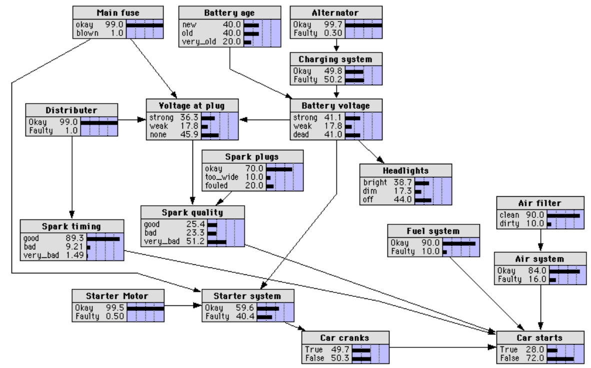
HTF: --
RN: 14
B: 8 – 8.3



Introduction to Bayesian Belief Nets

R Greiner
University of Alberta
Alberta Ingenuity Centre for Machine Learning

aicml

Outline

- Motivation
- What is a Belief Net?
  - … use *connections*… just *some* connections
  - Factored Distribution
  - Reasoning
  - Applications
  - Relation to other Models
- Learning a Belief Net
  - Goal?
  - Learning Parameters – Complete Data
  - Learning Parameters – Incomplete Data
  - Learning Structure

Terms from Probability Theory

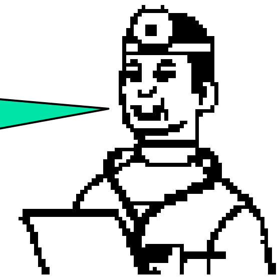
Repeat

- **Random Variable**:
  Weather $\in$ { Sunny, Rain, Cloudy, Snow }
- **Domain**: Possible values a random variable can take.
  (... finite set, $\Re$, ...)
- Probability distribution:
  mapping from domain to values in [0, 1]
- $P(\text{ Weather }) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$

means
$$\left\{ \begin{array}{l} P(\text{ Weather = Sunny }) = 0.7 \\ P(\text{ Weather = Rain }) = 0.2 \\ P(\text{ Weather = Cloudy }) = 0.08 \\ P(\text{ Weather = Snow }) = 0.02 \end{array} \right\}$$

- Event:
  Each assignment (eg, Weather = Rain) is "event"

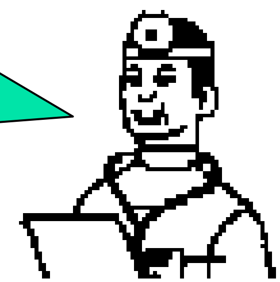
? Hepatitis?



Jaundiced

BloodTest

? Hepatitis,
not Jaundiced
but +BloodTest
?



What is P(+h | -j, +b) ?

Inference by Enumeration

- Using only joint probability distribution:

H	Hepatitis
J	Jaundice
B	(positive) Blood test

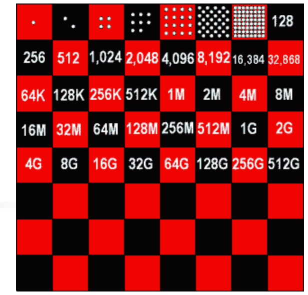
- Can compute *conditional probabilities*:

J	H	B	P(j,b,h)
0	0	0	0.03395
0	0	1	0.0095
0	1	0	0.0003
0	1	1	0.1805
1	0	0	0.01455
1	0	1	0.038
1	1	0	0.00045
1	1	1	0.722

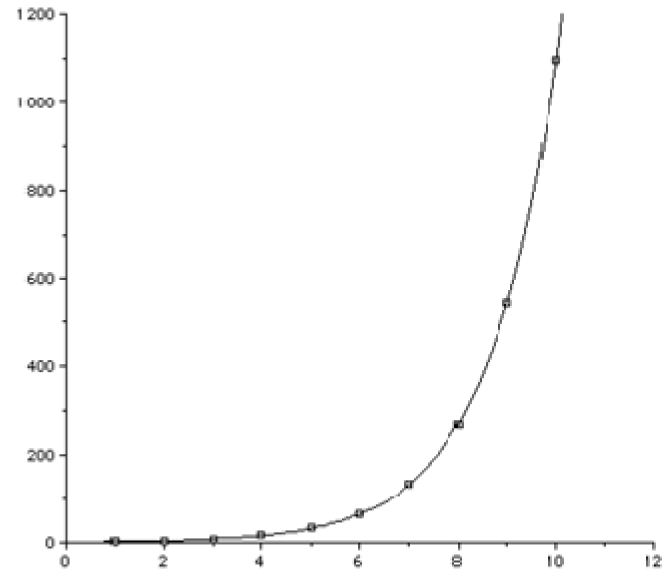
$$P(-h \mid +j) = \frac{P(-h \wedge +j)}{P(+j)} = \frac{0.01455 + 0.038}{0.01455 + 0.038 + 0.00045 + 0.722}$$

$$\approx 0.0678$$

Just use Joint ??



- Problems with full joint?
  - Too big ($\geq 2^n$)
    - How to acquire?
    - Too slow
      (inference requires adding 2^k...)



- Better:
  - Encode dependencies
  - Encode only *relevant* dependencies

[illegible]	[illegible]
:	:
30	1,073,741,824
:	:

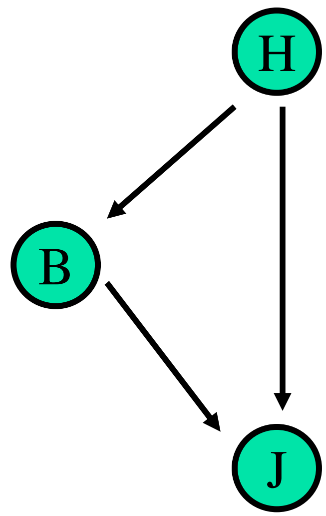
Table is Sufficient

- Just need single table!! But…
- Unnatural:
  - Easier to think about CORRELATIONS
    - P(Jaudice | Hepatitis)
    - P(DimLight | BadBattery), …
  - ⇒ better to use CONDITIONAL EVENTS
- Too MANY NUMBERS!!
  - Exponential size to store
    $O(2^N)$ numbers…
  - Exponential cost for inference
  - ⇒ only use some connections

⇒ Bayesian Belief Net

J	B	H	P(j,b,h)
0	0	0	0.03395
0	0	1	0.0095
0	1	0	0.0003
0	1	1	0.1805
1	0	0	0.01455
1	0	1	0.038
1	1	0	0.00045
1	1	1	0.722

Simple Belief Net



P(H=1)	P(H=0)
0.05	0.95

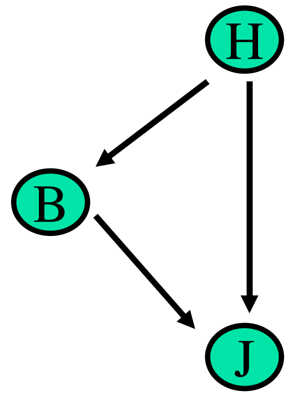
h	P(B=1 | H=h)	P(B=0 | H=h)
1	0.95	0.05
0	0.03	0.97

h	b	P(J=1|h,b)	P(J=0|h,b)
1	1	0.8	0.2
1	0	0.8	0.2
0	1	0.3	0.7
0	0	0.3	0.7

- Node ~ Variable
  Link ~ “Causal dependency”
- “CPTable” ~ P(child | parents)

Encoding Causal Links

h	P(B=1 | H=h)
1	0.95
0	0.03



P(H=1)
0.05

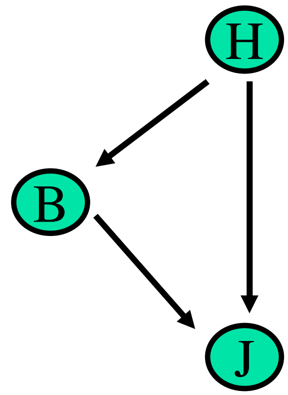
h	b	P(J=1|h , b)
1	1	0.8
1	0	0.8
0	1	0.3
0	0	0.3

- $P(J \mid H, B=0) \;=\; P(J \mid H, B=1) \quad \forall\, J,\; H\,!$

  $\Rightarrow \quad \mathbf{P(\,J \mid H, B) \;=\; P(J \mid H)}$
- J is <u>INDEPENDENT</u> of B, once we know H
- Don't need $B \rightarrow J$ arc!

Encoding Causal Links

h	P(B=1 | H=h)
1	0.95
0	0.03



P(H=1)
0.05

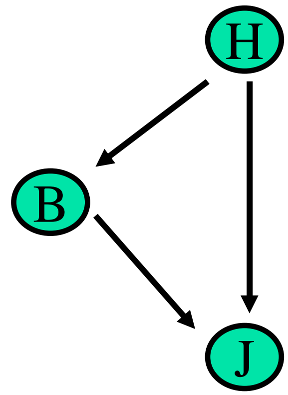
h		P(J=1|h)
1		0.8
1		
0		0.3
0		

- $P(J \mid H, B=0) \;=\; P(J \mid H, B=1) \quad \forall\, J,\; H\,!$

  $\Rightarrow \quad$ **$\mathbf{P(J \mid H, B) \;=\; P(J \mid H)}$**
- *J* is <u>INDEPENDENT</u> of *B*, once we know *H*
- Don't need $B \rightarrow J$ arc!

Encoding Causal Links

h	P(B=1 | H=h)
1	0.95
0	0.03



P(H=1)
0.05

h	P(J=1|h)
1	0.8
0	0.3

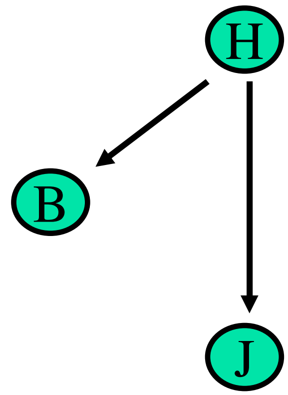
- $P(J \mid H, B=0) = P(J \mid H, B=1) \quad \forall J, H$!

  $\Rightarrow$ **$P(J \mid H, B) = P(J \mid H)$**
- J is <u>INDEPENDENT</u> of B, once we know H
- Don't need $B \rightarrow J$ arc!

Sufficient Belief Net

P(H=1)
0.05

h	P(B=1 | H=h)
1	0.95
0	0.03



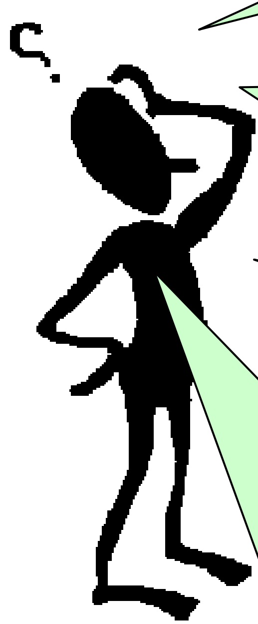
h	P(J=1|h)
1	0.8
0	0.3

- Requires: *P(H=1)* known
  P(J=1 | H=1) known
  P(B=1 | H=1) known

  (Only 5 parameters, not 7)

Hence: $$P(\text{H}=1 \mid \text{B}=1, \text{J}=0) = \frac{1}{\alpha}\ P(\text{H}=1)\ P(\text{B}=1 \mid \text{H}=1)\ \overset{P(\text{J}=0 \mid \text{H}=1)}{\cancel{P(\text{J}=0 \mid \text{B}=1,\text{H}=1)}}$$

What is probability that Fred is …
Jaundiced, given {}?



$P(+j) = 0.325$

… Jaundiced, given –BloodTest ?

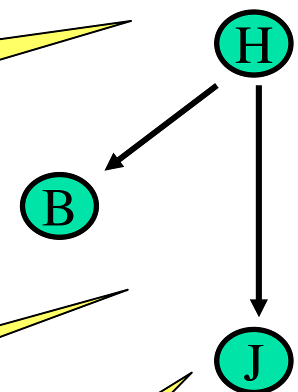
$P(+j \mid -b) = 0.301$

… Jaundiced, given +Hepatitis?

$P(+j \mid +h) = 0.8$

… Jaundiced, given +Hepatitis, -BloodTest ?

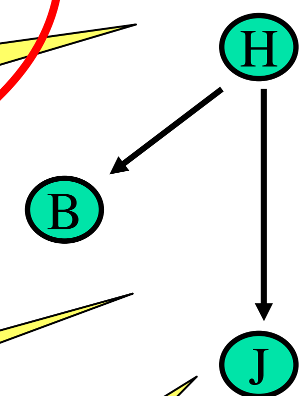
Same: $P(+j \mid +h, -b) = 0.8$



So Jaundice DOES depend on BloodTest, initially

$P(+j) = 0.325$

$P(+j \mid -b) = 0.301$

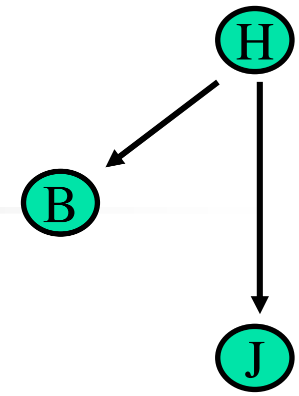


But Jaundice does NOT depend on BloodTest, given Hepatitis

$P(+j \mid +h) = 0.8$

Same: $P(+j \mid +h, -b) = 0.8$

Dependencies…



- B *does* depend on J:

  If J=1, then likely that H=1 $\Rightarrow$ B =1
- *but… ONLY THROUGH* H:
  - If know H=1, then likely that B=1
  - … doesn't matter whether J=1 or J=0 !

$$\Rightarrow \quad \boxed{\mathbf{P(J{=}0 \mid B{=}1,\ H{=}1) = P(J{=}0 \mid H{=}1)}}$$

N.b., B *and* J ARE *correlated* a priori $P(J \mid B \quad) \neq P(J)$
GIVEN H, *they become uncorrelated* $P(J \mid B, H) = P(J \mid H)$

Factored Distribution

- Symptoms *independent,* given Disease

H	Hepatitis
J	Jaundice
B	(positive) Blood test

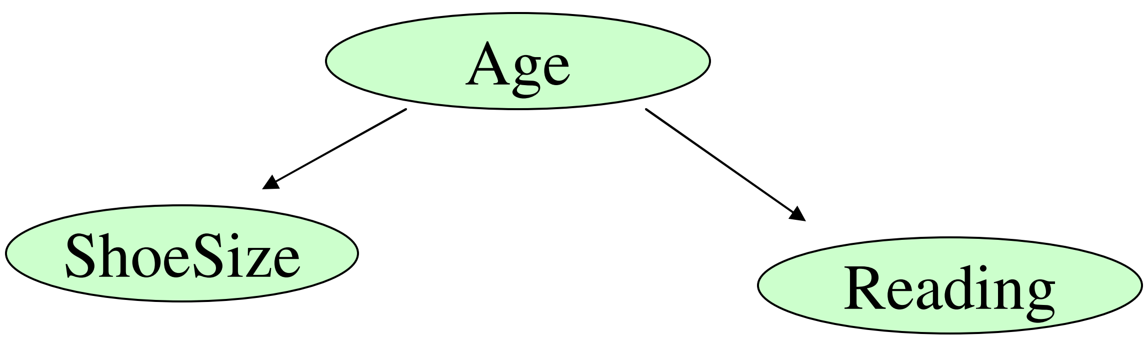
$P(\,B \mid J\,) \neq P(\,B\,)$ but
$P(\,B \mid J,H\,) = P(\,B \mid H\,)$

- **ReadingAbility** and **ShoeSize** are dependent,

$P(\text{ ReadAbility} \mid \text{ShoeSize }) \neq P(\text{ ReadAbility })$

but become independent, given Age

$P(\text{ ReadAbility} \mid \text{ShoeSize, Age }) = P(\text{ ReadAbility} \mid \text{Age })$



(a) Independence

- Coin tosses:
  - T_1: the first toss is a head; T_2: the second toss is a tail
  - $P(T_2 \mid T_1) = P(T_2)$
- α and β ***independent*** iff $P(\beta|\alpha)=P(\beta)$
  - $P \models (\alpha \perp \beta)$
  - … distr'n P entails α independent of β
- **Proposition:** α and β *independent*
  if and only if
  $P(\alpha,\beta) = P(\alpha)\, P(\beta)$

Independence

- Events α and B are independent *iff*
  - $P(\alpha \,\&\, \beta) = P(\alpha)\, P(\beta)$
  - $P(\alpha \mid \beta) = P(\alpha)$
  - $P(\alpha \vee \beta) = 1 - (1 - P(\alpha))\,(1 - P(\beta))$

- Variables independent

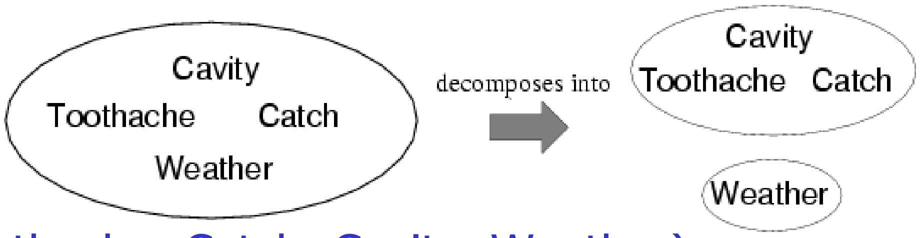
  $\Leftrightarrow$ independent for all values

$\forall a, b \quad P(A = a, B = b) = P(A = a)\; P(B = b)$

Hide

Independence

- *A* and *B* are independent iff
  $P(A|B) = P(A)$ or $P(B|A) = P(B)$ or $P(A, B) = P(A)\,P(B)$



$P(Toothache, Catch, Cavity, Weather)$
$= P(Toothache, Catch, Cavity)\,P(Weather)$

- 16 entries reduced to 9;
  for *n* independent biased coins, $O(2^n) \rightarrow O(n)$

- Absolute independence powerful… but rare

- Dentistry is a large field with hundreds of variables, none of which are independent.
  … What to do?

