# CS 452 – PROBABILISTIC GRAPHICAL MODELS

Unit # 3

# WHAT IS BAYESIAN NETWORK?

A BN is a Directed Acyclic Graph (DAG) in which:

- A set of random variables makes up the nodes in the network.
- A set of directed links or arrows connects pairs of nodes.
- Each node has a conditional probability table that quantifies the effects the parents have on the node.

The intuitive meaning of an arrow from a parent to a child is that the parent directly influences the child.

The direction of this influence is often taken to represent casual influence.

These influences are quantified by conditional probabilities.

# EARTHQUAKE EXAMPLE (PEARL)

You have a new burglar alarm installed.

It is reliable about detecting burglary, but does not respond to minor earthquakes.

Two neighbors (John, Mary) promises to call you at work when they hear the alarm.

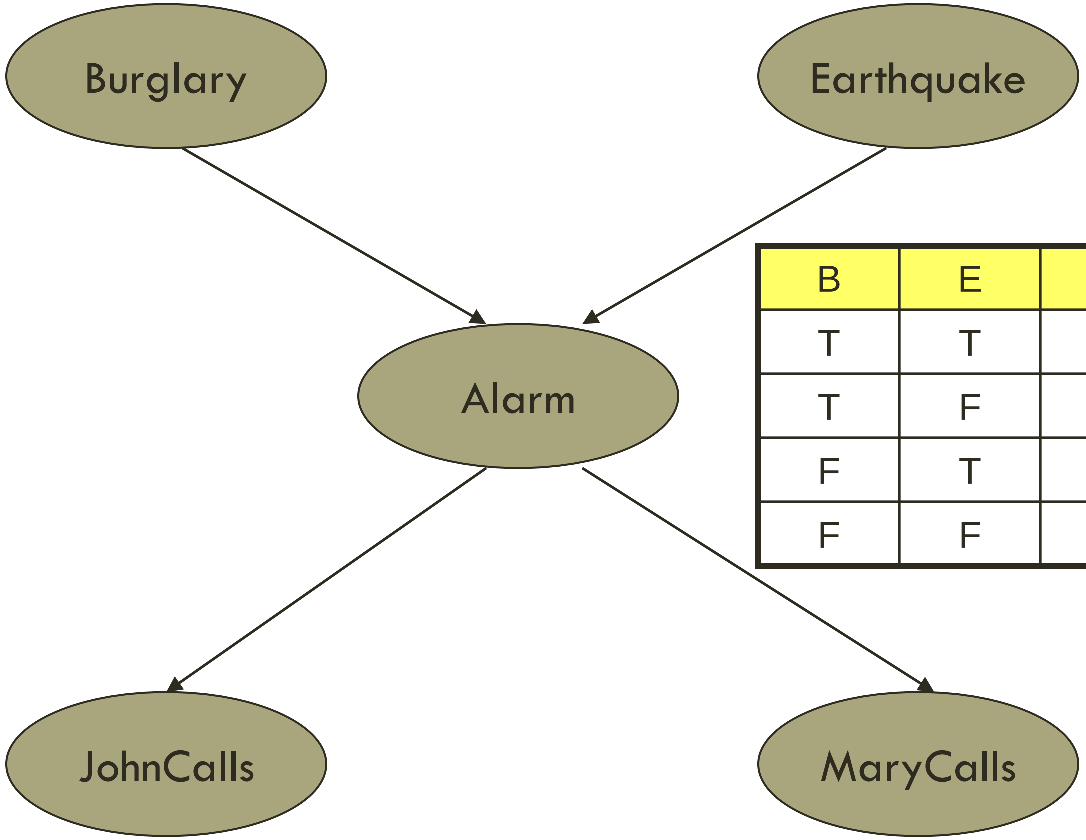
- John always call when hears alarm, but confuses alarm with phone ringing (and calls then also)
- Mary likes loud music and sometimes misses alarm

Given evidence about who has and hasn't called, estimate the probability of a burglary.

