causing fever

$P(\neg Fev \mid Cold, Flu) \approx P(\neg Fev \mid Cold) \times P(\neg Fev \mid Flu)$

Noisy-OR "CPTable" (2)



- $P(\texttt{Fev} \mid \neg\texttt{Col}, \neg\texttt{Flu}, \neg\texttt{Mal}) = 0$

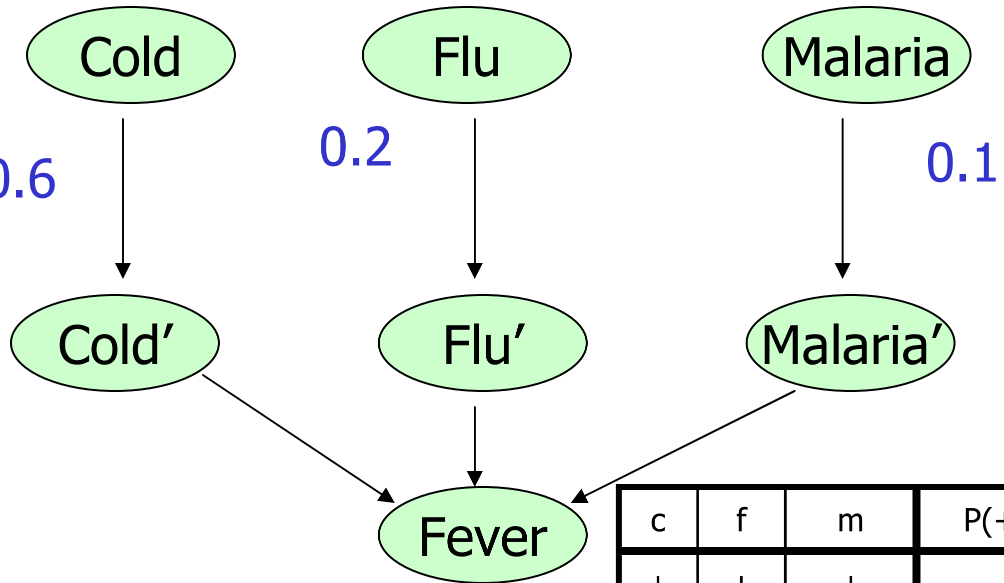
$$P(\neg\texttt{Fev} \mid \texttt{Col}) \approx q_{col} = 0.6$$
$$P(\neg\texttt{Fev} \mid \texttt{Flu}) \approx q_{flu} = 0.2$$
$$P(\neg\texttt{Fev} \mid \texttt{Mal}) \approx q_{mal} = 0.1$$

- Independent inhibiters:
  $P(\neg\texttt{Fev} \mid \texttt{Col},\ \texttt{Flu}) \approx P(\neg\texttt{Fev} \mid \texttt{Col}) \times P(\neg\texttt{Fev} \mid \texttt{Flu})$

$$P(\neg\texttt{Fever} \mid \pm_i d_i) = \prod_{i:+d_i} q_i$$

Cold	Flu	Malaria	$P(\neg\texttt{Fever} \mid \texttt{c,f,m})$	$P(\texttt{Fever} \mid \texttt{c,f,m})$
F	F	F	1.0	0.0
F	F	T	0.1	0.9
F	T	F	0.2	0.8
F	T	T	$0.02 = 0.2 \times 0.1$	0.98
T	F	F	0.6	0.4
T	F	T	$0.06 = 0.6 \times 0.1$	0.94
T	T	F	$0.12 = 0.6 \times 0.2$	0.88
T	T	T	$0.012 = 0.6 \times 0.2 \times 0.1$	0.988

Noisy-Or … expanded



c	P(+cold’ | c)	P(-cold’ | c)
+	$1\text{-}q_c = 0.4$	$q_c = 0.6$
–	0.0	1.0

c	f	m	P(+Fever|c, f, m)
+	+	+	1.0
+	+	–	1.0
+	–	+	1.0
+	–	–	1.0
+	+	+	1.0
–	+	–	1.0
–	–	+	1.0
–	–	–	0.0

Noisy-Or (Gen'l)

- Fever if Cold, Flu or Malaria

Want
$$\begin{cases} P(\texttt{Fev} \mid \neg\texttt{Col}, \neg\texttt{Flu}, \neg\texttt{Mal}) = 0 \\ P(\neg\texttt{Fev} \mid \texttt{Col}) \approx q_{col} = 0.6 \\ P(\neg\texttt{Fev} \mid \texttt{Flu}) \approx q_{flu} = 0.2 \\ P(\neg\texttt{Fev} \mid \texttt{Mal}) \approx q_{mal} = 0.1 \end{cases}$$

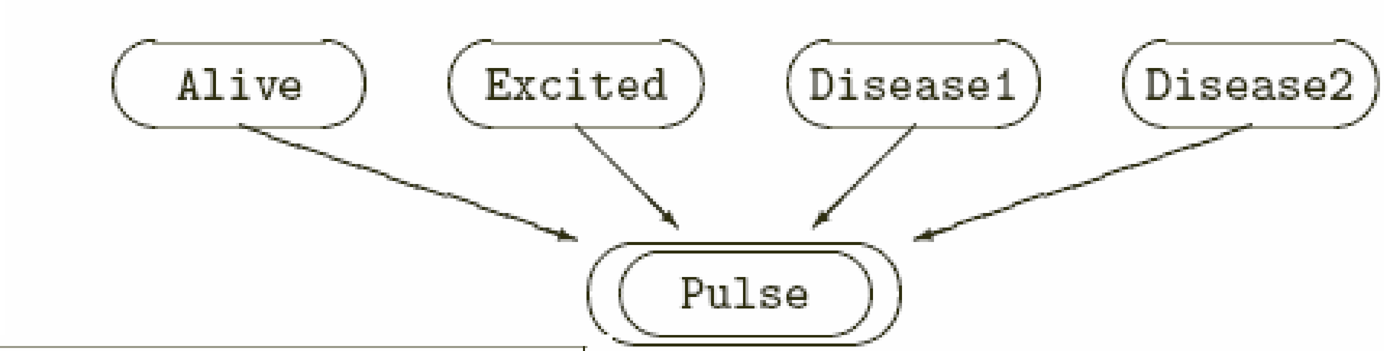
("noise" parameters)

CPCS Network:
- Modeling disease/symptom for internal medicine
- Using Noisy-Or & Noisy-Max
- 448 nodes, 906 links
- Required 8,254 values (not 13,931,430) !