(b) Conditional independence

- Independence is rarely true unconditionally… but is *conditionally*…
  - Shoe size is NOT independent of Reading Ability
  - But is independent, given AGE…

- α and β ***conditionally independent*** given γ if $P(\beta \mid \alpha, \gamma) = P(\beta \mid \gamma)$
  - $P \models (\alpha \perp \beta \mid \gamma)$

Proposition: $P \models (\alpha \perp \beta \mid \gamma)$ if and only if

$$P(\alpha, \beta \mid \gamma) = P(\alpha \mid \gamma)\, P(\beta \mid \gamma)$$

Conditional Independence

- **P**(*Hep, Jaun, BT*) has $2^3 - 1 = 7$ independent entries
- Given +Hep, Jaun doesn't depend on blood test :

  (1) P(Jaun | +h, BT) = P(Juan| +h)
- Given –Hep, Jaun doesn't depend on blood test :

  (2) P(Jaun | –h, BT) = P(Juan| –h)

Conditional Independence

- Events E_1 and E_2 are conditionally independent given E iff

$$P(E_1 \mid E, E_2) = P(E_1 \mid E)$$

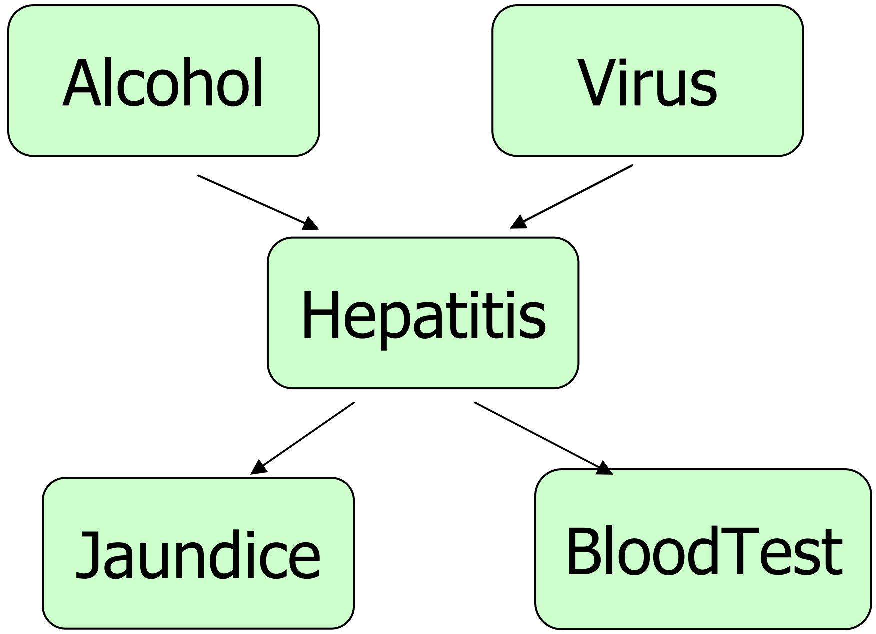
- Given E, knowing E_2 does not change the probability of E_1
- Equivalent formulations:

$$P(E_1, E_2 \mid E) = P(E_1 \mid E)\, P(E_2 \mid E)$$
$$P(E_2 \mid E, E_1) = P(E_2 \mid E)$$

Bigger Networks

P(A=1)
0.32

P(V=1)
0.20



a	v	P(H=1|a ,v)
1	1	0.82
1	0	0.10
0	1	0.45
0	0	0.04

h	P(J=1| h)
1	0.8
0	0.3

h	P(B=1| h)
1	0.98
0	0.01

- Intuition: Show *CAUSAL* connections:
  Alcohol CAUSES Hepatitis; Hepatitis CAUSES Jaundice
- If Alcohol, then expect Jaundice: $\boxed{\text{Alcohol} \Rightarrow \text{Hepatitis} \Rightarrow \text{Jaundice}}$

But only via Hepatitis: $\boxed{\text{Alcohol and not Hepatitis} \not\Rightarrow \text{Jaundice}}$

$$P(\text{J} \mid \text{A}) \neq P(\text{J}) \quad \text{but}$$
$$P(\text{J} \mid \text{A, H}) = P(\text{J} \mid \text{H})$$

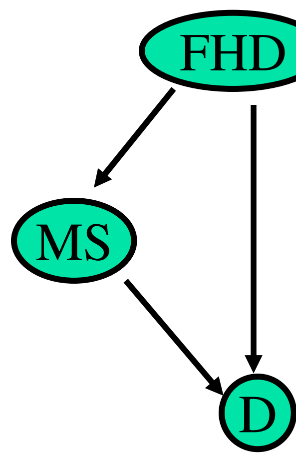
Less Trivial Situations

- *N.b.,* obs_1 is *not* always independent of obs_2 given H
- *Eg,* **FamilyHistoryDepression** 'causes' **MotherSuicide** and **Depression**

  MotherSuicide causes **Depression** (w/ or w/o **F.H.Depression)**

P(FHD=1)
0.001

f	P(MS=1 | FHD=f)
1	0.10
0	0.03



f	m	P(D=1 | FHD=f, MS=m)
1	1	0.97
1	0	0.90
0	1	0.08
0	0	0.04

- Here, $P(\text{ D} \mid \text{MS, FHD }) \neq P(\text{ D} \mid \text{FHD })$!

- Can be done using Belief Network,
  but need to specify:

$P(\ FHD\ )$	1
$P(\ MS \mid FHD\ )$	2
$P(\ D \mid MS, FHD\ )$	4

Advantages of Belief Net

- All of advantages of *Probability Theory*
  - Not CertaintyFactor, Fuzzy, Dempster-Schaeffer, …
  - Formal understanding of how things relate
  - Well-defined inference

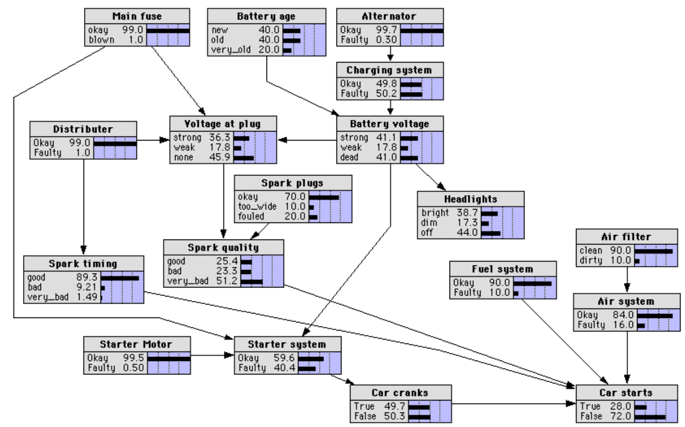
- Explanatory power
  - What is related to what? … and how strongly?

- Efficient encoding
  - 10 values, not 32…
  - 8,254 values, not 13,931,430… not 2^{422}
    (CPCS Network: Modeling disease/symptom for internal medicine)

- Effective learning…

What to do with a Belief Net?

- Examine its connections
  - What depends on what?

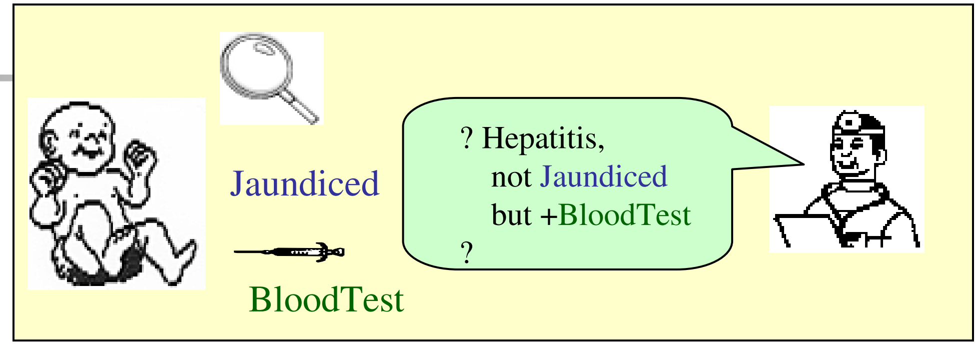


- Get answers to specific questions
  - What is P(Cancer | G_3=+, Age>52) ?
  - What is most likely cause of symptoms?
  - …

Outline

- Motivation
- What is a Belief Net?
  - Example
  - Inference
  - Semantics
  - Applications
  - Relation to other Models
- Learning a Belief Net

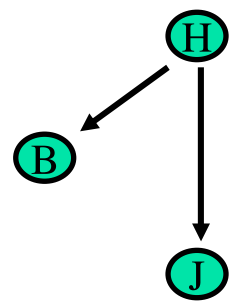
Classification



- Which is more likely: $+h$ vs $-h$?

- Given independencies:
  + values:

h	P(+b | h)	P(-b | h)
1	0.95	0.05
0	0.03	0.93



P(+h)	P(-h)
0.05	0.95

h	P(+j |h)	P(-j |h)
1	0.8	0.2
0	0.3	0.7

- $\text{argmax}_h\ P(\ h \mid +b, -j\ )$
  $= \text{argmax}_h\ P(\ h\ ) \times P(+b \mid h\ ) \times P(-j \mid h)$
  $= \text{argmax}_h\ \{\ 0.05 \times 0.95 \times 0.2,\ \ 0.95 \times 0.03 \times 0.7\ \}$

$\boxed{-h \quad \text{as} \quad 0.0095 \quad < \quad 0.01995}$

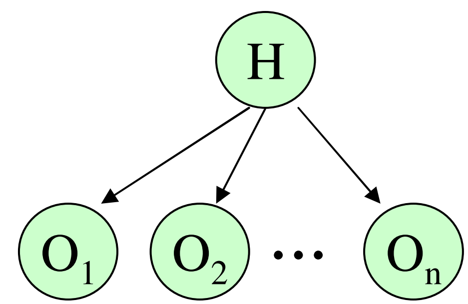
"Naïve Bayes"

- Classification Task:

  Given $\{ O_1 = v_1, \ldots, O_n = v_n \}$

  Find h_i that maximizes $P(H = h_i \mid O_1 = v_1, \ldots, O_n = v_n)$

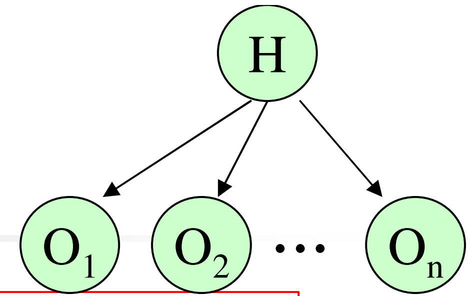
- Given $\left\{ \boxed{\begin{array}{l} P(H = h_i) \\ P(O_j = v_k \mid H = h_j) \\ \\ \textit{Independent:} \quad P(O_j \mid H, O_k, \ldots) = P(O_j \mid H) \end{array}} \right.$



$$\boxed{P(H = h_i \mid O_1 = v_1 \ldots, O_n = v_n) = \frac{1}{\alpha} P(H = h_i) \prod_j P(O_j = v_j \mid H = h_i)}$$

- Find argmax $\{h_i\}$

Naïve Bayes (con't)



$$P(H = h_i \mid O_1 = v_1, \ldots, O_n = v_n) \;=\; \frac{1}{\alpha}\, P(H = h_i) \prod_j P(O_j = v_j \mid H = h_i)$$

- *Normalizing term*

$$\alpha = P(O_1 = v_1, \ldots, O_n = v_n) \;=\; \sum_i P(H = h_i) \prod_j P(O_j = v_j \mid H = h_i)$$

(No need to compute, as same for all h_i)