# Earthquake Example (Pearl)

| P(B) |
|---|
| 0.1 |

| P(E) |
|---|
| 0.02 |




| B | E | P(A\|B,E) |
|---|---|---|
| T | T | 0.95 |
| T | F | 0.94 |
| F | T | 0.29 |
| F | F | 0.001 |

| A | P(J\|A) |
|---|---|
| T | 0.9 |
| F | 0.05 |

| A | P(M\|A) |
|---|---|
| T | 0.7 |
| F | 0.01 |

# JOINT PROBABILITY TABLE FOR 'EARTHQUAKE' EXAMPLE

| B | E | A | J | M | P(B, E, A, J, M) |
|---|---|---|---|---|---|
| T | T | T | T | T | 0.0001197 |
| T | T | T | T | F | 0.0000513 |
| T | T | T | F | T | 0.0000133 |
| T | T | T | F | F | 0.0000057 |
| T | T | F | T | T | 5E-09 |
| T | T | F | T | F | 4.95E-07 |
| T | T | F | F | T | 9.5E-08 |
| T | T | F | F | F | 9.405E-06 |
| T | F | T | T | T | 0.0058036 |
| T | F | T | T | F | 0.0024872 |
| T | F | T | F | T | 0.0006448 |
| T | F | T | F | F | 0.0002764 |
| T | F | F | T | T | 2.94E-07 |
| T | F | F | T | F | 2.911E-05 |
| T | F | F | F | T | 5.586E-06 |
| T | F | F | F | F | 0.000553 |
| F | T | T | T | T | 0.0036175 |
| F | T | T | T | F | 0.0015503 |
| F | T | T | F | T | 0.0004019 |
| F | T | T | F | F | 0.0001723 |
| F | T | F | T | T | 7.029E-06 |
| F | T | F | T | F | 0.0006959 |
| F | T | F | F | T | 0.0001336 |
| F | T | F | F | F | 0.0132215 |
| F | F | T | T | T | 0.0006112 |
| F | F | T | T | F | 0.000262 |
| F | F | T | F | T | 6.791E-05 |
| F | F | T | F | F | 2.911E-05 |
| F | F | F | T | T | 0.0004846 |
| F | F | F | T | F | 0.0479769 |
| F | F | F | F | T | 0.0092077 |
| F | F | F | F | F | 0.9115606 |

# PARAMETERS OF A BN

A problem domain is modeled by a list of variables $X_1, X_2, \ldots, X_n$.

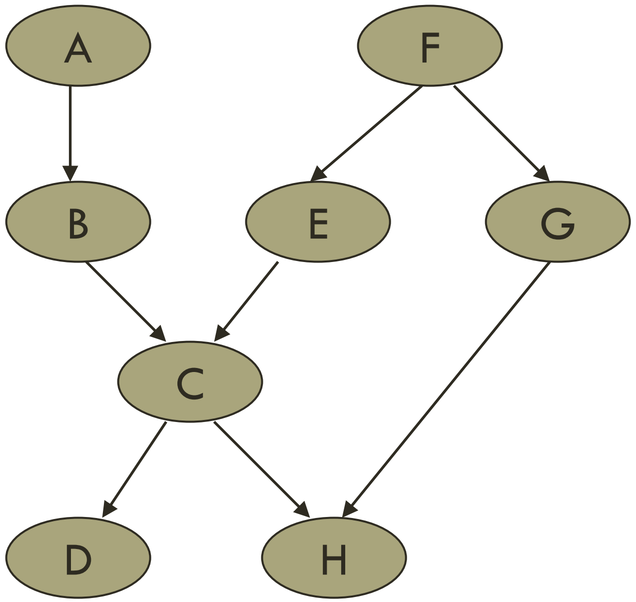
Knowledge about the problem domain is represented by a joint probability $P(X_1, X_2, \ldots, X_n)$.

General probability distribution of 8 variables with 2 states each has $2^8 = 256$ possible values and $2^8 - 1$ probabilities need to be specified.

Assumes that each node is conditionally independent of all its non-descendants given its parents.

Product of all conditional probabilities is the joint probability of all variables.