Assumes: – each cause has … effect
– all causes listed
(Leak node, to handle ALL … s . . .)
– inhibiting factors independent

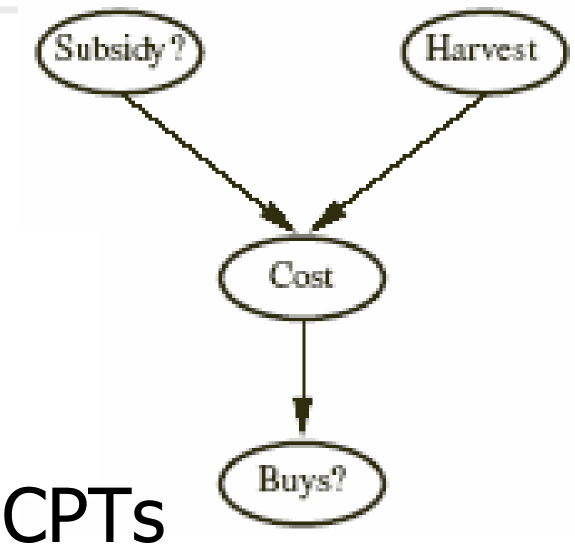
Note: Only k parameters, not 2^k

DecisionTree CPTable



A	E	D1	D2	χ s.t. $P(\text{Pulse}=\chi \mid \text{A,E,D1,D2}) = 1.0$
Y	Y	Y	Y	vhigh
Y	Y	Y	N	vhigh
Y	Y	N	Y	vhigh
Y	Y	N	N	vhigh
Y	N	Y	Y	high
Y	N	Y	N	med
Y	N	N	Y	med
Y	N	N	N	ok
N	Y	Y	Y	none
N	Y	Y	N	none
N	Y	N	Y	none
N	Y	N	N	none
N	N	Y	Y	none
N	N	Y	N	none
N	N	N	Y	none
N	N	N	N	none

Hybrid (discrete+continuous) Networks



- **Discrete:** Subsidy?, Buys?
  Continuous: Harvest, Cost

Option 1: Discretization

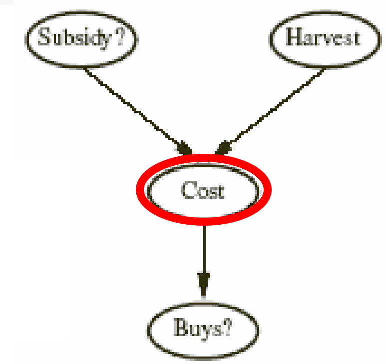
but possibly large errors, large CPTs

Option 2: Finitely parameterized canonical families

Problematic cases to consider. . .

- Continuous variable, discrete+continuous parents
  Cost
- Discrete variable, continuous parents
  Buys?

Continuous Child Variables



- For each "continuous" child E,
  - with continuous parents C
  - with discrete parents D
- Need conditional density function

  P(E = e | C = c, D = d) = $P_{D=d}$(E = e | C = c)

  for each assignment to discrete parents D=d
- Common: linear Gaussian model

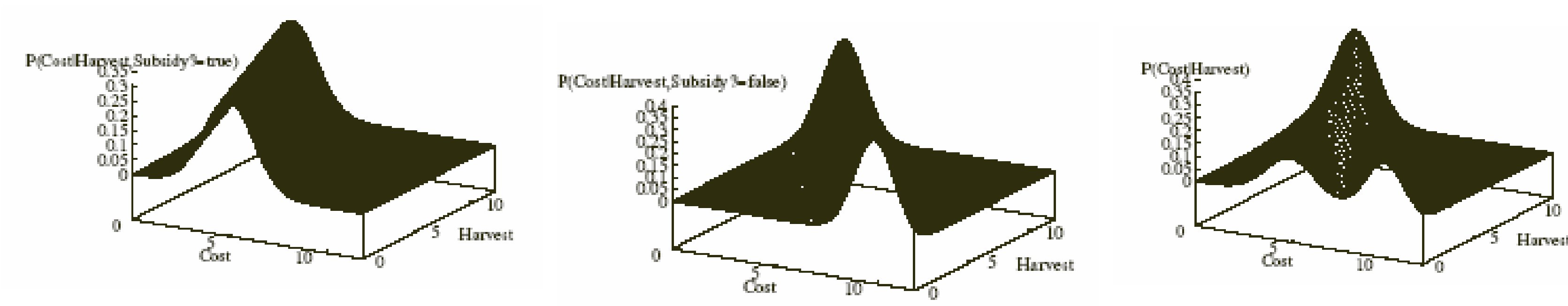
  f(Harvest, Subsidy?) = "dist over Cost"

$$P(\text{Cost} = c \mid \text{Harvest} = h, \text{Subsidy?} = \text{true})$$
$$= \mathcal{N}[a_t h + b_t, \sigma_t](c)$$
$$= \frac{1}{\sigma_t\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{c-(a_t h + b_t)}{\sigma_t}\right)^2\right)$$
$$P(\text{Cost} = c \mid \text{Harvest} = h, \text{Subsidy?} = \text{false})$$
$$= \mathcal{N}[a_f h + b_f, \sigma_f](c)$$

Need parameters:
$\sigma_t \quad a_t \quad b_t$
$\sigma_f \quad a_f \quad b_f$

If everything is Gaussian...