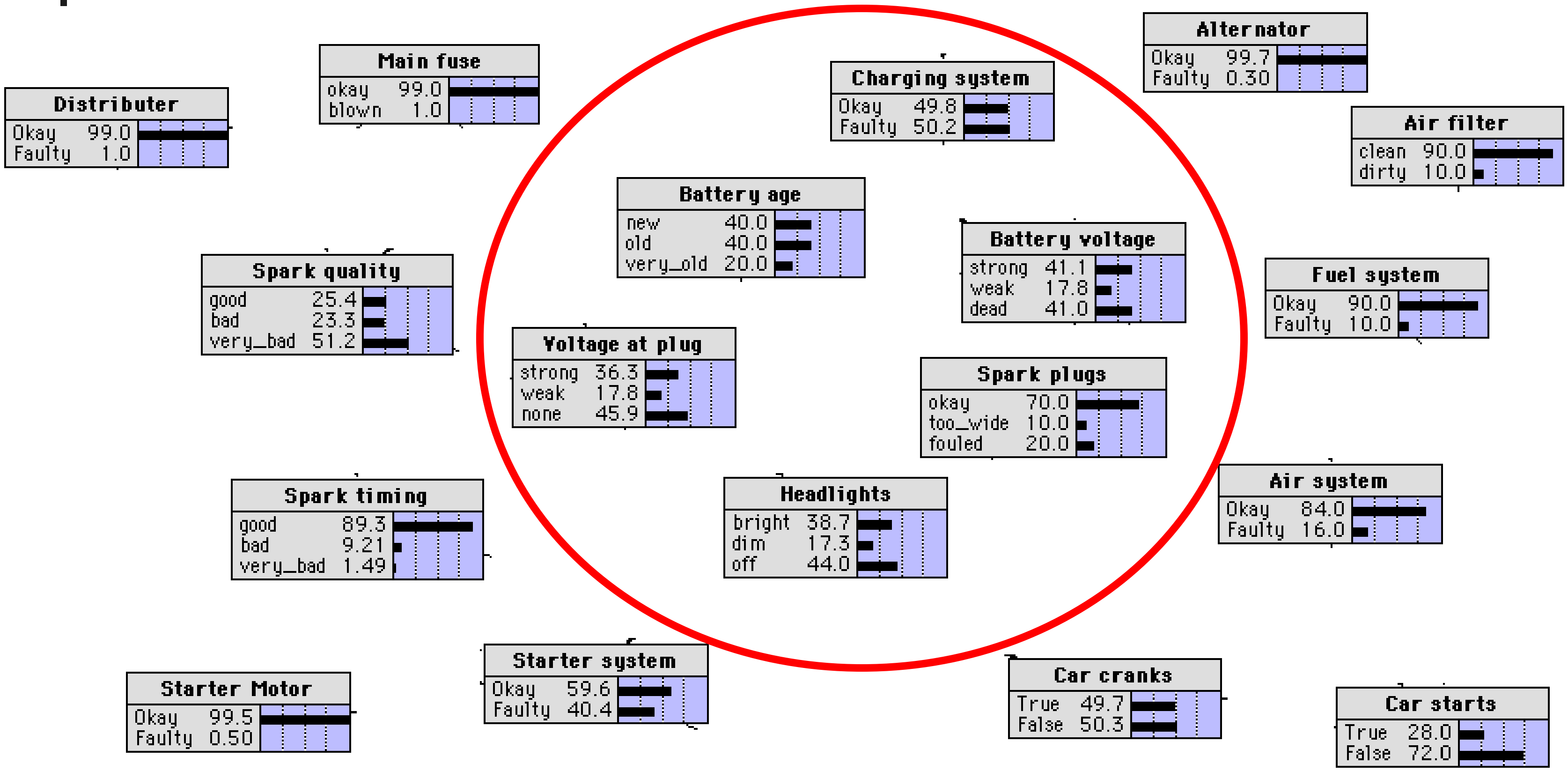
- Easy to use for Classification
- Can use even if some v_js not specified
- If k Dx's and n O_is,

  requires only k priors, $n \times k$ pairwise-conditionals

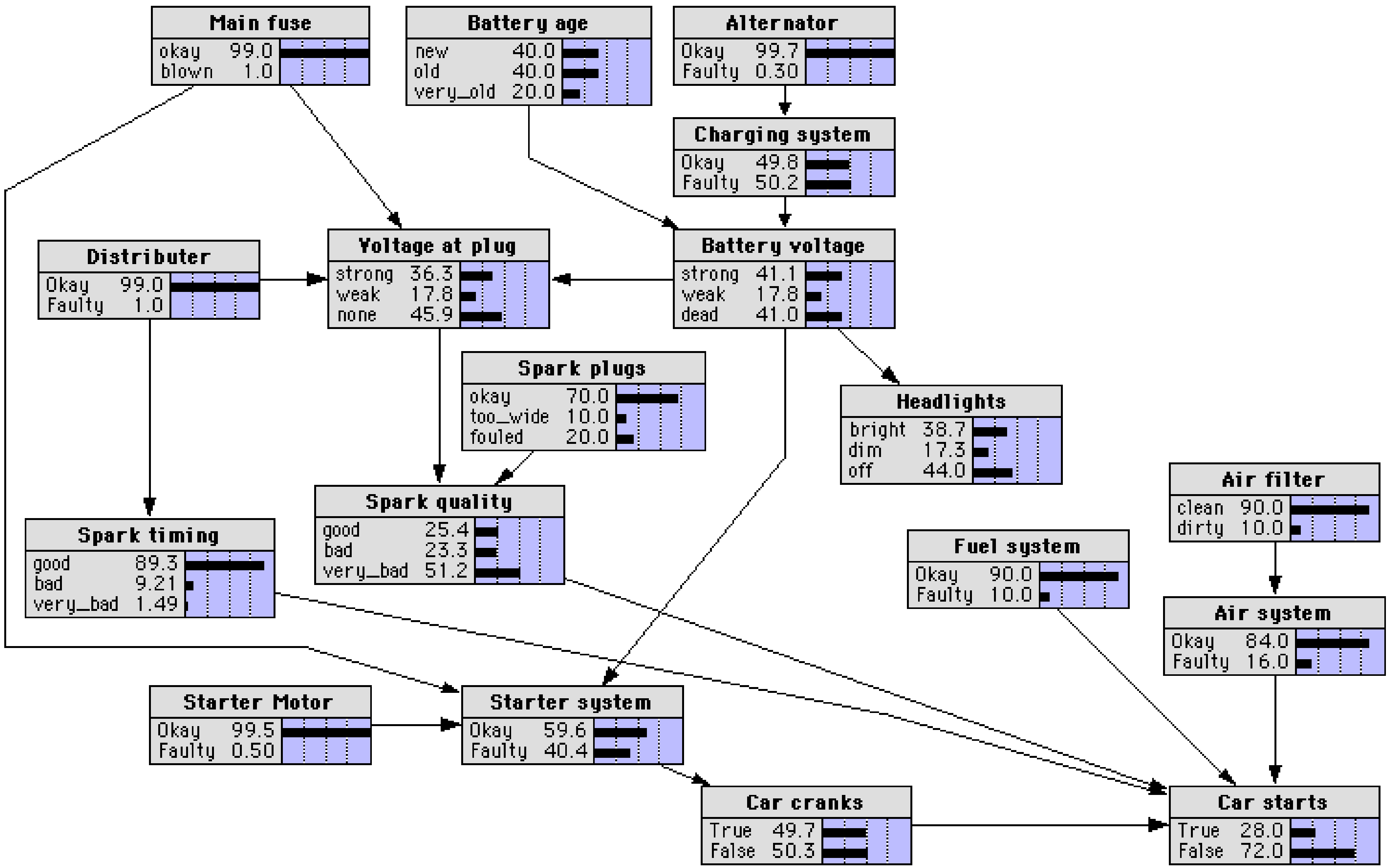
  (Not 2^{n+k}… relatively easy to learn)

n	1+2n	$2^{n+1} - 1$
10	21	2,047
30	61	2,147,438,647

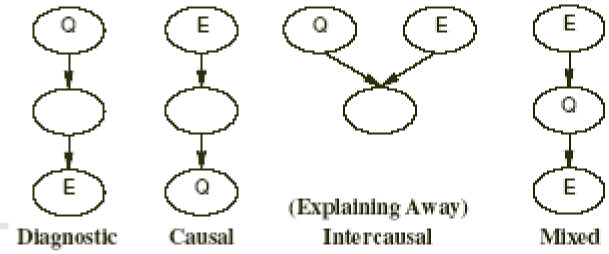
Engineer a Belief Net



Example: Car Diagnosis



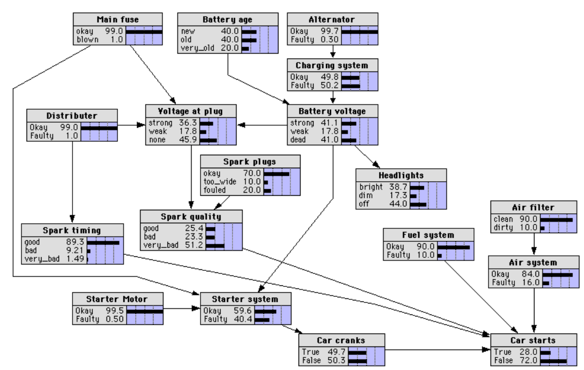
Types of Reasoning



- Typical case: P(QueryVar | EvidenceVars = vals)
  - Eg: P(+starts | +fuel, -voltage)
- **Diagnostic**: from effect to (possible) causes
  - P(-fuse | -starts) = 0.016
- **Causal**: from cause to effects
  - P(-starts | -fuse) = 0.86
- **InterCausal**: between causes of common effect
  - P(-fuel | -starts) = 0.376
  - P(-fuel | -starts, -filter) = 0.003

  Bad_Filter EXPLAINS no_start, and so

  Bad_Filter EXPLAINS AWAY low-fuel
- **Mixed**: combinations of . . .
  - P(+headlights | +voltage, -starts) = 0.03



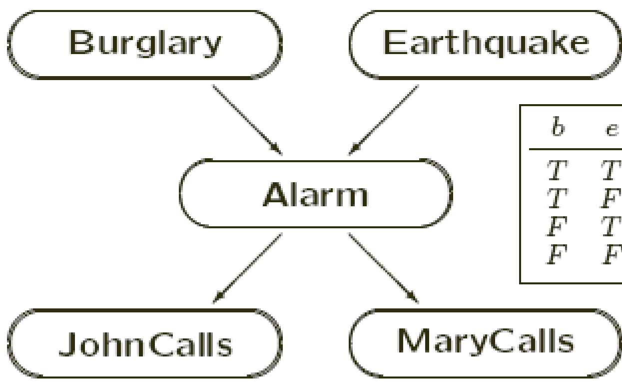
Outline

- Motivation
- What is a Belief Net?
  - Example
  - Inference
  - Semantics
  - Applications
  - Relation to other Models
- Learning a Belief Net

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

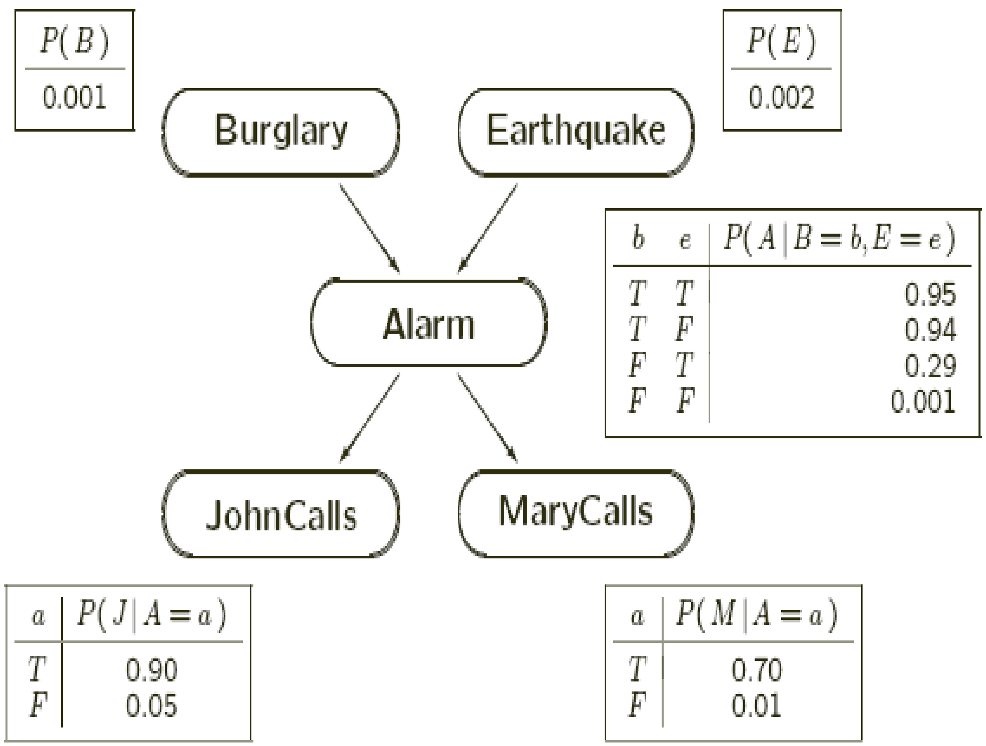
a	$P(M \mid A=a)$
T	0.70
F	0.01

- **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



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

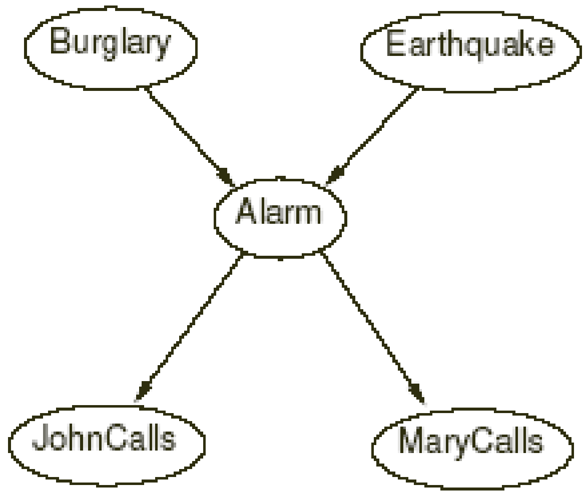
(CPTable =

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

)

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

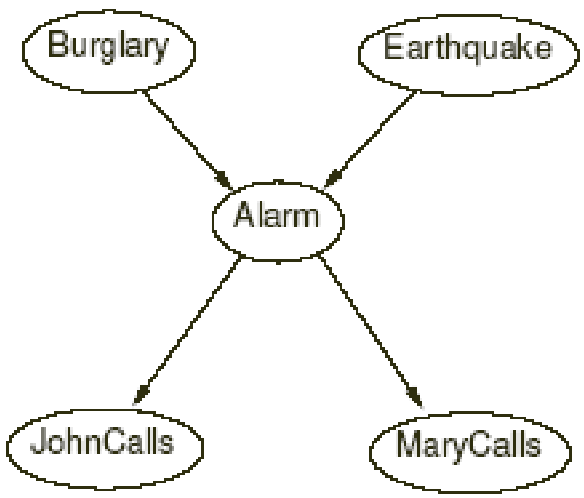
Independence in a Belief Net



- Burglary, Earthquake independent
  - B ⊥ E
- Given Alarm, JohnCalls and MaryCalls independent
  - J ⊥ M | A

  - JohnCalls is correlated with MaryCalls ¬(J ⊥ M) as suggest Alarm
  - But given Alarm, JohnCalls gives no NEW evidence wrt MaryCalls

The Independence Assumption

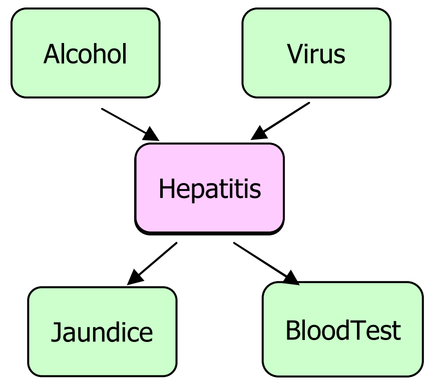


Local Markov Assumption:
A variable X is independent of its non-descendants given its parents
$(X_i \perp NonDescendants_{Xi} \mid Pa_{Xi})$

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

Belief Nets

P(A=1)
0.32



P(V=1)
0.20

a	v	P(H=1|a ,v)
1	1	0.82
1	0	0.10
0	1	0.45
0	0	0.04

h	P(J=1| h)
1	0.8
0	0.3

h	P(B=1| h)
1	0.98
0	0.01

- DAG structure
  - Each node ≡ Variable *v*
  - *v* depends (only) on its parents

  \+ conditional prob: $P(v_i \mid \text{parent}_i = \langle 0,1,\ldots \rangle)$

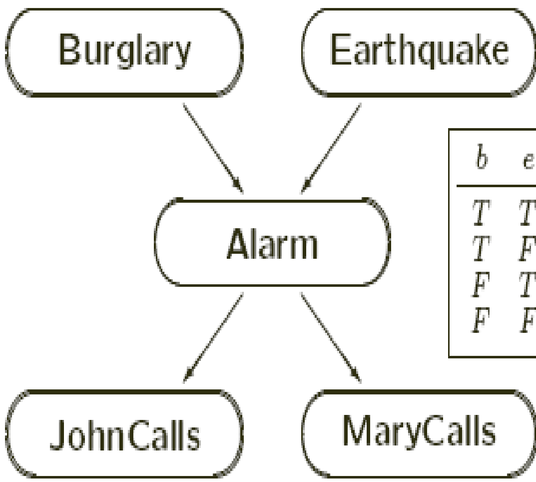
- *v is INDEPENDENT* of non-descendants, given assignments to its parents

- Given H = 1,
  - A has no influence on J
  - J has no influence on B
  - etc.