- $P(X_1, X_2, \ldots, X_n) = \prod P(X_i \mid parents(X_i))$



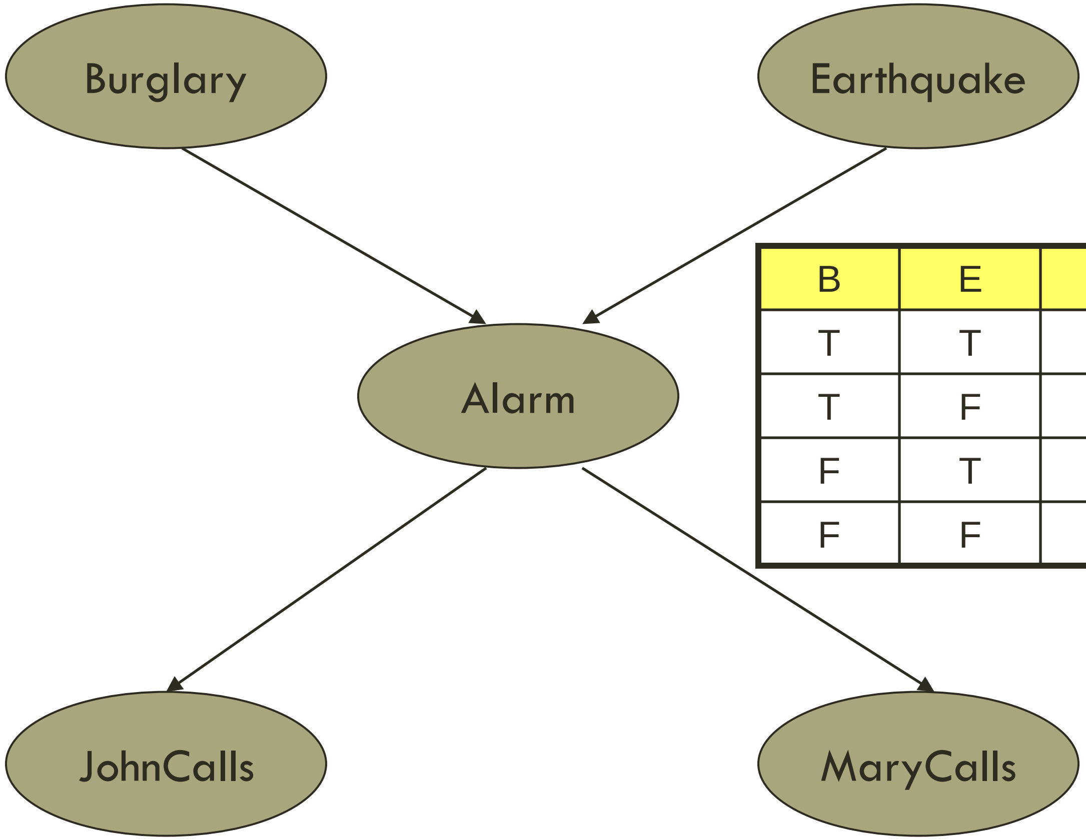

18 probabilities are required to specify the joint distribution

# WHY BAYESIAN NETWORKS?

There are two main reasons.

- It quickly becomes intractable to compute probabilities using the joint distribution, since the number of probabilities ($2^n$ ) is exponential in the number of variables (n).
- It is unnatural and tedious to specify all the probabilities

# Earthquake Example (Pearl)




| P(B) |
|---|
| 0.1 |

| P(E) |
|---|
| 0.02 |

| B | E | P(A\|B,E) |
|---|---|---|
| T | T | 0.95 |
| T | F | 0.94 |
| F | T | 0.29 |
| F | F | 0.001 |

| A | P(J\|A) |
|---|---|
| T | 0.9 |
| F | 0.05 |

| A | P(M\|A) |
|---|---|
| T | 0.7 |
| F | 0.01 |

## JOINT PROBABILITY TABLE FOR 'EARTHQUAKE' EXAMPLE

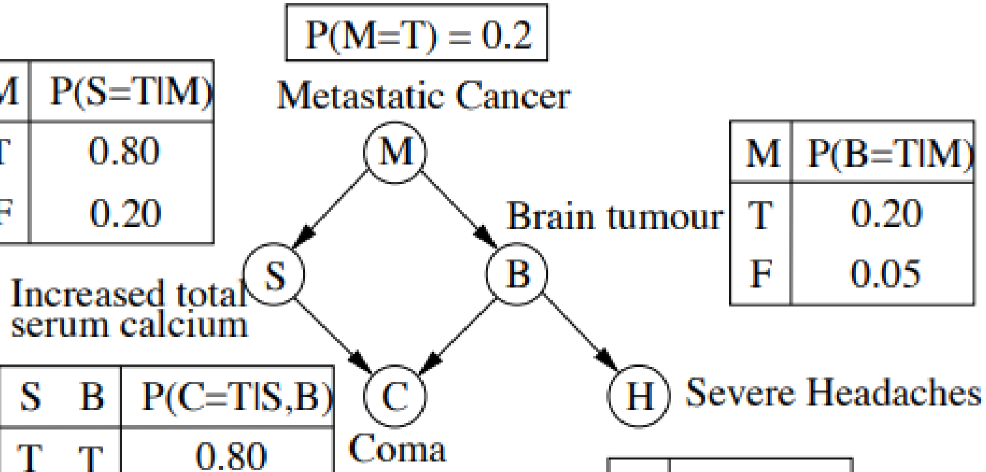
| B | E | A | J | M | P(B) | P(E) | P(A \|B,E) | P(J\|A) | P(M\|A) | P(B, E, A, J, M) |
|---|---|---|---|---|---|---|---|---|---|---|
| T | T | T | T | T | 0.01 | 0.02 | 0.95 | 0.90 | 0.70 | 0.0001197 |
| T | T | T | T | F | 0.01 | 0.02 | 0.95 | 0.90 | 0.30 | 0.0000513 |
| T | T | T | F | T | 0.01 | 0.02 | 0.95 | 0.10 | 0.70 | 0.0000133 |
| T | T | T | F | F | 0.01 | 0.02 | 0.95 | 0.10 | 0.30 | 0.0000057 |
| T | T | F | T | T | 0.01 | 0.02 | 0.05 | 0.05 | 0.01 | 5E-09 |
| T | T | F | T | F | 0.01 | 0.02 | 0.05 | 0.05 | 0.99 | 4.95E-07 |
| T | T | F | F | T | 0.01 | 0.02 | 0.05 | 0.95 | 0.01 | 9.5E-08 |
| T | T | F | F | F | 0.01 | 0.02 | 0.05 | 0.95 | 0.99 | 9.405E-06 |
| T | F | T | T | T | 0.01 | 0.98 | 0.94 | 0.90 | 0.70 | 0.0058036 |
| T | F | T | T | F | 0.01 | 0.98 | 0.94 | 0.90 | 0.30 | 0.0024872 |
| T | F | T | F | T | 0.01 | 0.98 | 0.94 | 0.10 | 0.70 | 0.0006448 |
| T | F | T | F | F | 0.01 | 0.98 | 0.94 | 0.10 | 0.30 | 0.0002764 |
| T | F | F | T | T | 0.01 | 0.98 | 0.06 | 0.05 | 0.01 | 2.94E-07 |
| T | F | F | T | F | 0.01 | 0.98 | 0.06 | 0.05 | 0.99 | 2.911E-05 |
| T | F | F | F | T | 0.01 | 0.98 | 0.06 | 0.95 | 0.01 | 5.586E-06 |
| T | F | F | F | F | 0.01 | 0.98 | 0.06 | 0.95 | 0.99 | 0.000553 |
| F | T | T | T | T | 0.99 | 0.02 | 0.29 | 0.90 | 0.70 | 0.0036175 |
| F | T | T | T | F | 0.99 | 0.02 | 0.29 | 0.90 | 0.30 | 0.0015503 |
| F | T | T | F | T | 0.99 | 0.02 | 0.29 | 0.10 | 0.70 | 0.0004019 |
| F | T | T | F | F | 0.99 | 0.02 | 0.29 | 0.10 | 0.30 | 0.0001723 |
| F | T | F | T | T | 0.99 | 0.02 | 0.71 | 0.05 | 0.01 | 7.029E-06 |
| F | T | F | T | F | 0.99 | 0.02 | 0.71 | 0.05 | 0.99 | 0.0006959 |
| F | T | F | F | T | 0.99 | 0.02 | 0.71 | 0.95 | 0.01 | 0.0001336 |
| F | T | F | F | F | 0.99 | 0.02 | 0.71 | 0.95 | 0.99 | 0.0132215 |
| F | F | T | T | T | 0.99 | 0.98 | 0.001 | 0.90 | 0.70 | 0.0006112 |
| F | F | T | T | F | 0.99 | 0.98 | 0.001 | 0.90 | 0.30 | 0.000262 |
| F | F | T | F | T | 0.99 | 0.98 | 0.001 | 0.10 | 0.70 | 6.791E-05 |
| F | F | T | F | F | 0.99 | 0.98 | 0.001 | 0.10 | 0.30 | 2.911E-05 |
| F | F | F | T | T | 0.99 | 0.98 | 0.999 | 0.05 | 0.01 | 0.0004846 |
| F | F | F | T | F | 0.99 | 0.98 | 0.999 | 0.05 | 0.99 | 0.0479769 |
| F | F | F | F | T | 0.99 | 0.98 | 0.999 | 0.95 | 0.01 | 0.0092077 |
| F | F | F | F | F | 0.99 | 0.98 | 0.999 | 0.95 | 0.99 | 0.9115606 |