- All nodes continuous w/ LG dist'ns
  ⇒ full joint is a multivariate Gaussian



- Discrete+continuous LG network

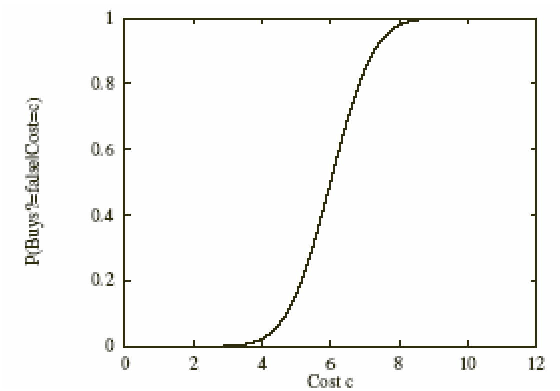
⇒ conditional Gaussian network

multivariate Gaussian over all continuous variables
for each combination of discrete variable values

Discrete variable w/ Continuous Parents

- Probability of Buys? given Cost

  $\approx^{?}$ "soft" threshold:



- Probit distribution uses integral of Gaussian:

$$\Phi(x) = \int_{-\infty}^{x} \mathcal{N}[0,1](x)\, dx$$

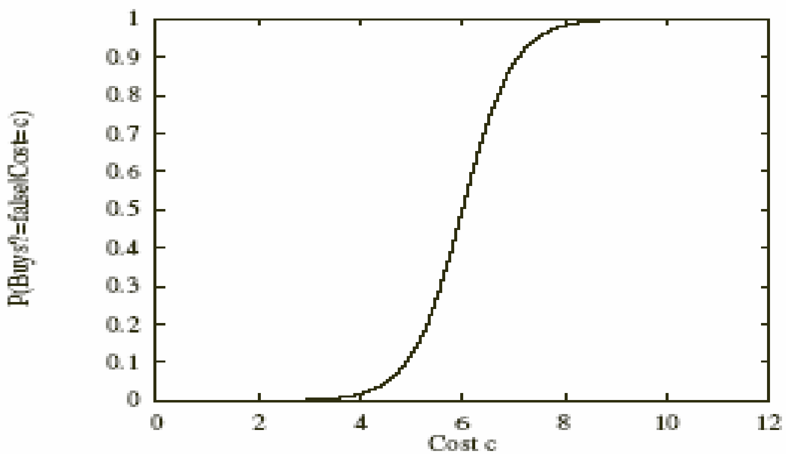
$$P(\text{Buys?} = \text{true} \mid \text{Cost} = c) = \Phi\left(\frac{\mu - c}{\sigma}\right)$$

$\approx$ hard threshold, whose location is subject to noise

Logit vs Probit

- Logit (Sigmoid) used in neural networks:

$$P(\text{Buys?} = \text{true} \mid \text{Cost} = c) = \frac{1}{1 + exp(-2\frac{\mu - c}{\sigma})}$$

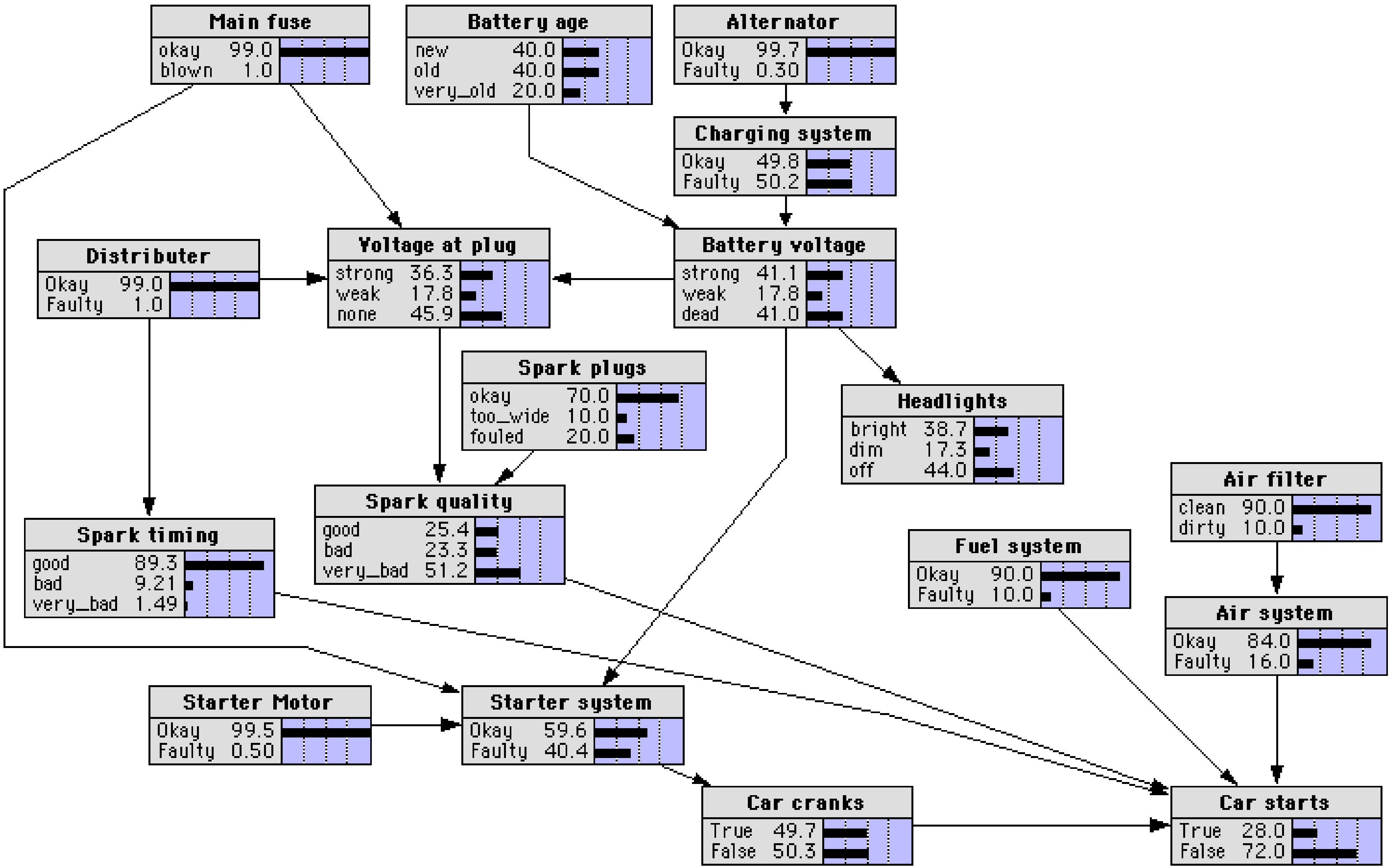


- Shapes:

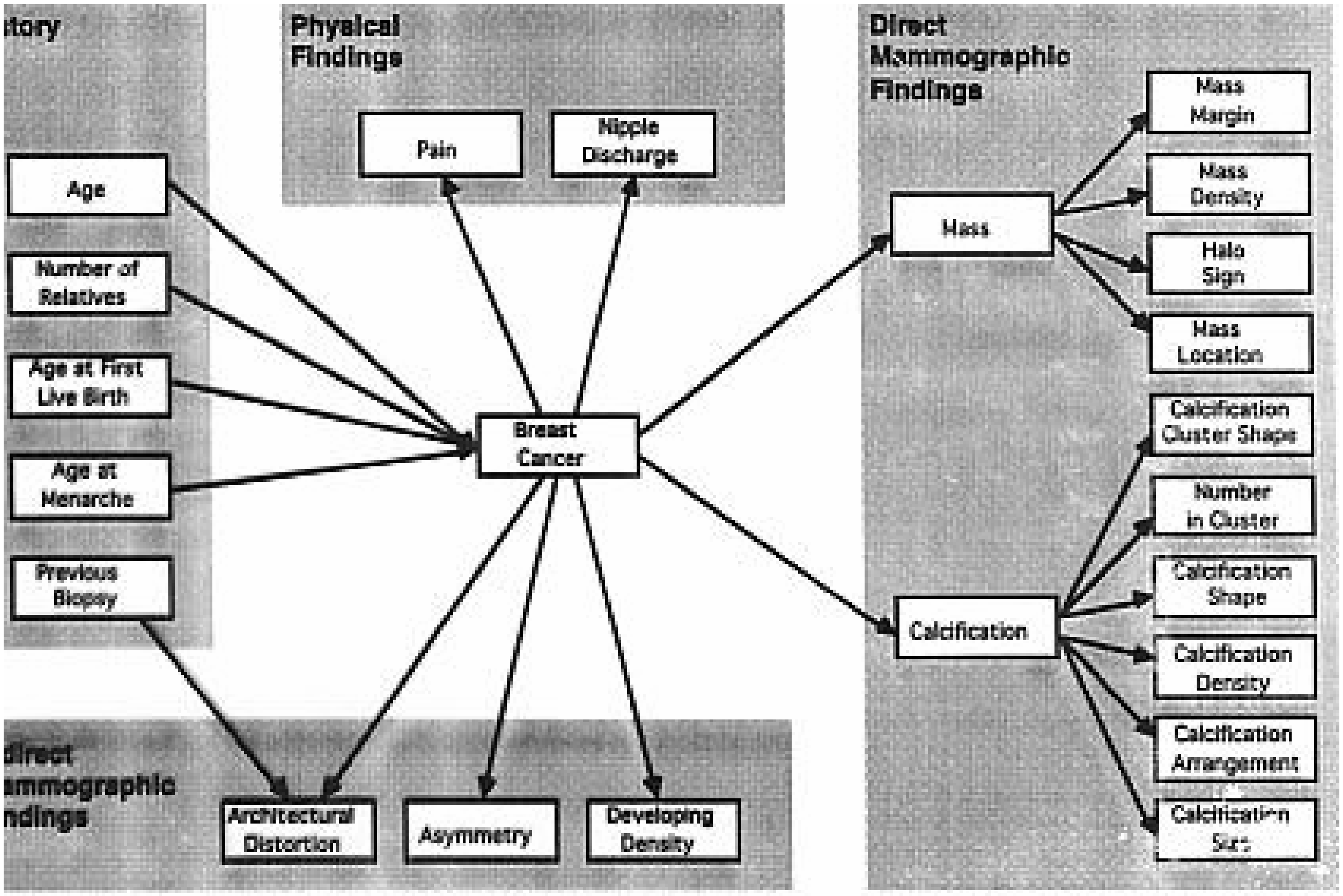
  Logit ≈ Probit

  but Logit has much longer tails

Example: Car Diagnosis

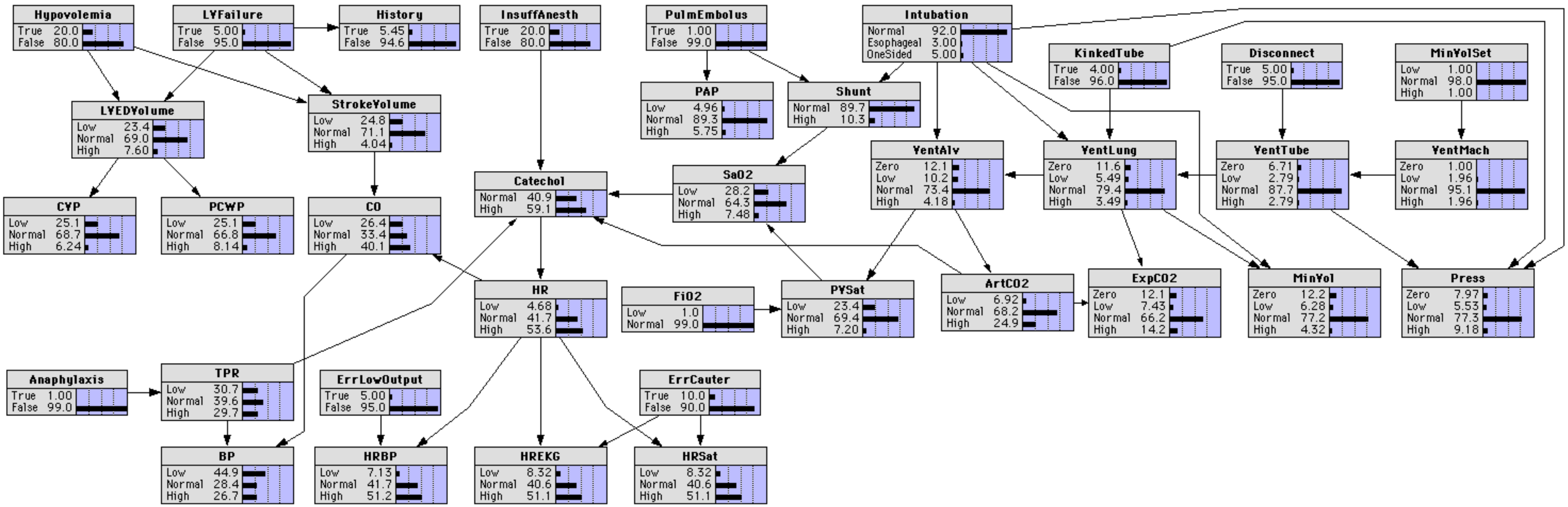


MammoNet



story
Physical Findings
Direct Mammographic Findings
Age
Number of Relatives
Age at First Live Birth
Age at Menarche
Previous Biopsy
Pain
Nipple Discharge
Breast Cancer
Mass
Mass Margin
Mass Density
Halo Sign
Mass Location
Calcification
Calcification Cluster Shape
Number in Cluster
Calcification Shape
Calcification Density
Calcification Arrangement
direct
mmographic
ndings
Architectural Distortion
Asymmetry
Developing Density