What about probabilities? Conditional probability tables (CPTs)

$P(B)$
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

- Each CPTable is called a “Factor”

Factoid…

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

  $= P(A \mid B,C)\ P(B|C)\ P(C)$
- In general:

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

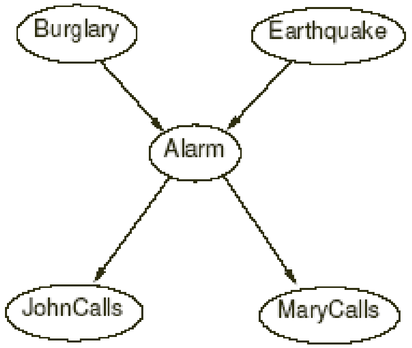
  $P(X_1 \mid X_2, \ldots, X_m)\ P(X_2, \ldots, X_m) =$

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

  $=$

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

Joint Distribution

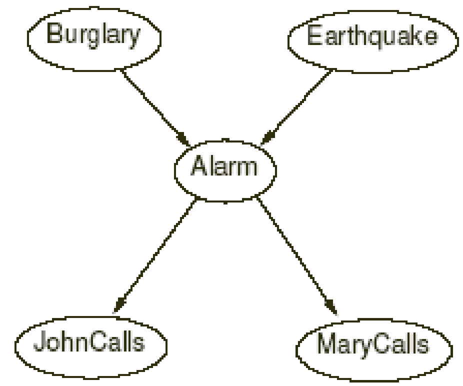


- In gen'l, $P(X_1, X_2, \ldots, X_m) =$
  $P(X_1 | X_2, \ldots, X_m)\, P(X_2, \ldots, X_m) =$
  $P(X_1 | X_2, \ldots, X_m)\, P(X_2 | X_3, \ldots, X_m)\, P(X_3, \ldots, X_m) =$
  $\prod_i P(X_i | X_{i+1}, \ldots, X_m)$

- Independence means.
  $P(X_i | X_{i+1}, \ldots, X_m) = P(X_i \mid \text{Parents}(X_i))$
  Node independent of predecessors, given parents

- So... $P(X_1, X_2, \ldots, X_m) = \prod_i P(X_i \mid \text{Parents}(X_i))$

Joint Distribution



Node is INDEPENDENT of non-descendants, given assignments to its parents

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

$$= \cancel{P(+j \mid +m, +a, -b, -e)} \xrightarrow{J \perp \{M,B,E\} \mid A} P(+j \mid +a)$$

$$\cancel{P(+m \mid +a, -b, -e)} \xrightarrow{M \perp \{B,E\} \mid A} P(+m \mid +a)$$

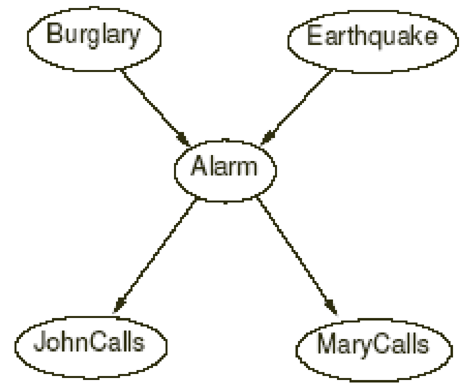
$$\cancel{P(+a \mid -b, -e)} \longrightarrow P(+a \mid -b, -e)$$

$$\cancel{P(-b \mid -e)} \xrightarrow{B \perp E} P(-b)$$

$$\cancel{P(-e)} \longrightarrow P(-e)$$

Joint Distribution

Node is INDEPENDENT of non-descendants, given assignments to its parents



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

= P(+j | +a)

P(+m | +a)

P(+a| -b, -e)

P(-b)

P(-e)

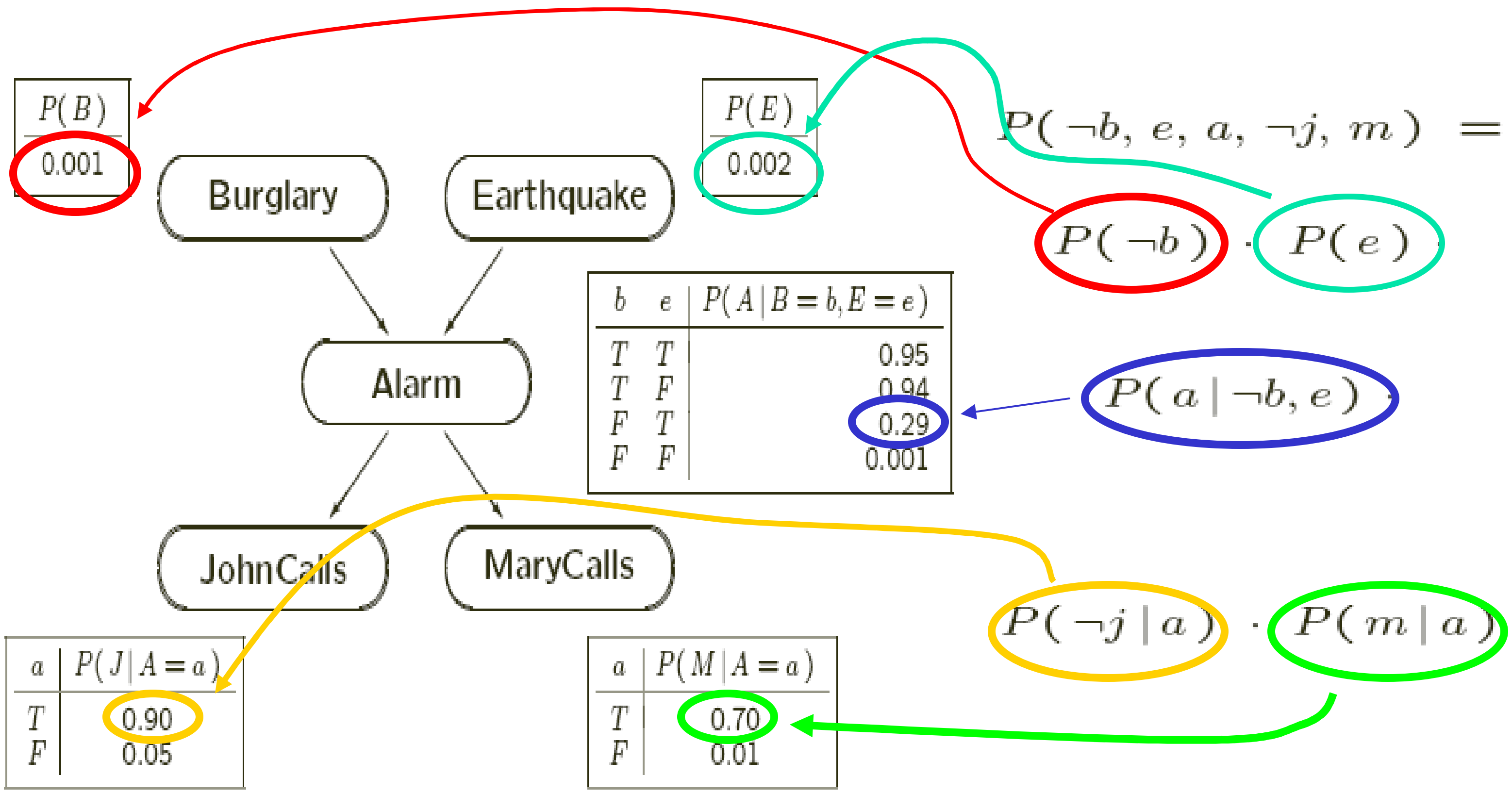
Recovering Joint

$$P(\neg b, e, a, \neg j, m) = P(\neg b)\, P(e \mid \neg b)\, P(a \mid e, \neg b)\, P(\neg j \mid a, e, \neg b)\, P(m \mid \neg j, a, e, \neg b)$$

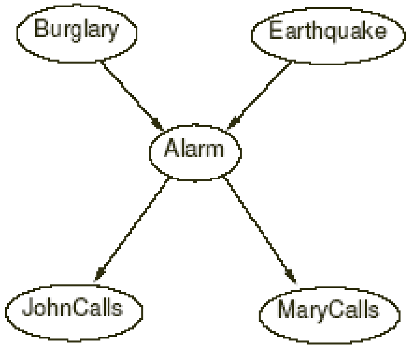
$$= P(\neg b)\, P(e)\, P(a \mid e, \neg b)\, P(\neg j \mid a)\, P(m \mid a)$$

$$= 0.99 \times 0.02 \times 0.29 \times 0.1 \times 0.70$$

Node independent of predecessors, given parents

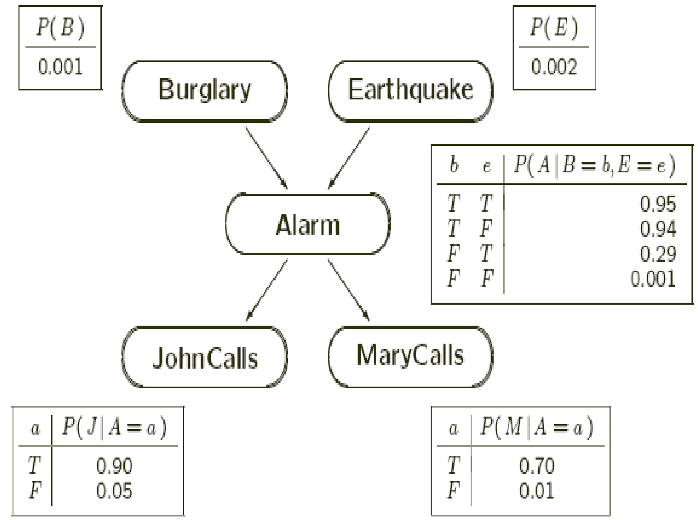


Meaning of Belief Net



- A BN represents
  - joint distribution
  - condition independence statements
- P(+j, +m, +a, -b, -e)
  = P(-b) P(-e) P(+a|-b, -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, $P(X_1, X_2, \ldots, X_m) = \prod_i P(X_i \mid X_{i+1}, \ldots, X_m)$
- Independence means
  $P(X_i \mid X_{i+1}, \ldots, X_m) = P(X_i \mid \text{Parents}(X_i))$
  Node independent of predecessors, given parents
- So... $P(X_1, X_2, \ldots, X_m) = \prod_i P(X_i \mid \text{Parents}(X_i))$

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) :

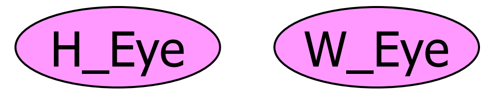
Get joint, then marginalize, conditionalize, ...

∃ better ways. . .

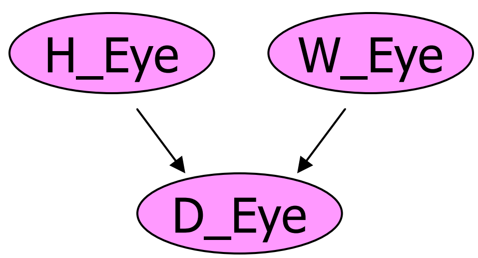
- Note: Given structure, ANY CPT is consistent.

  ∄ redundancies in BN. . .

"V"-Connections



- What color 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 BROWN eyes
  Now do you want to know my eye_color?

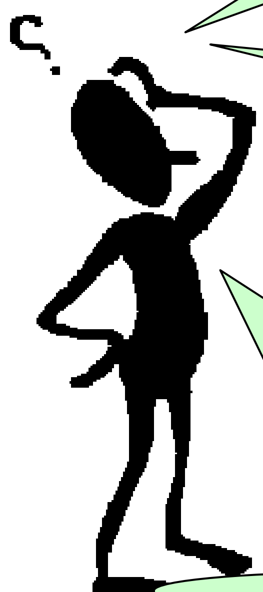


h	w	P(D= bl | h , w)
bl	bl	1.0
bl	br	0.5
br	bl	0.5
br	br	0.25

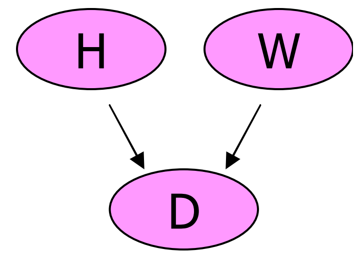
- H_Eye and W_Eye became dependent!

What color is W?

Prior is $P(W = br) = 0.8$?



But I know H!
Should I tell you?



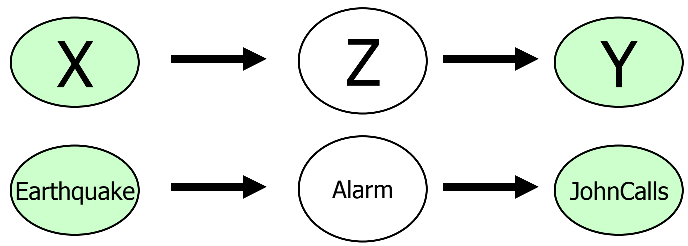
Don't bother; it doesn't matter
$P(W = br \mid H = bl) = 0.8$
$P(W = br \mid H = br) = 0.8$

I also know $D = br$. Now do you care?

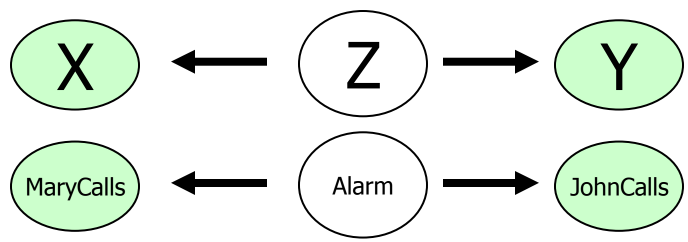
Yes, yes!!! Tell me H!
$P(W = br \mid H = bl, D = br) = 0.50$
$P(W = br \mid H = br, D = br) = 0.22$