# EXAMPLE PROBLEM: METASTATIC CANCER

Metastatic cancer is a possible cause of a brain tumor and is also an explanation for increased total serum calcium. In turn, either of these could explain a patient falling into a coma. Severe headache is also possibly associated with a brain tumor.

# EXAMPLE PROBLEM: METASTATIC CANCER

Metastatic cancer is a possible cause of a brain tumor and is also an explanation for increased total serum calcium. In turn, either of these could explain a patient falling into a coma. Severe headache is also possibly associated with a brain tumor.




| M | P(S=T\|M) |
|---|---|
| T | 0.80 |
| F | 0.20 |

| M | P(B=T\|M) |
|---|---|
| T | 0.20 |
| F | 0.05 |

| S | B | P(C=T\|S,B) |
|---|---|---|
| T | T | 0.80 |
| T | F | 0.80 |
| F | T | 0.80 |
| F | F | 0.05 |

| B | P(H=T\|B) |
|---|---|
| T | 0.80 |
| F | 0.60 |

FIGURE 2.7: Metastatic cancer BN.

# INFERENCE PROBLEM

The canonical inference problem: find the posterior probability distribution for some variable(s) given direct or virtual evidence about other variable(s).

Diagnostic Reasoning

- Reasoning from symptoms to cause
- This reasoning occurs in the opposite direction to the network arcs

Predictive Reasoning

- Reasoning from new information about causes to new beliefs about effects
- Follows the directions of the network arcs.