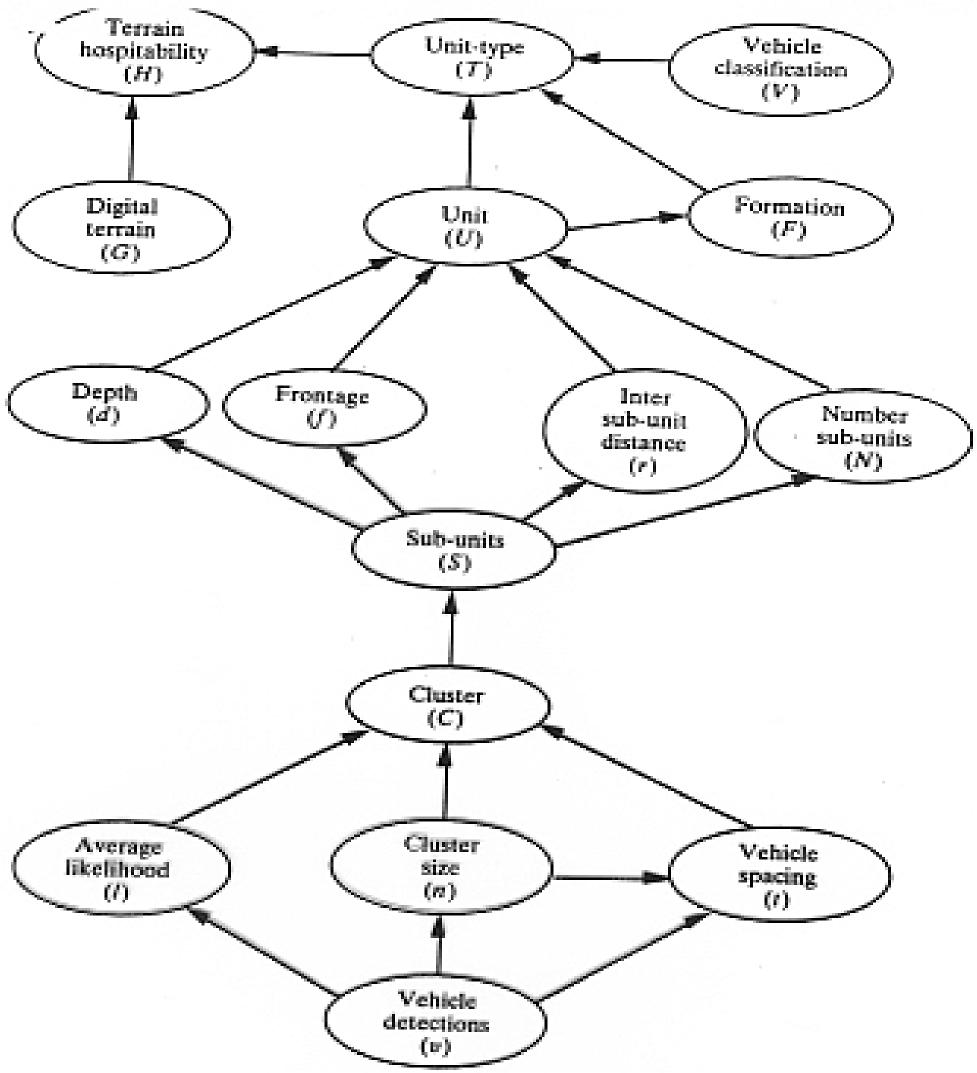
ALARM



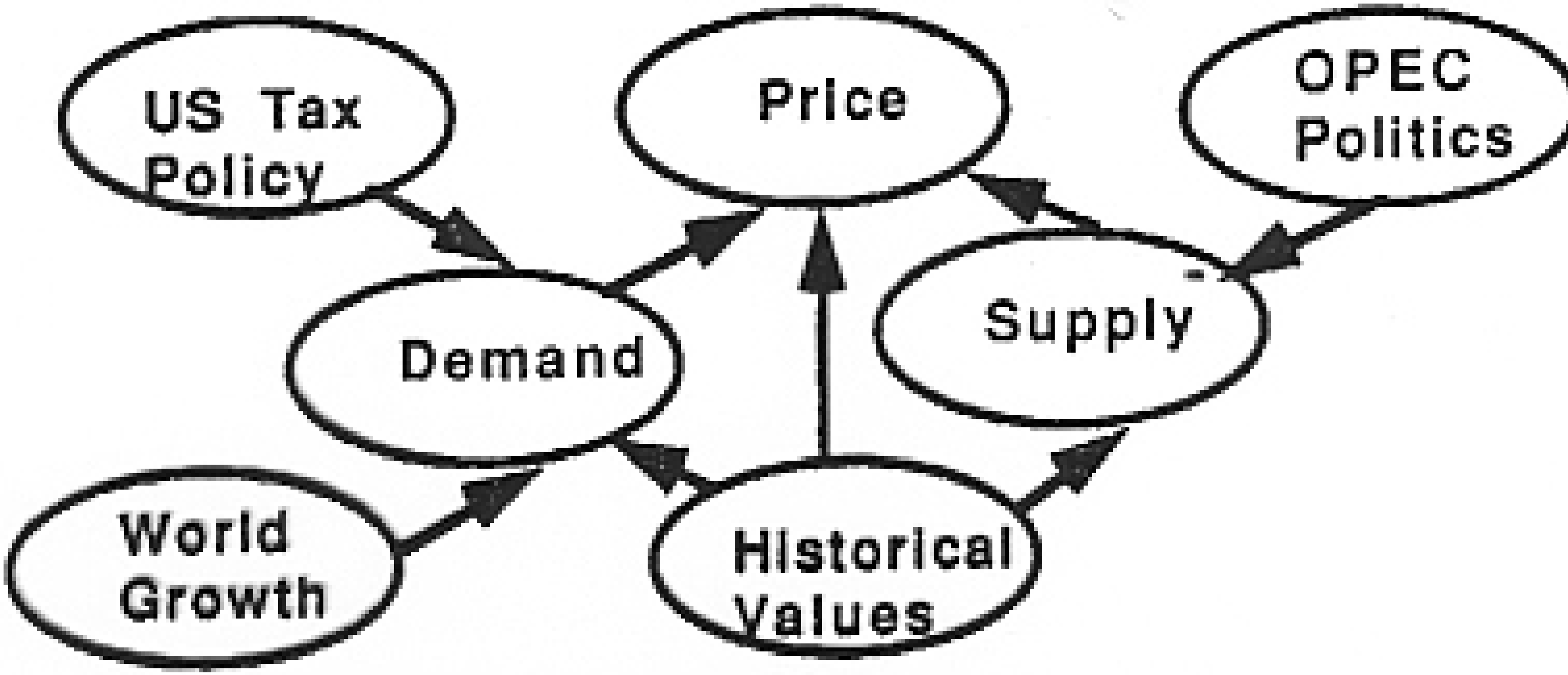
A Logical Alarm Reduction Mechanism

- 8 diagnoses, 16 findings, …

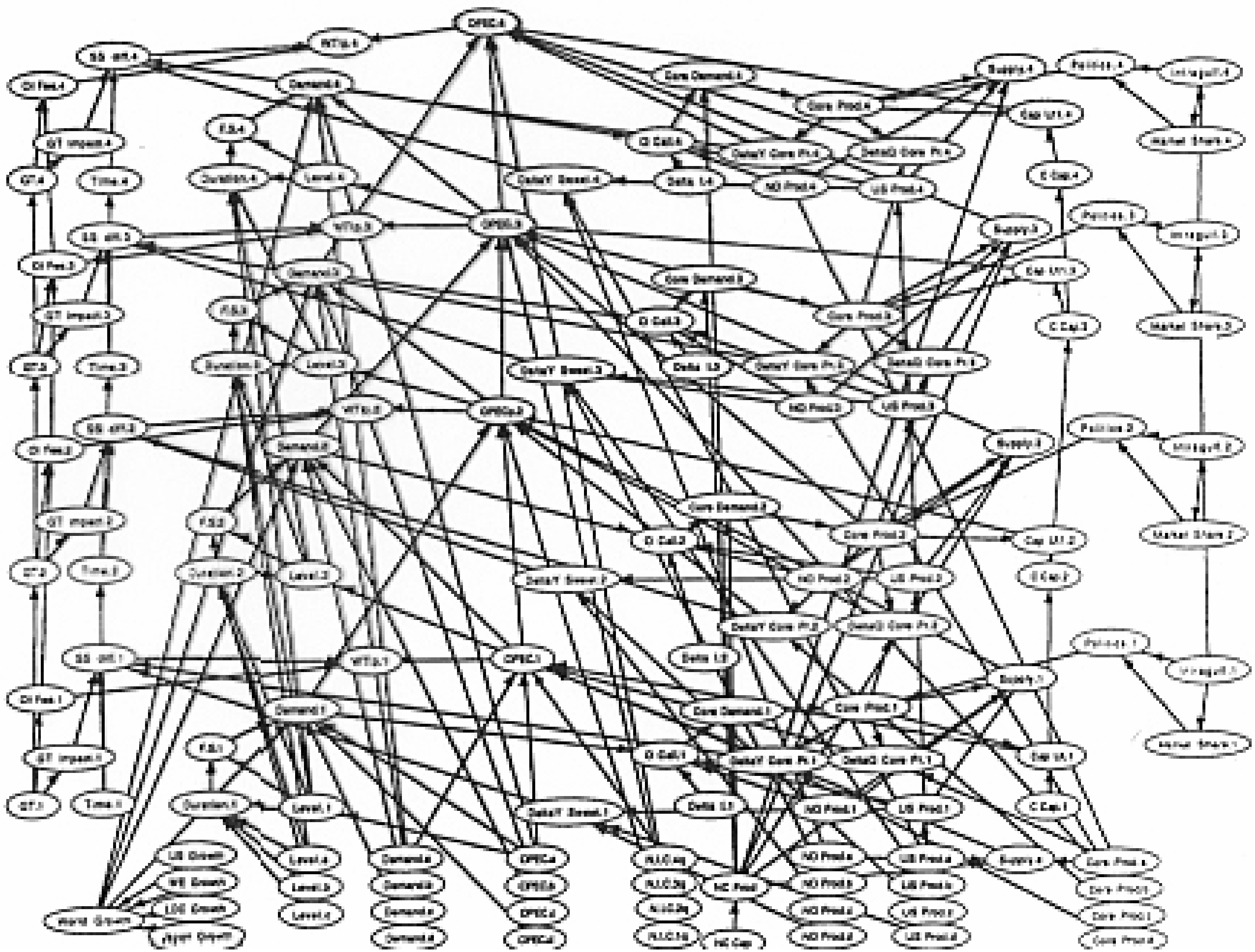
Troup Detection



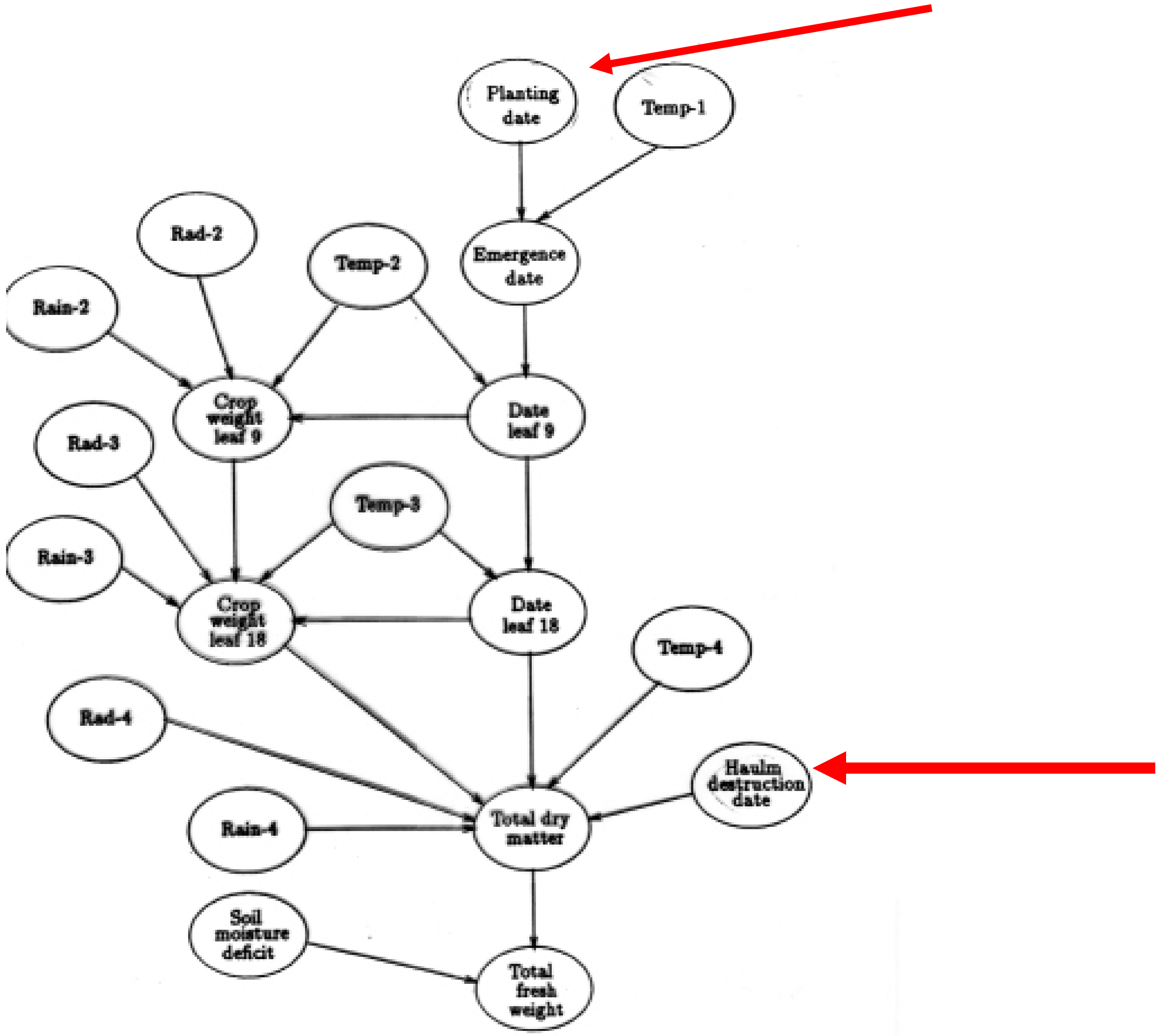
ARCO1: Forecasting Oil Prices



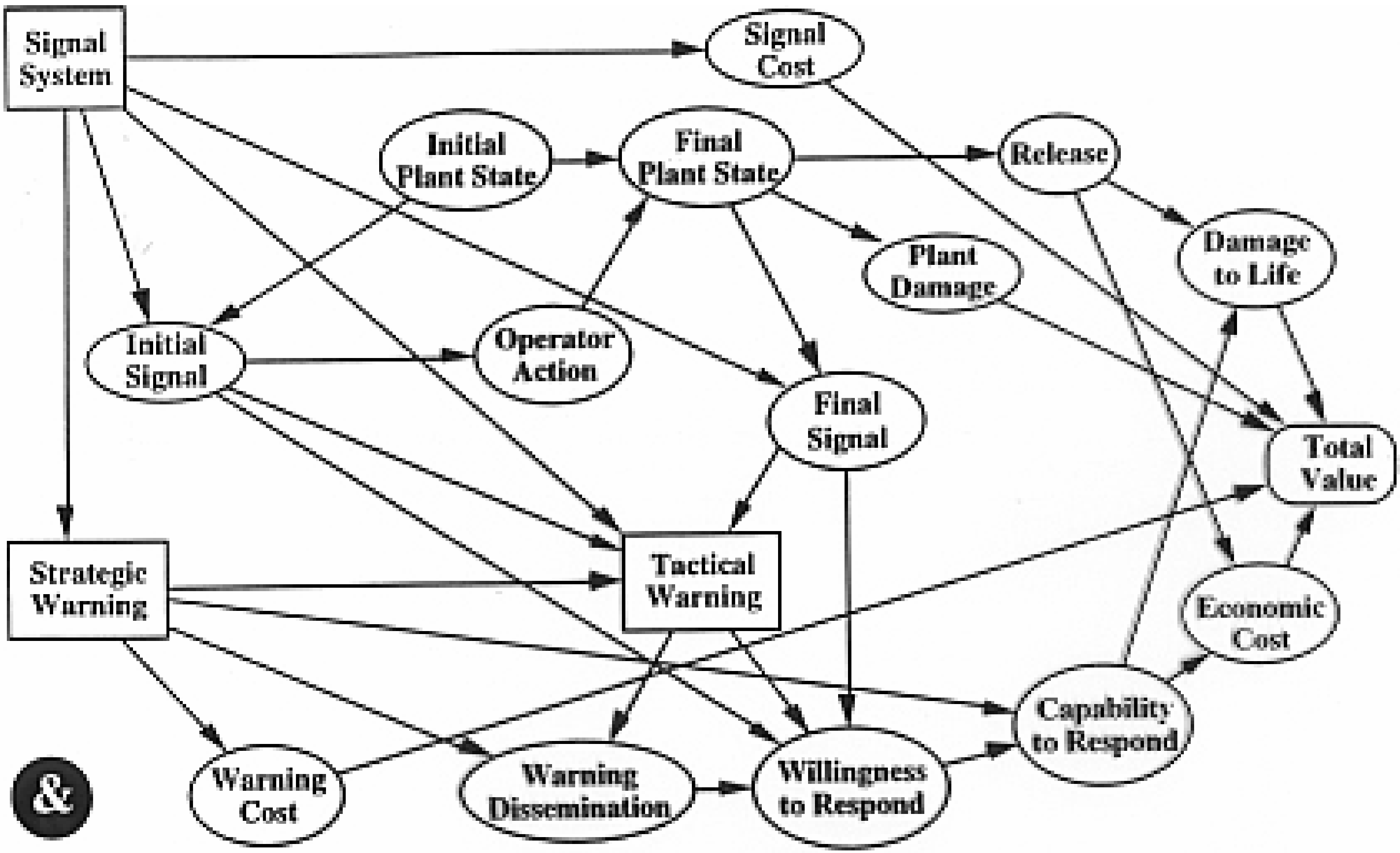
ARCO1: Forecasting Oil Prices



Forecasting Potato Production



Warning System



Uses of Belief Nets #1

- Medical Diagnosis: "Assist/Critique" MD
  - identify diseases not ruled-out
  - specify additional tests to perform
  - suggest treatments appropriate/cost-effective
  - react to MD's proposed treatment

- Decision Support: Find/repair faults in complex machines

  [Device, or Manufacturing Plant, or …]

  … based on sensors, recorded info, history,…

- Preventative Maintenance:

  Anticipate problems in complex machines

  [Device, or Manufacturing Plant, or …]

  …based on sensors, statistics, recorded info, device history,…

Uses (con’t)

- Logistics Support: Stock warehouses appropriately
  …based on (estimated) freq. of needs, costs,
- Diagnose Software:
  Find most probable bugs, given
  program behavior, core dump, source code, …
- Part Inspection/Classification:
  … based on multiple sensors, background, model of production,…
- Information Retrieval:
  Combine information from various sources,
  based on info from various “agents”,…

General: Partial Info, Sensor fusion

-Classification	**-Interpretation**
-Prediction	**-…**

Belief Nets vs Rules

- Both have "*Locality*"
  Specific clusters (rules / connected nodes)
- Often *same nodes* (rep'ning Propositions) but

BN:	Cause	⇒	Effect	
	"Hep	⇒	Jaundice"	$P(J \mid H)$
Rule:	Effect	⇒	Cause	
	"Jaundice	⇒	Hep"	

WHY?: Easier for people to reason ***CAUSALLY***
even if use is ***DIAGNOSTIC***

- BN *provide OPTIMAL way to deal with*
  + ***Uncertainty***
  + ***Vagueness*** ***(var not given, or only dist)***
  + ***Error***

  …Signals meeting Symbols …

- BN *permits* different "direction"'s of inference

Belief Nets vs Neural Nets

- Both have "*graph structure*" but

BN:	Nodes have SEMANTICs Combination Rules: Sound Probability
NN:	Nodes: arbitrary Combination Rules: Arbitrary

- So harder to
  - *Initialize NN*
  - *Explain NN*

  (But perhaps easier to learn NN from examples only?)
- BNs can deal with
  - *Partial Information*
  - *Different "direction"s of inference*

Belief Nets vs Markov Nets

- Each uses "*graph structure*"