d-separation Conditions

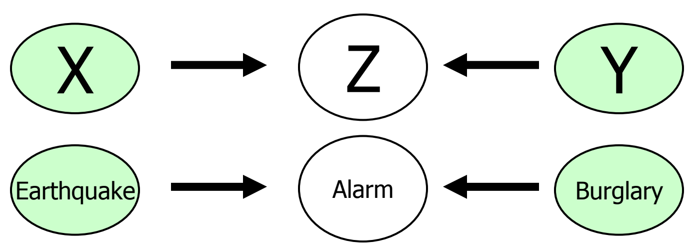
$\neg(X \perp Y)$



$\neg(X \perp Y)$

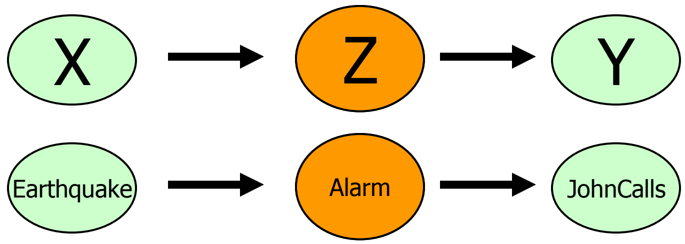


$X \perp Y$



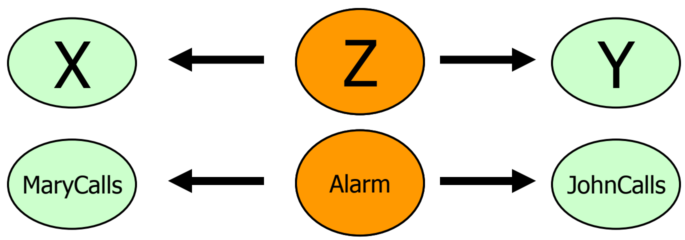
d-separation Conditions

$\neg(X \perp Y)$



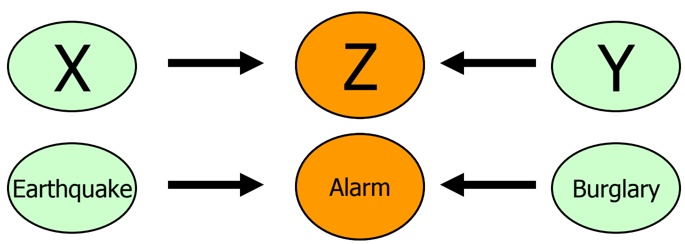
$X \perp Y \mid Z$

$\neg(X \perp Y)$



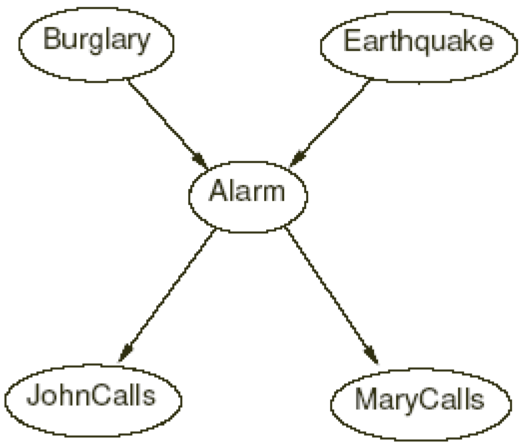
$X \perp Y \mid Z$

$X \perp Y$



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

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

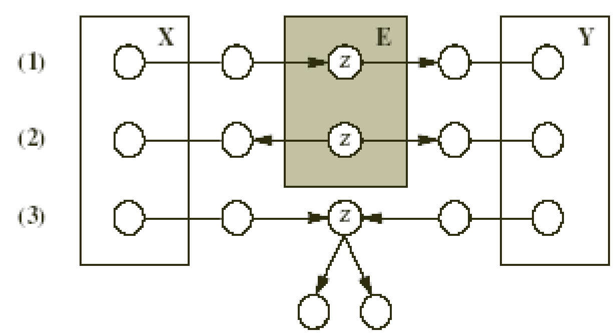
Conditional Independence

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

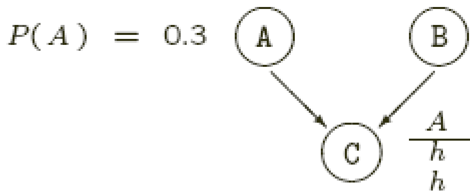
- **General** question: "X ⊥ Y | **E**"
  - Are nodes **X** independent of nodes **Y**, given assignments to (evidence) nodes **E**?

- **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**



"V"-Connections, con't



A	B	$P(C \mid A, B)$
h	h	0.0
h	t	1.0
t	h	1.0
t	t	0.0

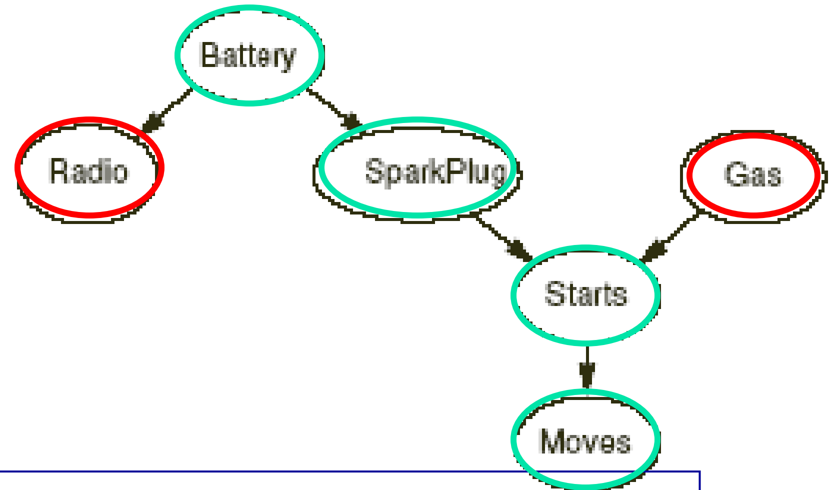
- $A \perp B \mid \{\}$
  $P(A) = 0.3 = P(A \mid B)$
- But: $\neg[A \perp B \mid C]$
  $P(A \mid B) = 0.3$; $P(A \mid B, C) = 0$

- Proof:
  - $P(+a, +b, +c) = P(+a)\,P(+b)\,P(+c \mid +a, +b) = 0.3 \times 0.5 \times 0 = 0$
  - $P(\neg a, +b, +c) = P(\neg a)\,P(+b)\,P(+c \mid \neg a, +b) = 0.7 \times 0.5 \times 1 = 0.35$
  - $P(+b, +c) = P(+a, +b, +c) + P(\neg a, +b, +c) = 0 + 0.35 = 0.35$
  - $P(+a \mid +b, +c) = P(+a, +b, +c) / P(+b, +c) = 0/035 = 0$

- P(Cold | Sneeze) vs P(Cold | Sneeze, Purr)

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 . . .

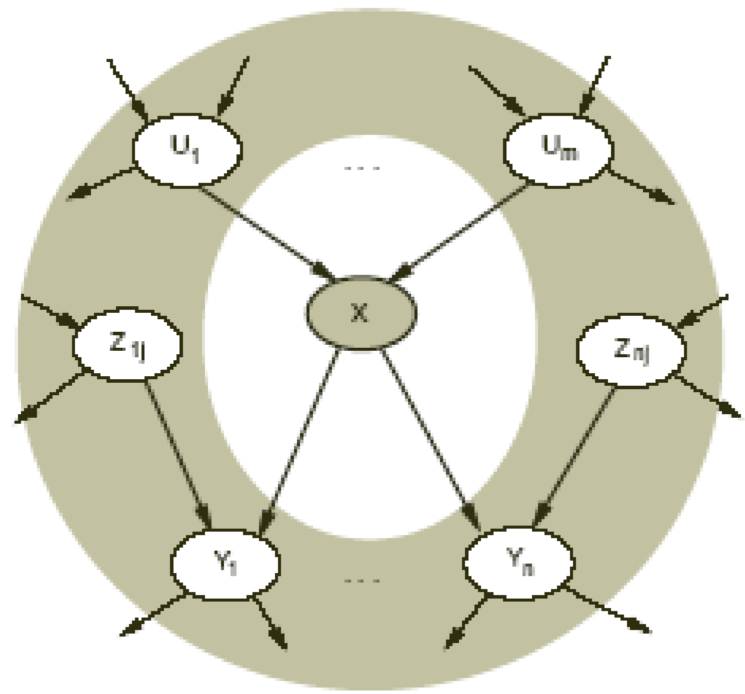
1. **E** = {} ?
   YES: P(R | G) = P(R)
   Starts ∉ **E**, and Starts has 2 in-links
2. **E** = Starts ?
   NO!! P(R | G, S) ≠ P(R| S)
   Starts ∈ **E**, and Starts has 2 in-links
3. **E** = Moves ?
   NO!! P(R | G, M) ≠ P(R| M)
   Moves ∈ **E**, Moves child-of Starts, and Starts has 2 in-links (on path)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

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



Outline

- Motivation
- What is a Belief Net?
  - Example
  - Inference
  - Semantics
  - Applications
  - Relation to other Models
- Learning a Belief Net

Deployed Applications

- 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

Deployed Applications (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

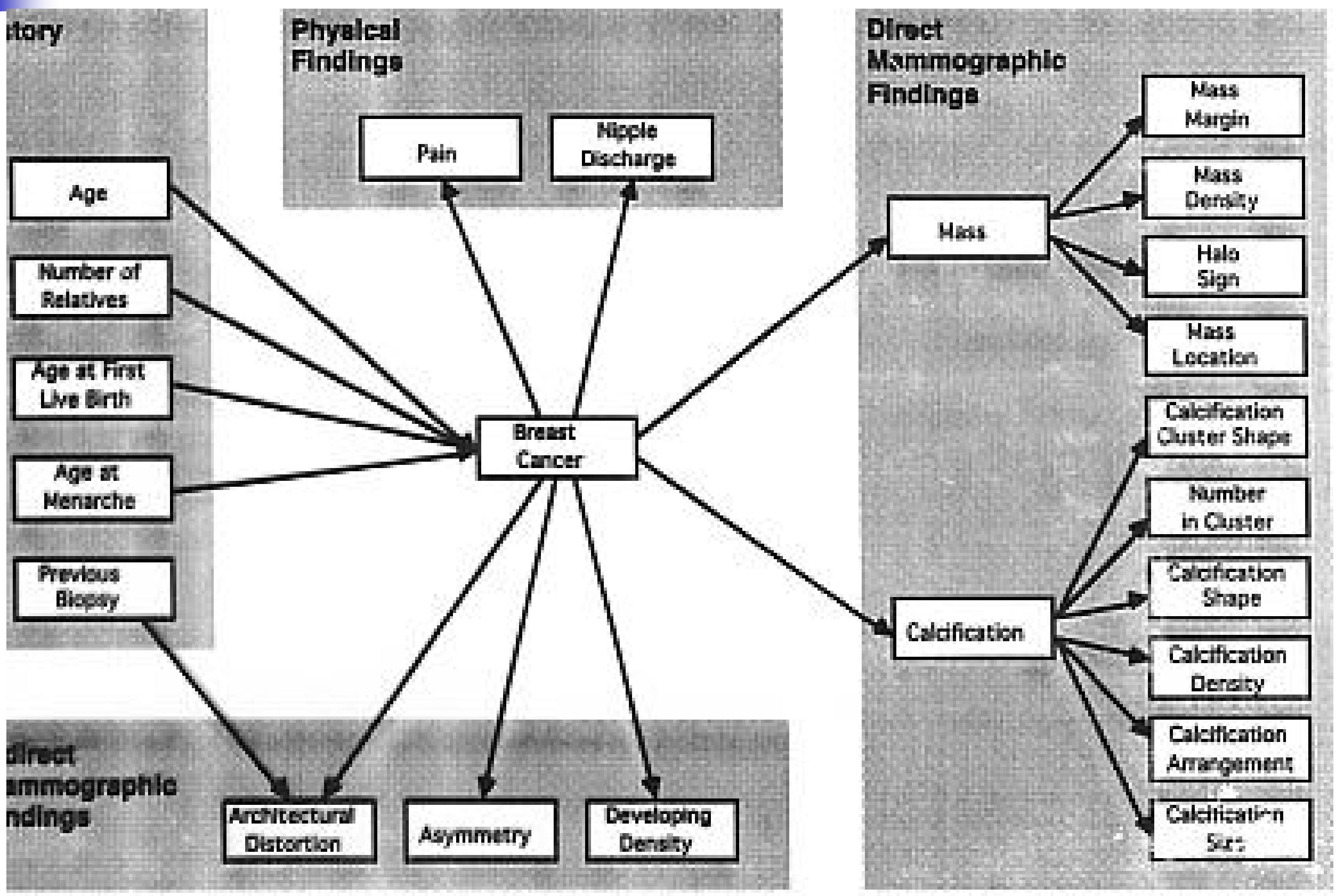
Deployed Applications (III)

- Speech recognition
- Human genome analysis
- Robot mapping
- Identify meteorites to study
- Modeling fMRI data
- Anomaly detection
- Fault diagnosis
- Modeling sensor network data

Deployed Applications (IV)

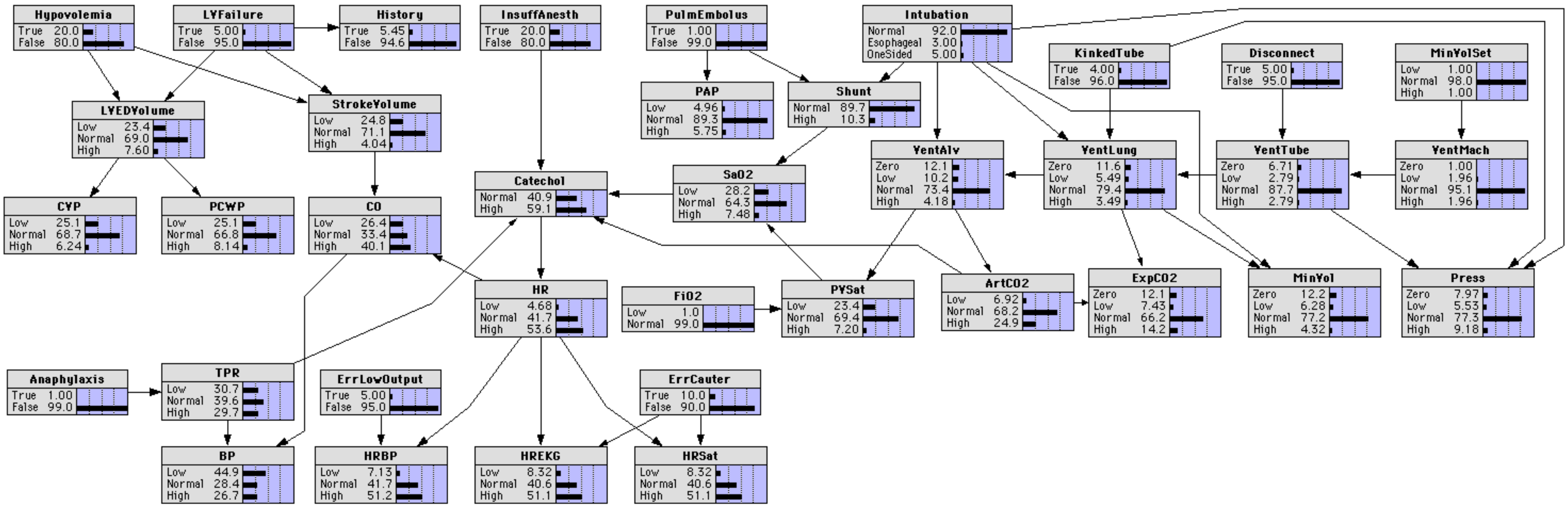
- Lymph-node pathology diagnosis
- Manufacturing control
- Software diagnosis
- Information retrieval
- *Types of tasks*
  - *Classification/Regression*
  - *Sensor Fusion*
  - *Prediction/Forecasting*

MammoNet



tory
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 Cluster Shape
Number in Cluster
Calcification Shape
Calcification
Calcification Density
Calcification Arrangement
direct
ammographic
ndings
Architectural Distortion
Asymmetry
Developing Density