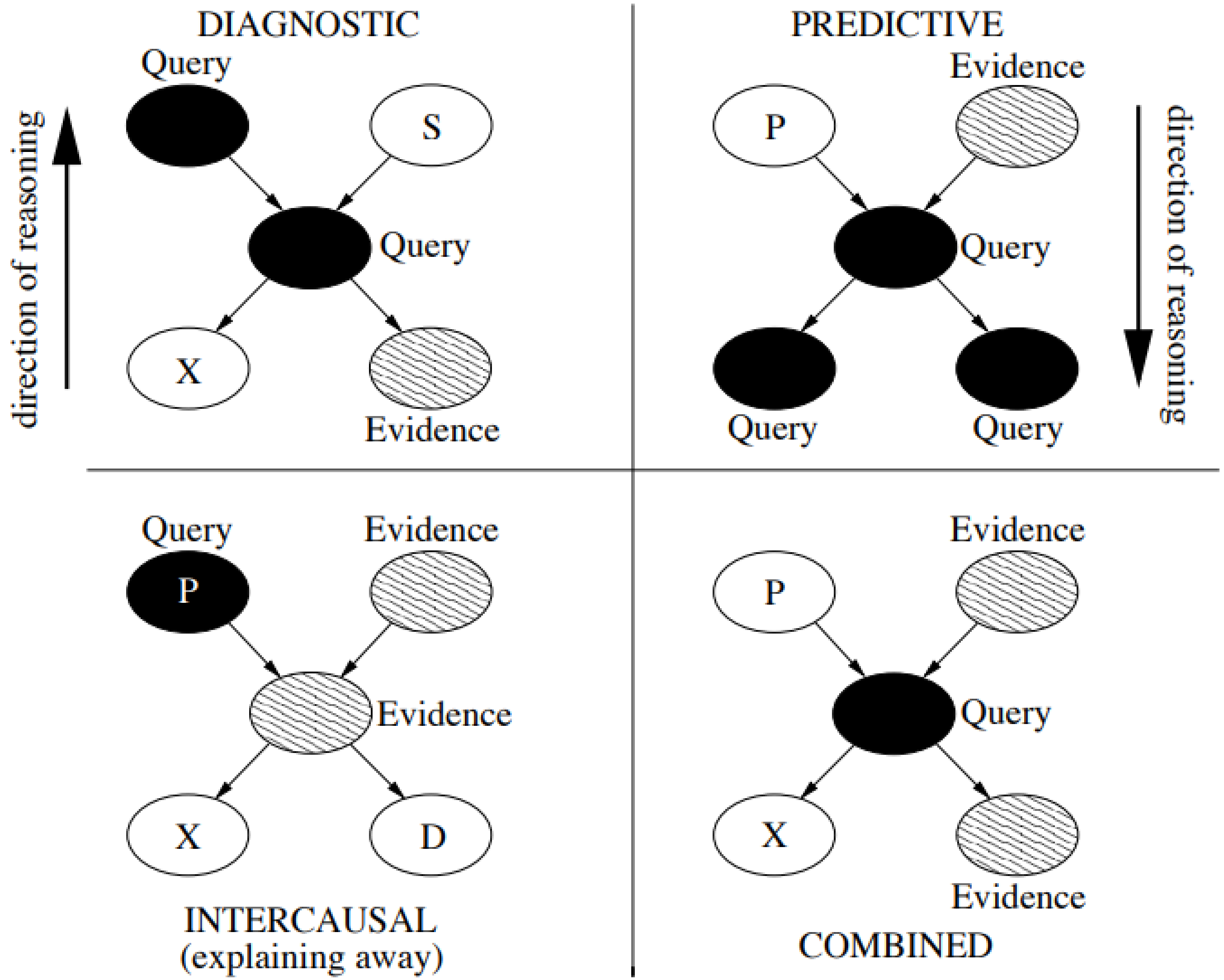



FIGURE 2.2: Types of reasoning.

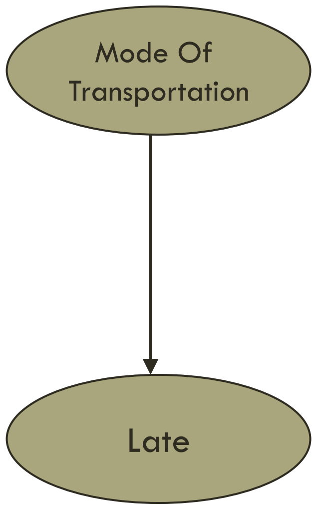
# REVISED "BOB LATE TO WORK" EXAMPLE

Mode of Transportation: {Car, Not Car (Public Transport)}

Late: {True, False}

$P(C) = 1/3$, $P(\neg C) = 2/3$

$P(L \mid C) = 0.4$, $P(L \mid \neg C) = 0.2$




**$P(L) = ?$**

**$P(C \mid L)$, $P(\neg C \mid L)$**

# APPROACH # 1 “BOB LATE TO WORK” EXAMPLE

One approach is to compute the joint distribution through

- $P(X_1, X_2, \ldots, X_n = \prod P(X_i \mid parents(X_i))$
- $P(L, C) = P(L \mid C)\, P(C) = 0.4 \times 0.33$
- $P(L, \neg C) = P(L \mid \neg C)\, P(\neg C) = 0.2 \times 0.67$
- $P(\neg L, C) = P(\neg L \mid C)\, P(C) = 0.6 \times 0.33$
- $P(\neg L, \neg C) = P(\neg L \mid \neg C)\, P(\neg C) = 0.8 \times 0.67$

Now compute P(L), P(C | L)

- $P(L) = P(L, C) + P(L, \neg C)$
- $P(C \mid L) = \dfrac{P(C, L)}{P(L)}$

# APPROACH # 2 “BOB LATE TO WORK” EXAMPLE

Or we could use Bayes Theorem based propagation.