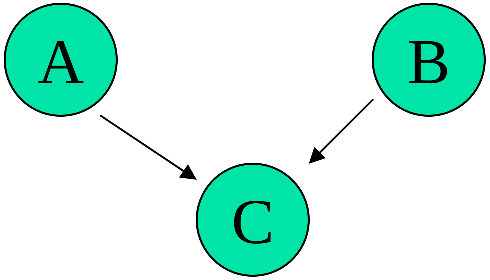
  to FACTOR a distribution

  … explicitly specify dependencies, implicitly independencies…

- but subtle differences…
  - BNs capture "causality", "hierarchies"
  - MNs capture "temporality"

Technical: BNs use DIRECTRED arcs

$\Rightarrow$ allow "induced dependencies"

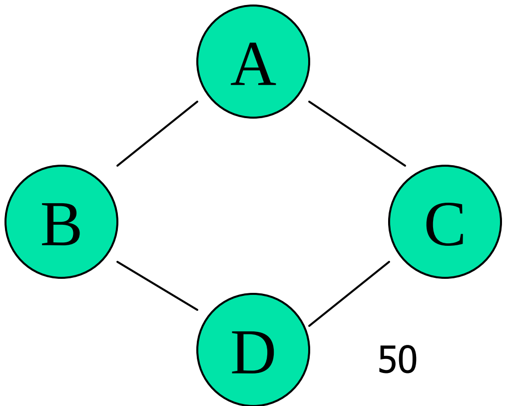


$I(A, \{\}, B)$ "*A* independent of *B*, given {}"

$\neg I(A, C, B)$ "*A* dependent on *B*, given *C*"

MNs use UNDIRECTED arcs

$\Rightarrow$ allow other independencies



$I(A, BC, D)$ *A* independent of *D*, given *B*, *C*

$I(B, AD, C)$ *B* independent of *C*, given *A*, *D*

Summary

- Components of Belief Net
- Conditional Independence
- *d*-separation
  - V-connections
  - Markov blanket
- CPtables
  - Special cases
  - Continuous
- Deployed Examples
- Comparison to other Rep'ns