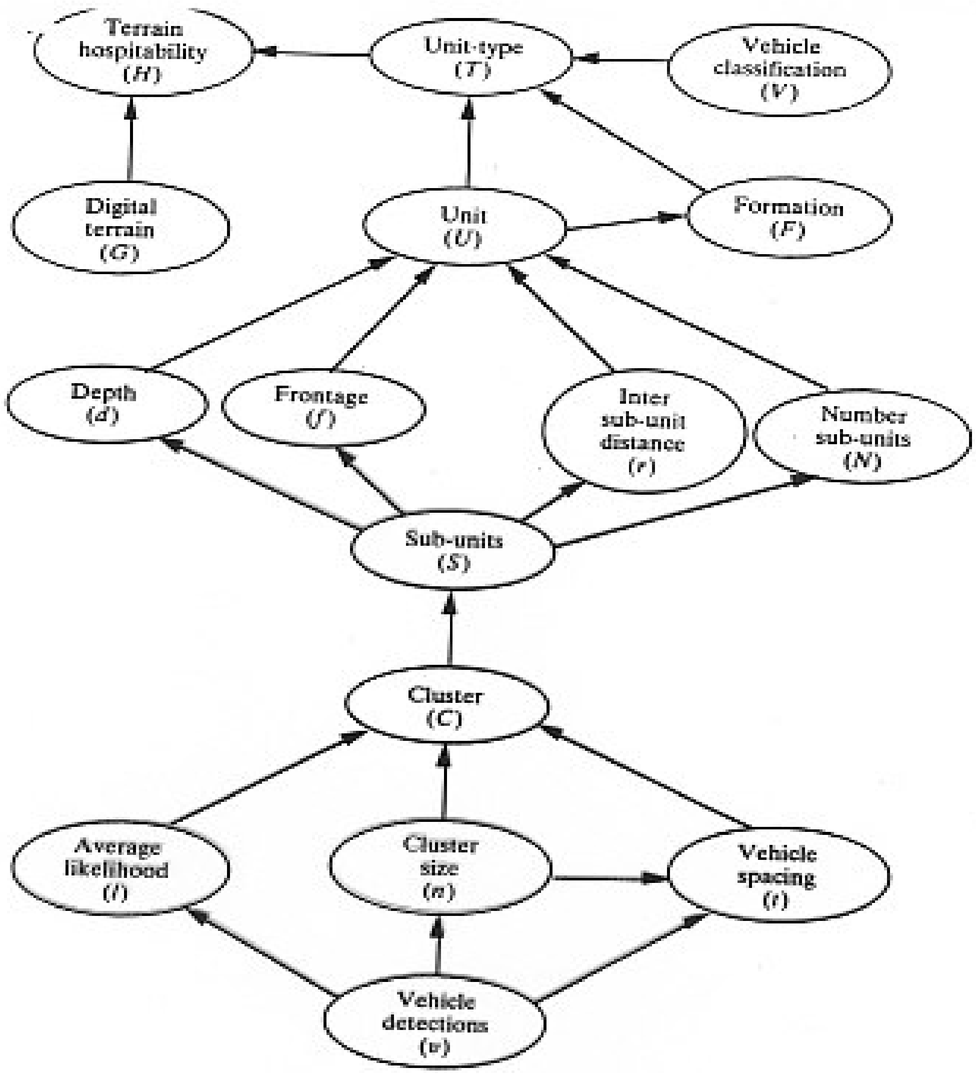
ALARM



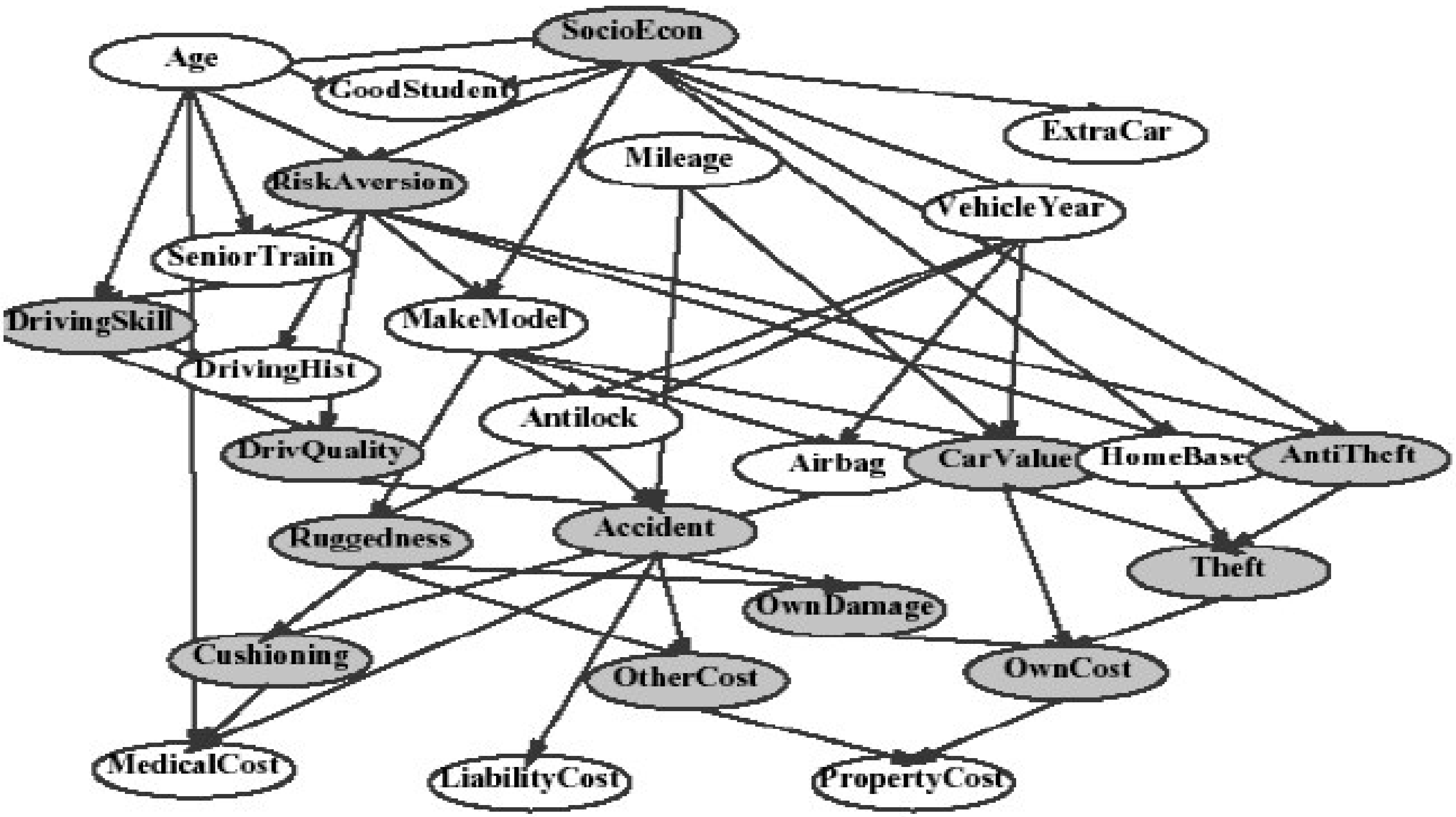
A Logical Alarm Reduction Mechanism

- 8 diagnoses, 16 findings, …

Troup Detection



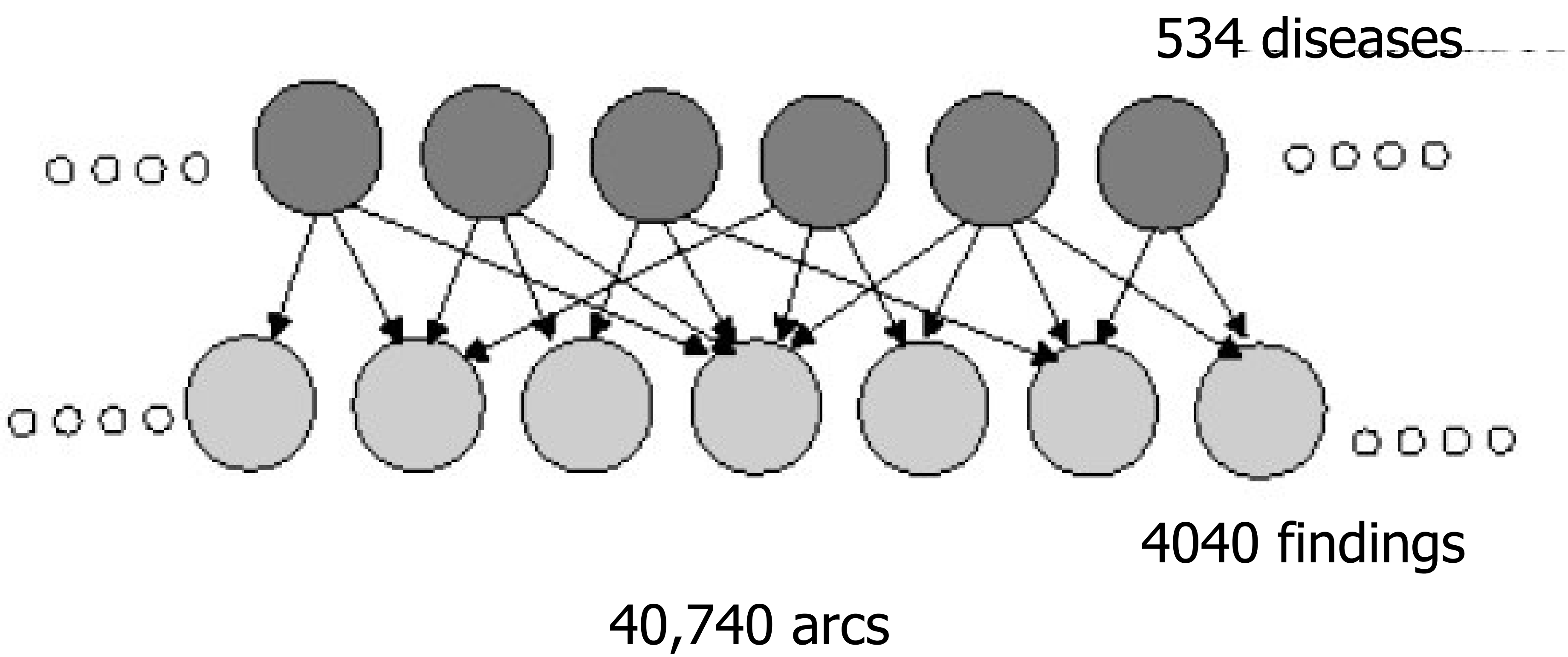
Car Insurance



Predict claim costs (medical, liability) based on application data

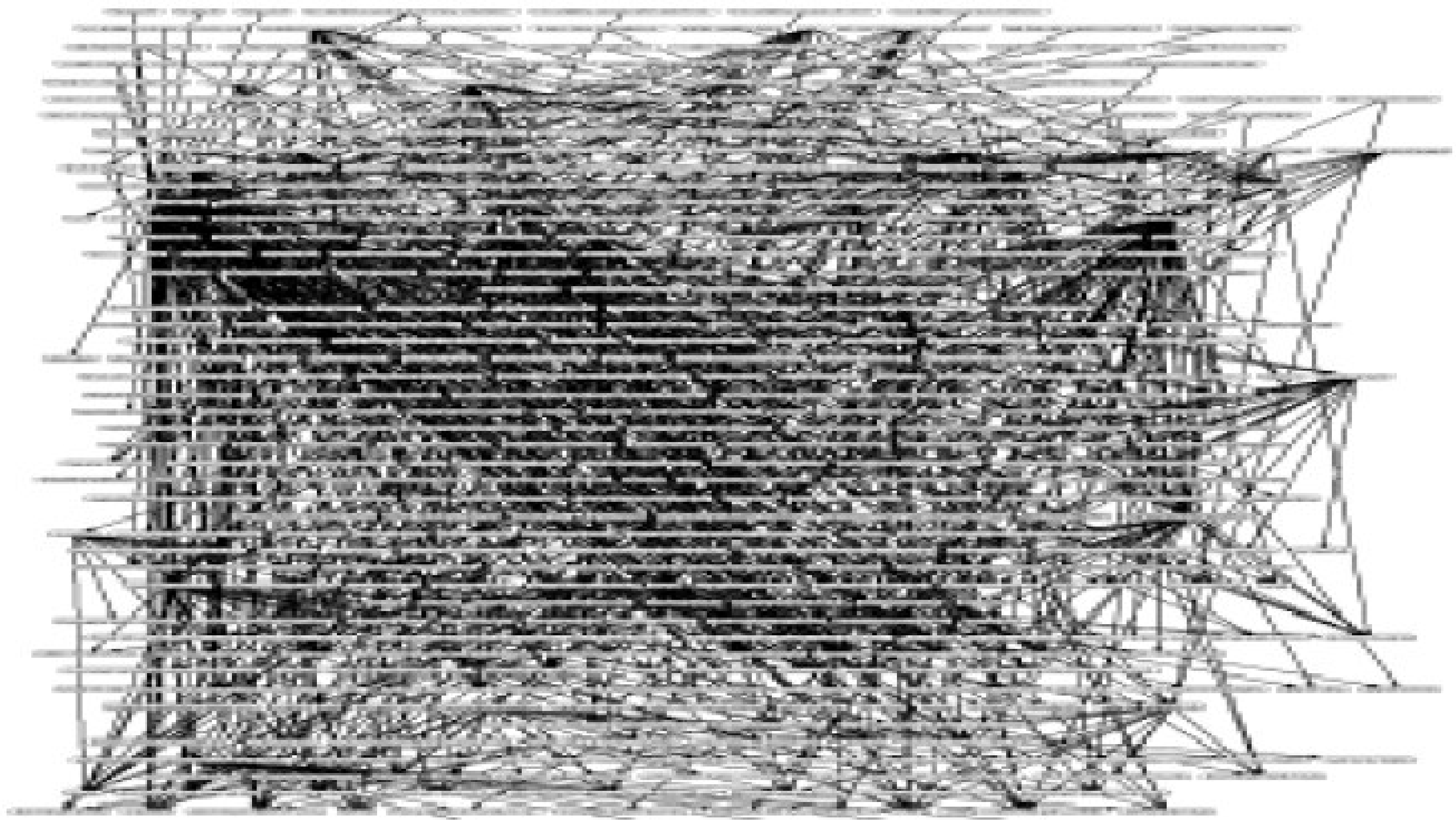
QMR-DT

- **Medical diagnosis in internal medicine**
- Bipartite network of disease/findings relations