$$P(C \mid L) = \frac{P(L \mid C)\, P(C)}{P(L)} = \frac{0.4 \times 0.33}{P(L)}$$

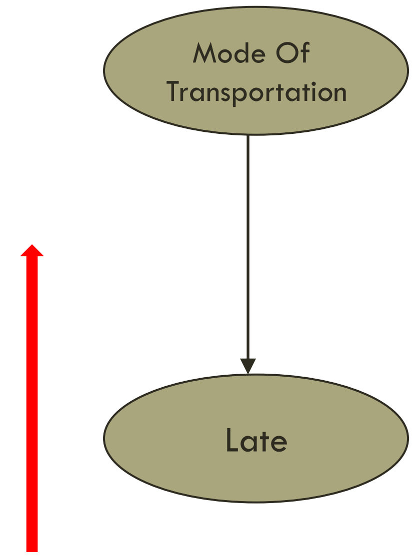
where

$$P(L) = P(L \wedge C) + P(L \wedge \neg C)$$

$$= P(L \mid C)P(C) + P(L \mid \neg C)\, P(\neg C)$$

$$= 0.4 \times 0.33 + 0.2 \times 0.67 = ?$$

Both approaches (on the previous and the current slides) should give the same result




Inference Direction

# MARKOV PROPERTY

- In a Bayesian network, each variable (node) is conditionally independent of its non-descendants given its parents.
- This means that for a given node Xi in the network, Xi is conditionally independent of all other nodes in the network that are not its descendants, given the values of its parent nodes.
- This property ensures that the joint probability distribution over all the variables can be factored into a product of conditional probabilities, making complex probabilistic reasoning tractable.

# REPRESENTING THE JOINT PROBABILITY DISTRIBUTION

Consider a BN containing the n nodes, X1 to Xn, taken in that order. A particular value in the joint distribution is represented by P(X1 = x1,X2 = x2,...,Xn = xn), or more compactly, P(x1, x2,..., xn). The chain rule of probability theory allows us to factorize joint probabilities so:

$$\begin{aligned} P(x_1, x_2, \ldots, x_n) &= P(x_1) \times P(x_2|x_1) \ldots, \times P(x_n|x_1, \ldots, x_{n-1}) \\ &= \prod_i P(x_i|x_1, \ldots, x_{i-1}) \end{aligned}$$

# REPRESENTING THE JOINT PROBABILITY DISTRIBUTION

However, for the structure of a BN implies that the value of a particular node is conditional only on the values of its parent nodes, this reduces to

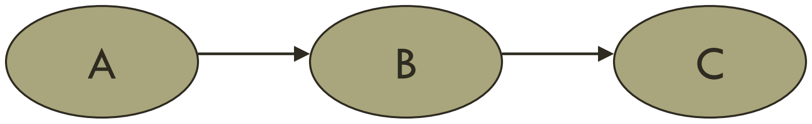
$$P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i | Parents(X_i))$$

# TYPES OF CONNECTIONS

- Causal Chain
- Common Cause
- Common Effect

# CAUSAL CHAIN

**Causal Chains (Serial Connection)**



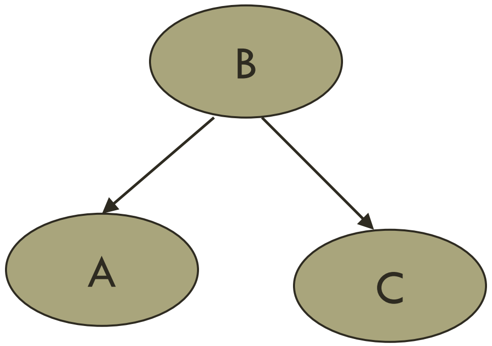

- smoking causes cancer which causes dyspnoea
- if we have evidence on B then A and C become independent
- $P(C|AB) = P(C|B)$


**Fig. 2.4.** A causal model for *Rainfall*, *Water level*, and *Flooding*.

# COMMON CAUSES / DIVERGING CONNECTION

**Common Causes (Diverging Connection)**

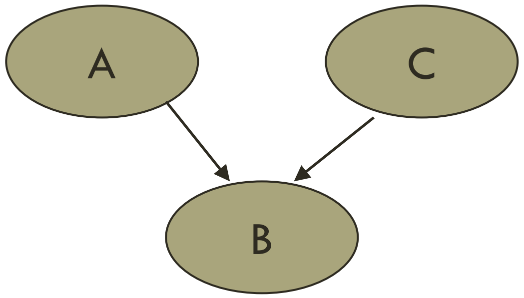
- cancer is a common cause of the two symptoms, a positive XRay result and dyspnoea
- If we have evidence on B then A and C become independent



# COMMON EFFECTS / CONVERGING CONNECTION

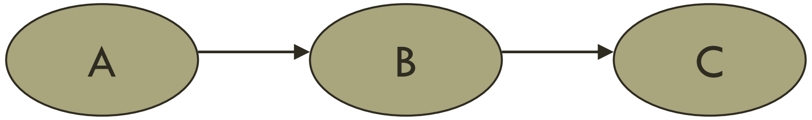
**Common Effects (Converging Connection)**

- Cancer is a common effect of pollution and smoking
- If we have evidence on B then A and C become dependent