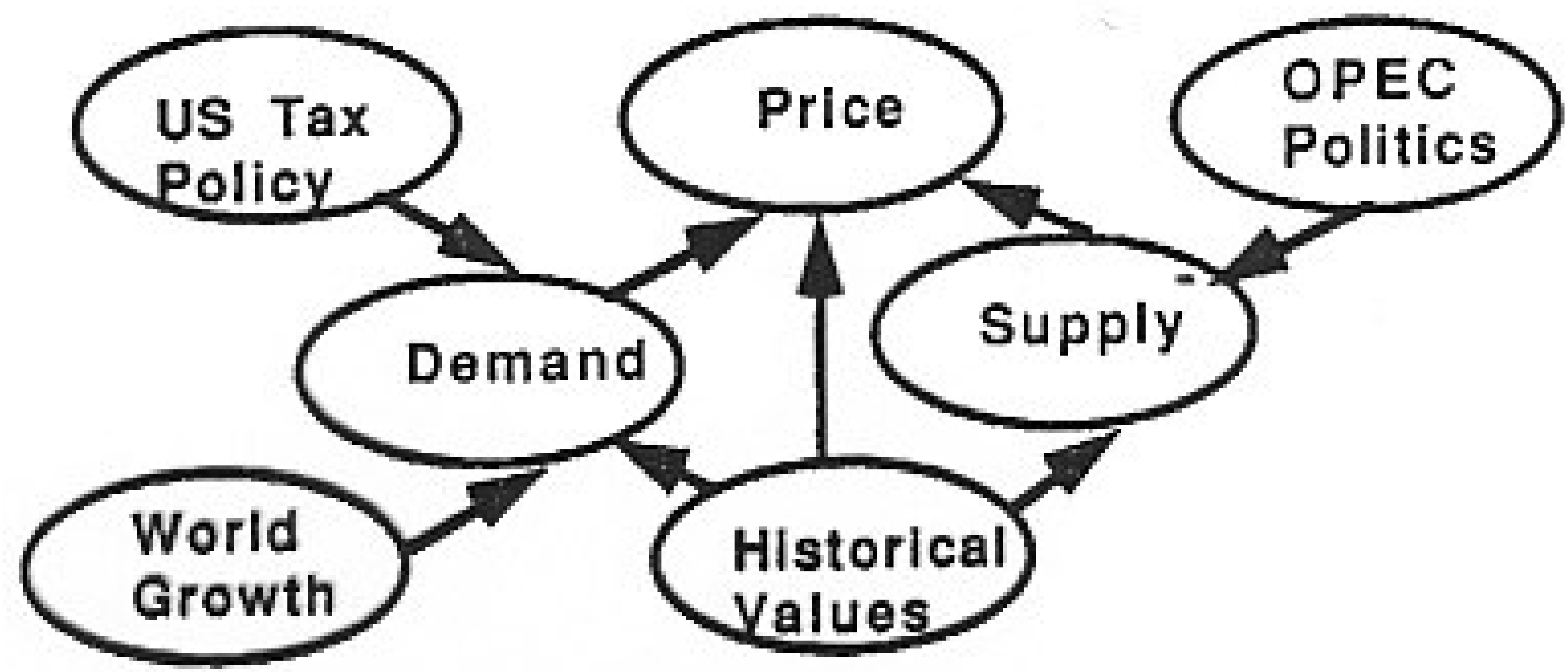


CPCS

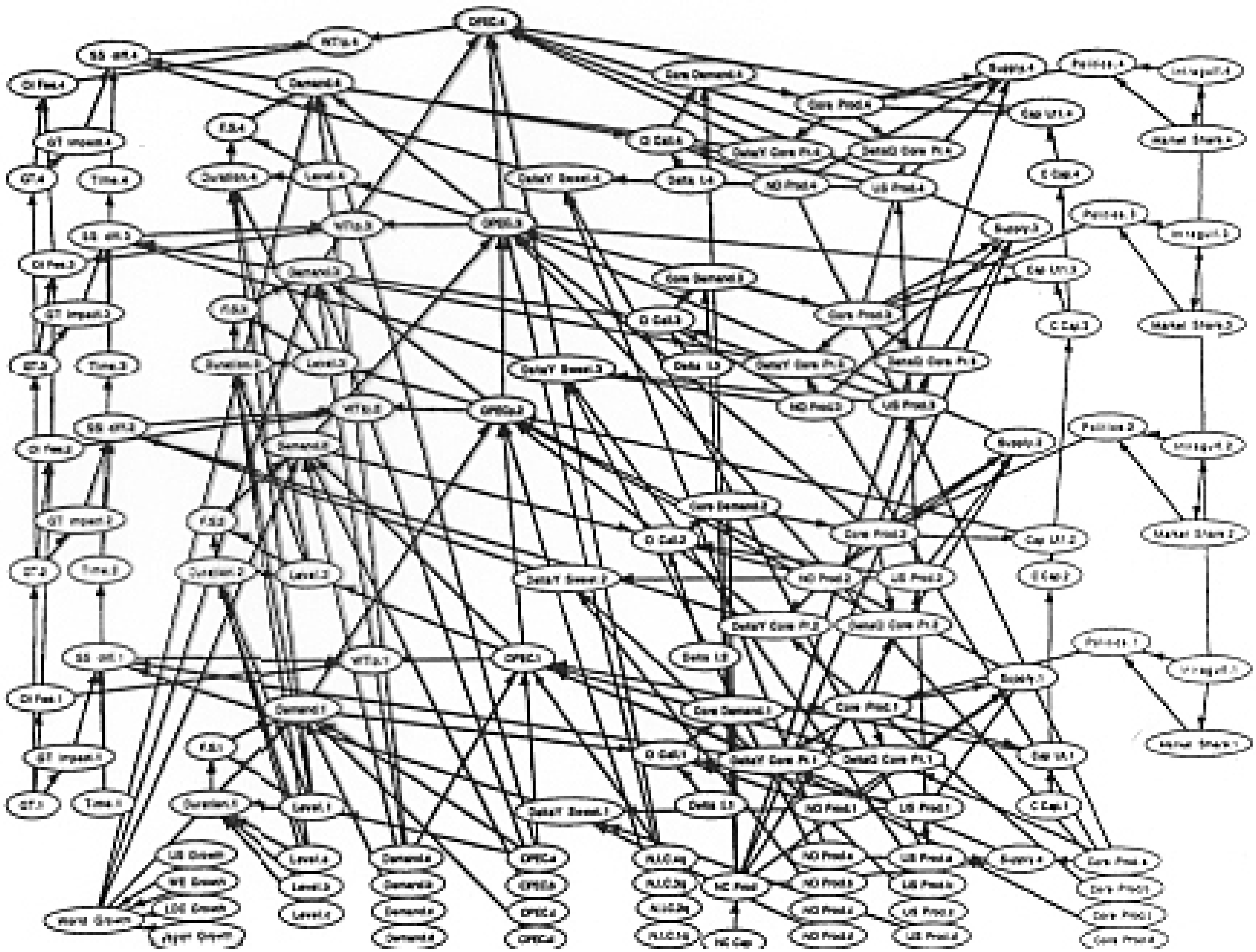
- **C**omputer-based **P**atient **C**ase **S**imulation system
- 422 nodes; 867 arcs



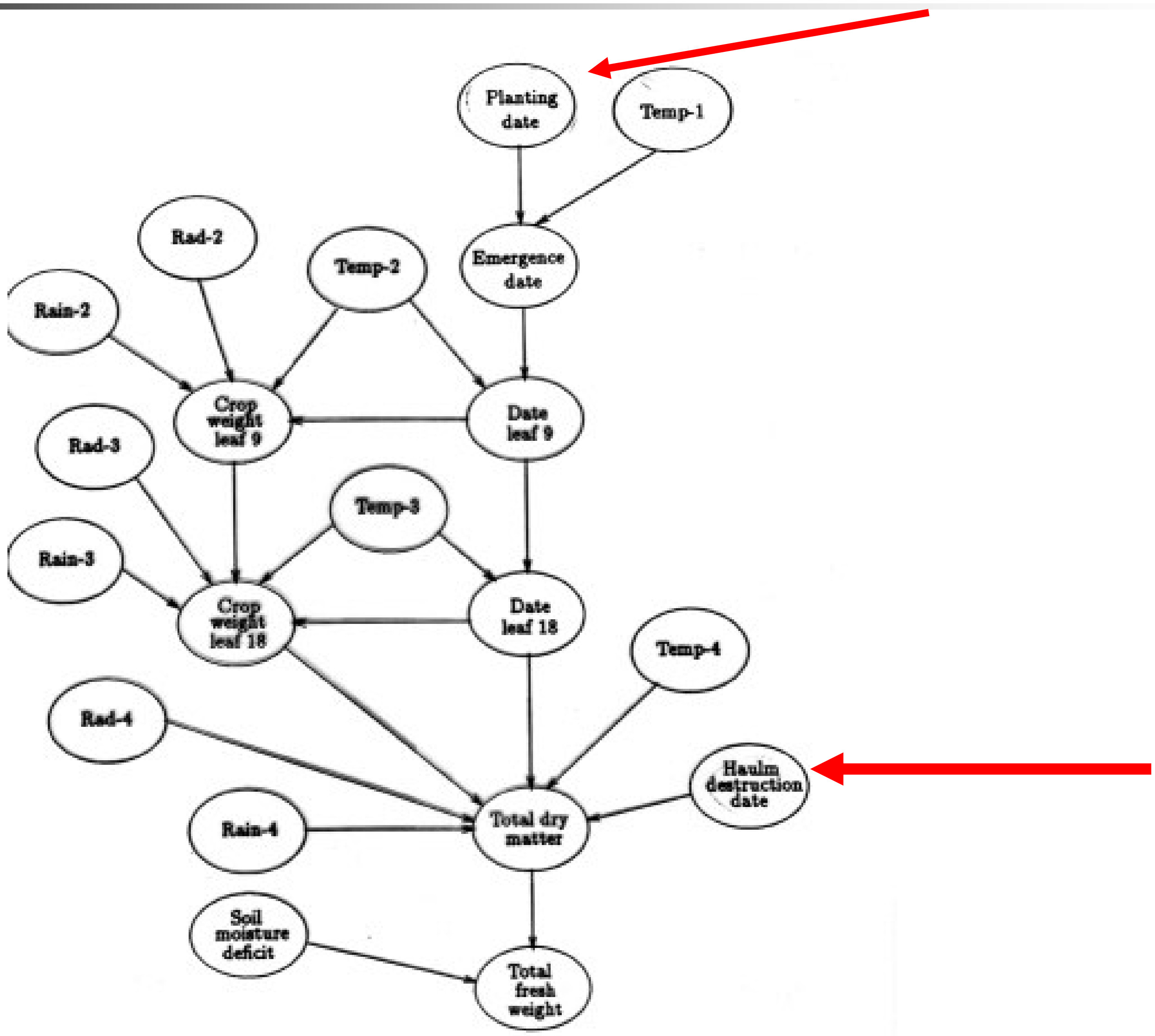
ARCO1: Forecasting Oil Prices



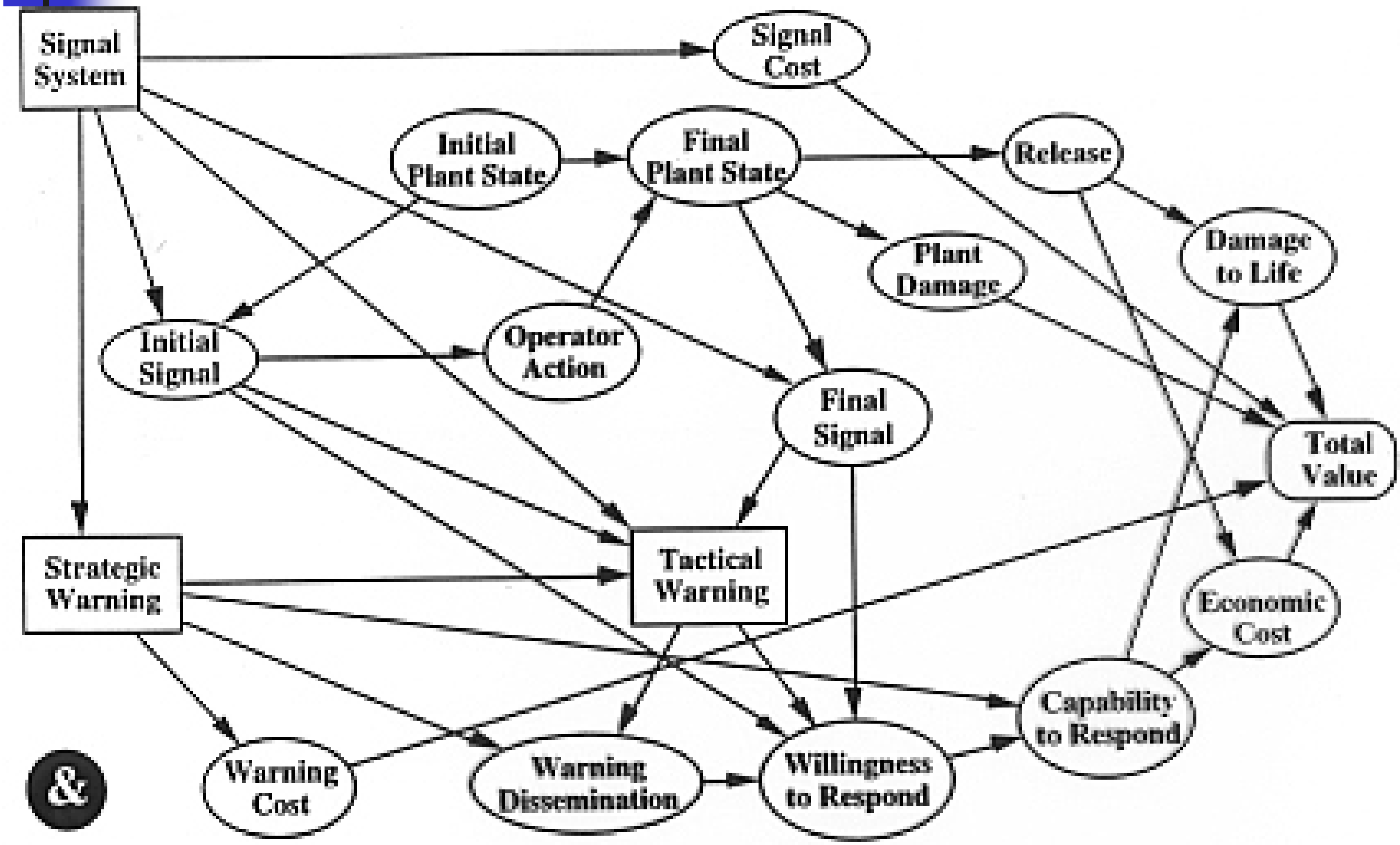
ARCO1: Forecasting Oil Prices



Forecasting Potato Production



Warning System



Utility-Based Agents

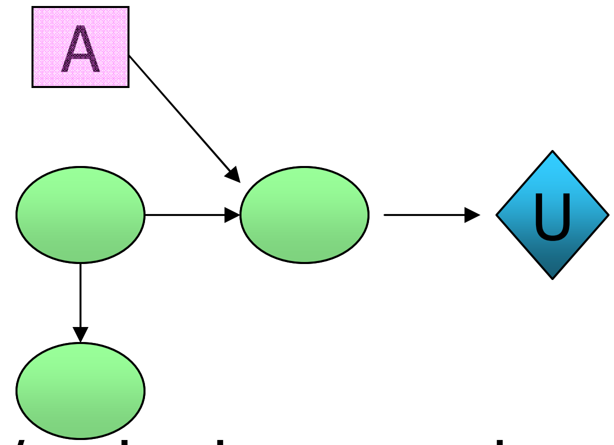
- MEU Principle:
  Agent should act to maximize expected utility
- Choose action that maximizes
  expected utility of state after A,
  given prior observations O:

$$A^* = \text{argmax}_A \{ EU(A|O) \}$$

$$EU(A \mid O) =$$
$$= \sum_{S'} P(S'|A,O)\ U(S')$$
$$= \sum_{S'} \sum_{S} P(S \mid O)\ P(S' \mid S,A)\ U(S')$$
$$= \sum_{S'} \sum_{S} [\alpha\, P(O \mid S)\ P(S)]\ \ P(S' \mid S,A)\ U(S')$$

- Given simple assumptions, this is best possible action!
  (Average of utility, not of ~~utility^2~~, not ~~minimaxing...~~)
- Good decision, bad outcome.

Decision Network



- Chance Nodes: S, O, S′
  - Bayesian net ≡ decision diagram w/ only chance nodes
  - Specify: P(S), P(O | S), P(S′ | S, A)
  - Here: S ≡ Current State   O ≡ Observation
    S′ ≡ Resulting State
- Decision Nodes: A
  - represents decision/action to make.
  - Specify: set of possible actions a ∈ Dom(A)
- Utility Node(s): U
  - represents utility of each value-set of its parent chance variables
  - Specify: set of U(s′) for each s′ ∈ Dom(S′)

Perform a Medical Treatment?

- EU(T = 1) =
  $\sum_r P(R = r \mid T = 1)\, U(R = r)$

  EU(T = 0) =
  $\sum_r P(R = r \mid T = 0)\, U(R = r)$

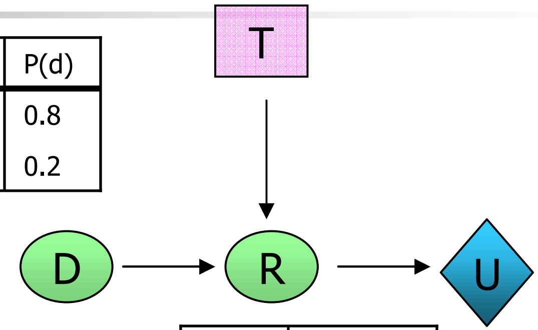
- P(R = 1 | T = 1) =
  $\sum_d \; P(R = 1, D = d \mid T = 1)$

  $= \sum_d P(R = 1 \mid D = d, T = 1)\, P(D = d)$

  $= P(R = 1 \mid D = 0, T = 1)\, P(D = 0) \;+ P(R = 1 \mid D = 1, T = 1)\, P(D = 1)$
  $= \quad (0.001 \times 0.8) + (0.01 \times 0.2) = 0.0028$

- $P(R = 0 \mid T = 1) = \; 1 - P(R = 1 \mid T = 1) \; = \; 0.9972$
- Similarly:
  - $P(R = 1 \mid T = 0) = 0.1908$
  - $P(R = 0 \mid T = 0) = 0.8092$

d	P(d)
0	0.8
1	0.2



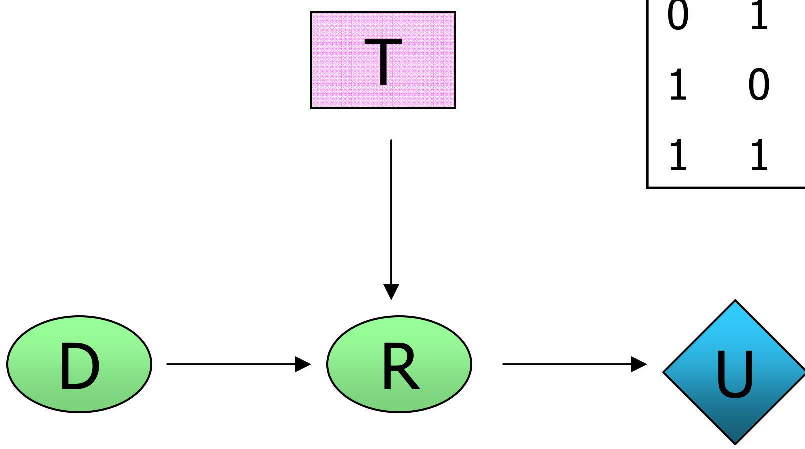
d	t	P(+r|d,t)
0	0	.001
0	1	.001
1	0	.950
1	1	.010

r	U(r)
0	0
1	-1000

Medical Treatment (con't)

d	t	P(+r|d,t)
0	0	.001
0	1	.001
1	0	.950
1	1	.010

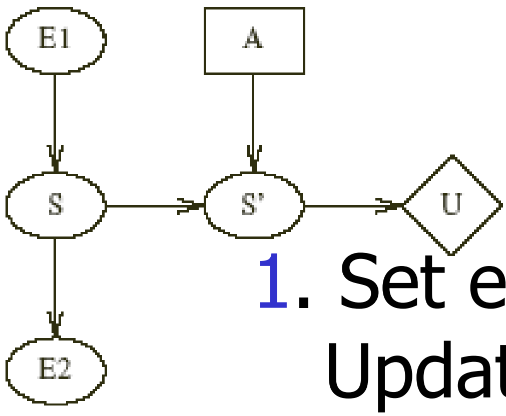
d	P(d)
0	0.8
1	0.2



r	U(r)
0	0
1	-1000

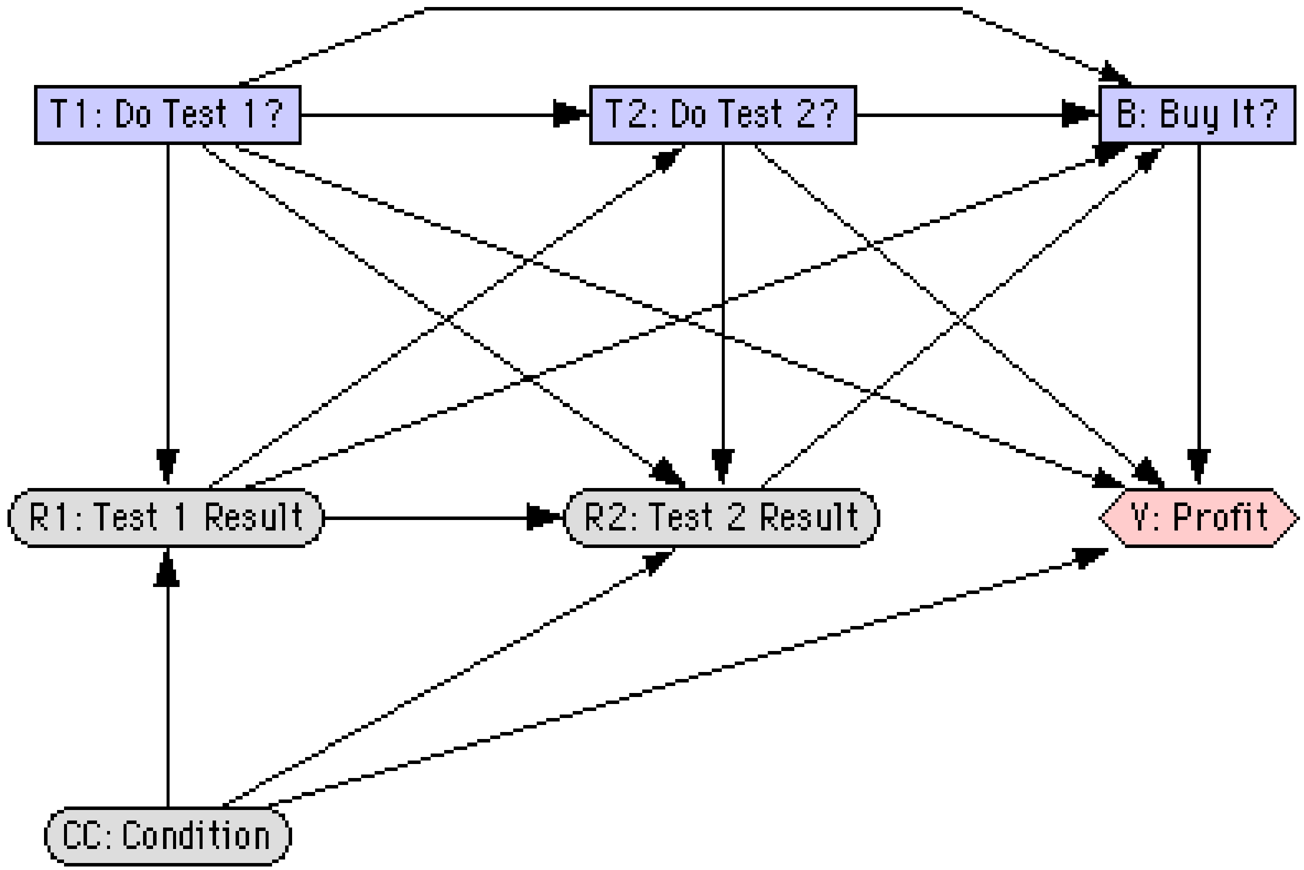
T	$P(R\|T)$ 0	$P(R\|T)$ 1	$U(R)$ 0	$U(R)$ 1	$EU(T)$
0	.8092	.1908	0	−1000	−190.8
1	.9972	.0028	0	−1000	−2.8 ⇐ chosen action

Evaluating a Decision Network



1. Set evidence variables E_1, E_2
   Update distribution over current state S
2. For each possible action a of decision node A
   (a) Set decision node A to a
   (b) For each parent $\{ S' \}$ of utility node U:
       Calculate posterior probability of S
       Here, just $P(S' \mid E_1, E_2, A = a)$
   (c) Calculate expected utility for action a:
       $EU(A \mid E_1, E_2) = \sum_{S'} P(S' \mid E_1, E_2, a) \, U(S')$
3. Choose action $a^* = \arg\max_a \{ EU(a \mid \ldots) \}$
   with highest expected utility

Decision Net: Test/Buy a Car



Extensions

- Find best values (posterior distr.) for
  SEVERAL (> 1) "output" variables
- *Partial* specification of "input" values
  - only subset of variables
  - only "distribution" of each input variable
- General Variables
  - Discrete, but domain > 2
  - Continuous (Gaussian: $x = \Sigma_i b_i y_i$ for parents $\{Y\}$)
- Decision Theory ⇒ *Decision Nets* (Influence Diagrams)
  Making Decisions, not just assigning prob's
- Storing $P(v \mid p_1, p_2,\ldots,p_k)$
  General "CP Tables" $0(2^k)$
  Noisy-Or, Noisy-And, Noisy-Max
  "Decision Trees"

Outline

- Motivation
- What is a Belief Net?
  - Example
  - Inference
  - Semantics
  - Applications
  - Relation to other Models
- Learning a Belief Net

Belief Nets vs Rules

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

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

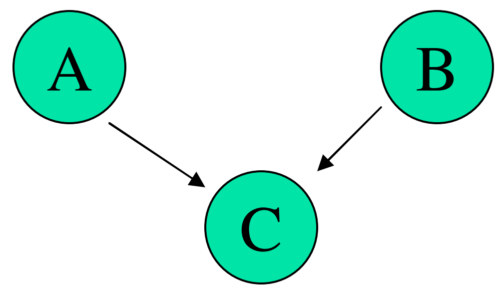
  to FACTOR a distribution

  … explicitly specify dependencies, implicitly independencies…

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

Technical: BNs use DIRECTRED arcs

$\Rightarrow$ allow "induced dependencies"

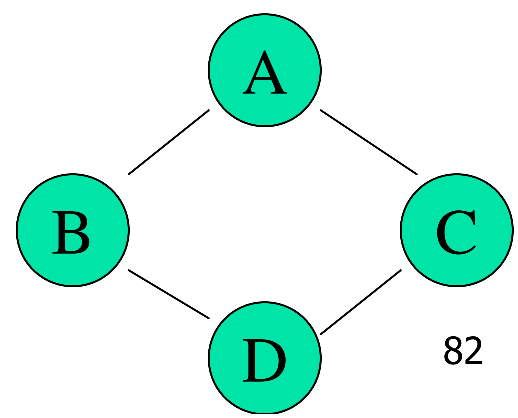


$I\,(A, \{\}, B)$ "*A* independent of *B*, given { }"

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

MNs use UNDIRECTED arcs

$\Rightarrow$ allow other independencies

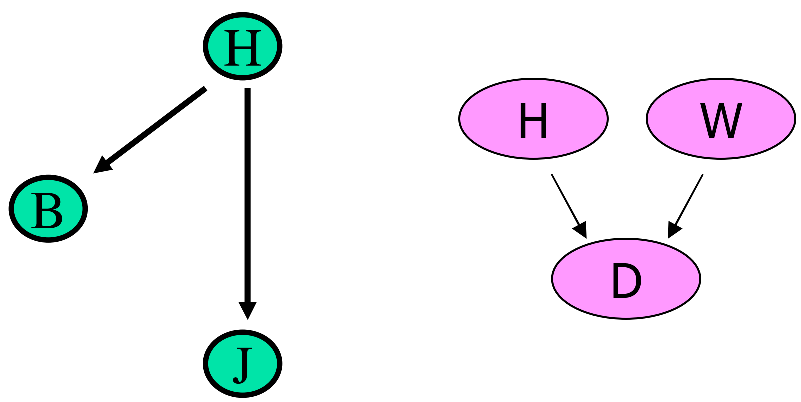


$I(A, BC, D)$ *A* independent of *D*, given *B*, *C*

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

Belief Nets vs Clusters

- Both "structure" the variables
  - Cluster: Put *similar* variables in same cluster
  - BN: Put *related* variables adjacent
- Cluster uses "first order" relationships
  - Put A and B together if A *directly correlated with* B
- BN can have higher order relationships, esp. independencies

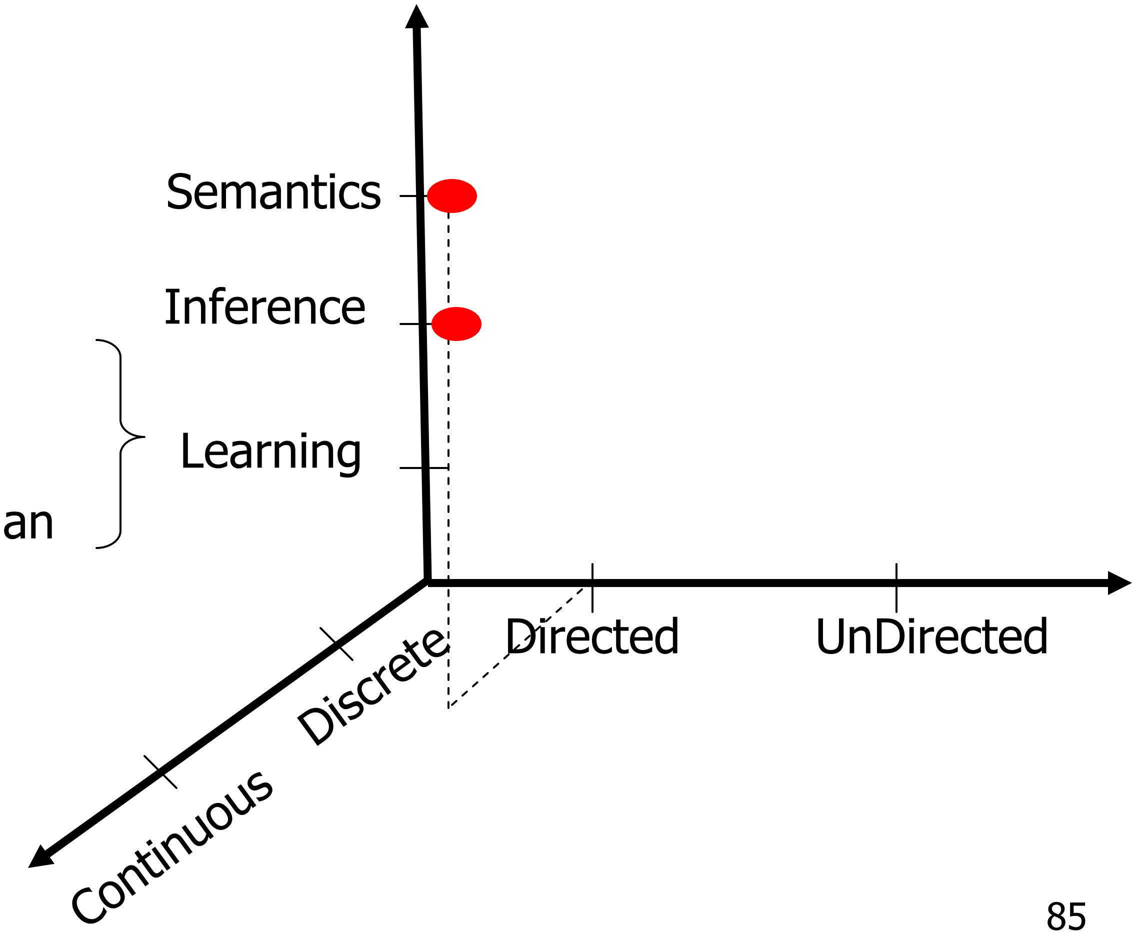


2^{nd} Order Statistics?

- Spse
  - ½ of kidney *donors* are Male (½ female)
  - ½ of kidney *recipients* are Male (½ female)
  - Transplant is SUCCCESSFUL iff
    Donor and Recipient are SAME gender (M/M or F/F)
- Here:
  - P(Success | Donor=m) = ½ = P(Success | Donor=f)
    ⇒ Success is independent of Donor Gender
  - P(Success | Recip=m) = ½ = P(Success | Recip=f)
    ⇒ Success is **independent** of Recipient Gender
- However:
  - P(Success | Donor=m, Recip=f) = 0
    P(Success | Donor=m, Recip=m) = 1
  - So Success is **dependent** on
    Recipient Gender and Donor Gender

Space of Topics

Learning…
- Parameter, Structure
- Data: Complete, Missing
- Framework: Frequentist, Bayesian



Summary

- Necessary to use Probabilistic Representation
  - … use *connections*… just *some* connections
  - Factored Distribution

  $\Rightarrow$ Belief Nets
- Proven Technology
  - Lots of deployed applications
- Challenge: Learning them!