# SERIAL CONNECTION EXAMPLE

Consider the following BN



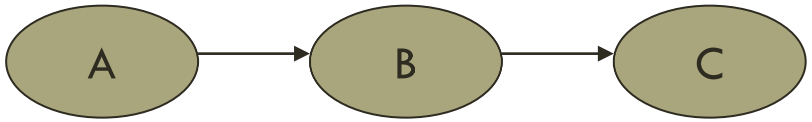

$P(A) = 0.4$

$P(B|A) = 0.7$
$P(B|\neg A) = 0.2$

$P(C|B) = 0.05$
$P(C|\neg B) = 0.8$

# SERIAL CONNECTION EXAMPLE

Consider the following BN




$P(A) = 0.4$

$P(B|A) = 0.7$
$P(B|\neg A) = 0.2$

$P(C|B) = 0.05$
$P(C|\neg B) = 0.8$

Using the Bayesian Chain Rule, compute the joint distribution (i.e., eight values: $P(A, B, C)$, $P(A, B, \neg C)$, ………., $P(\neg A, \neg B, C)$ and $P(\neg A, \neg B, \neg C)$)

Now compare

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

# PRACTICE QUESTIONS

Develop a simple diverging connection style BN (as shown in the earlier slide).

After computing the joint probability distribution, verify that

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

Develop a simple converging connection style BN. After computing the joint distribution, verify that

- $P(A \mid C) = P(A)$ but $P(A \mid B, C) \mathrel{!=} P(A \mid B)$
- $P(C \mid A) = P(C)$ but $P(C \mid A, B) \mathrel{!=} P(C \mid B)$

# D-SEPARATION

d-separation is short for direction-dependent separation

d-separation is the mathematical basis for efficient inference algorithms in Bayesian networks

Two nodes X and Y are d-separated by a set Z if all paths between X and Y are blocked by Z.

When X and Y are d-separated by Z, no information can be transmitted between X and Y given Z. Hence X and Y are conditional independent given Z.

# D-SEPARATION

**Definition 2.1 (d-separation).** *Two distinct variables A and B in a causal network are* d-separated *("d" for "directed graph") if for all paths between A and B, there is an intermediate variable V (distinct from A and B) such that either*

- *the connection is serial or diverging and V is instantiated*

  *or*
- *the connection is converging, and neither V nor any of V's descendants have received evidence.*

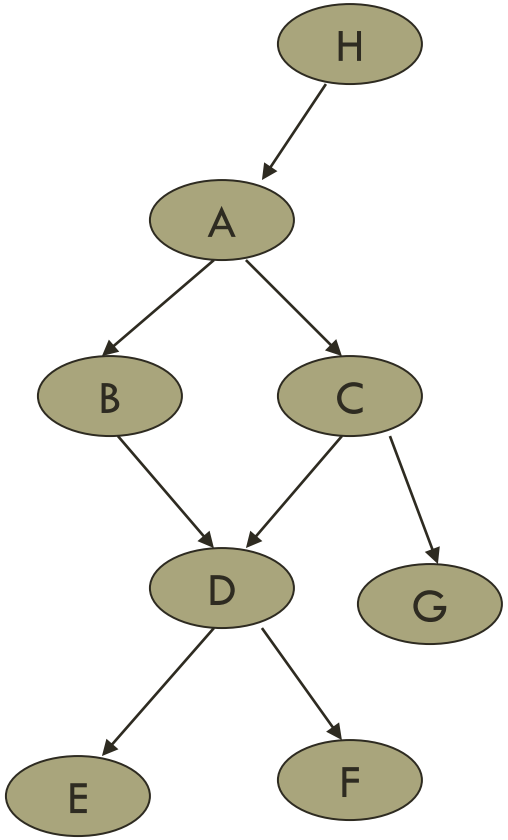
*If A and B are not d-separated, we call them* d-connected.

# MARKOV BLANKET

A node's Markov blanket consists of its:

- parents
- children
- other parents of its children (co-parents)

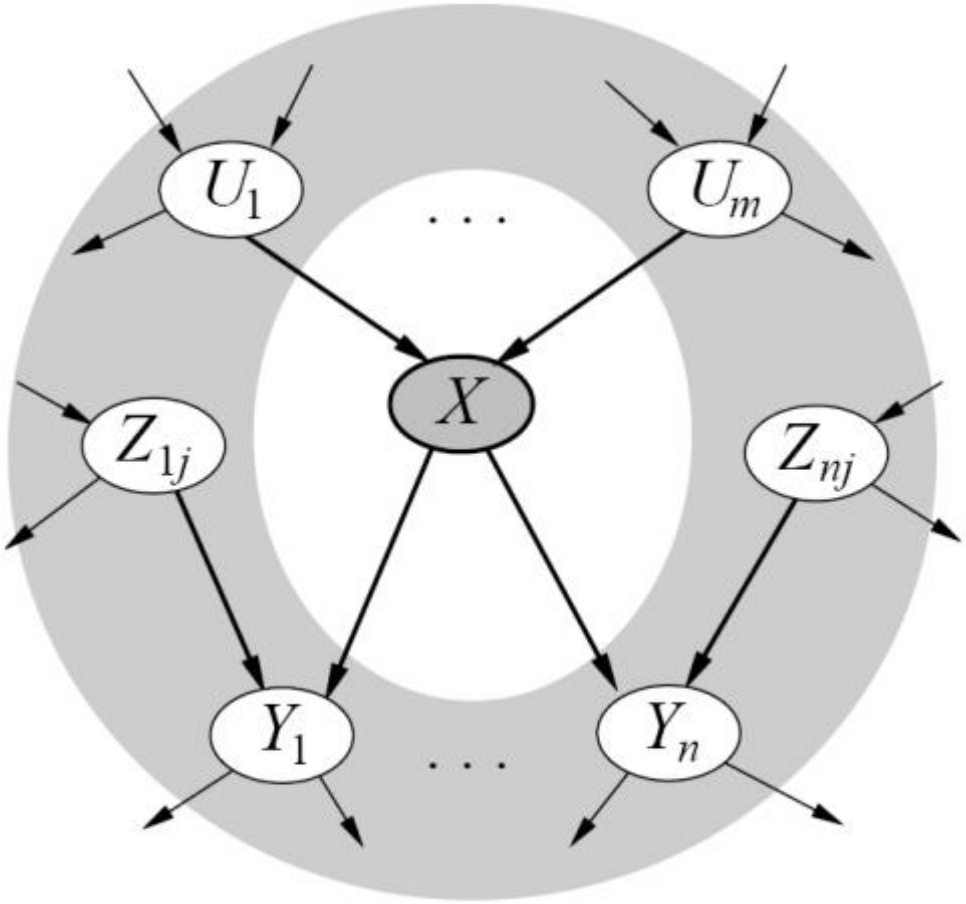
The Markov blanket of node B consists of all nodes whose local probability table mentions B and all nodes their local probability tables mention.



# MARKOV BLANKET (COTD.)

A node's Markov blanket d-separates it from all other nodes in the graph

- A node is conditionally independent of all other nodes given its Markov blanket



# INFERENCE MECHANISM

A and K are independent?

A and K are independent given G?

What's the Markov blanket of H?

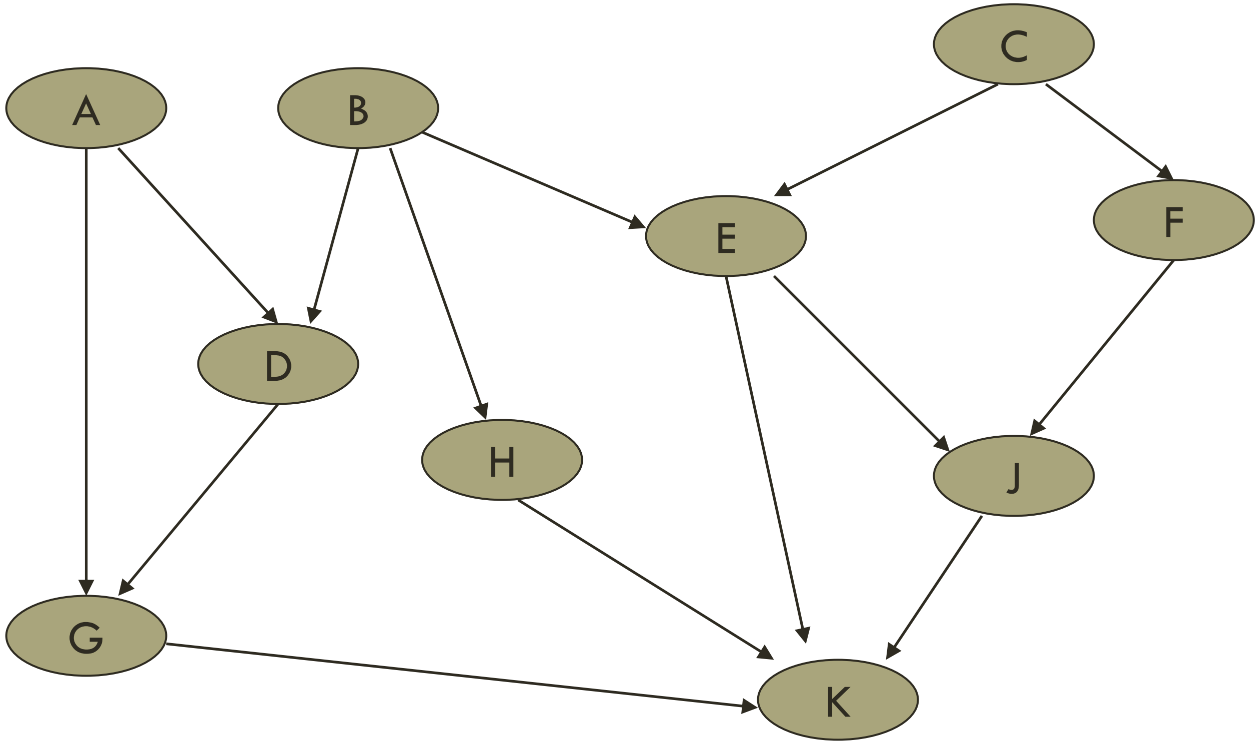
A and B are independent given D?

E and J are independent given K?

G and B are independent given D?

E and H are independent given B?

E and H are independent given B and K?



# INFERENCE MECHANISM

A and K are independent? (F)

A and K are independent given G? (F)

What's the Markov blanket of H? {B,K,G,E,J}

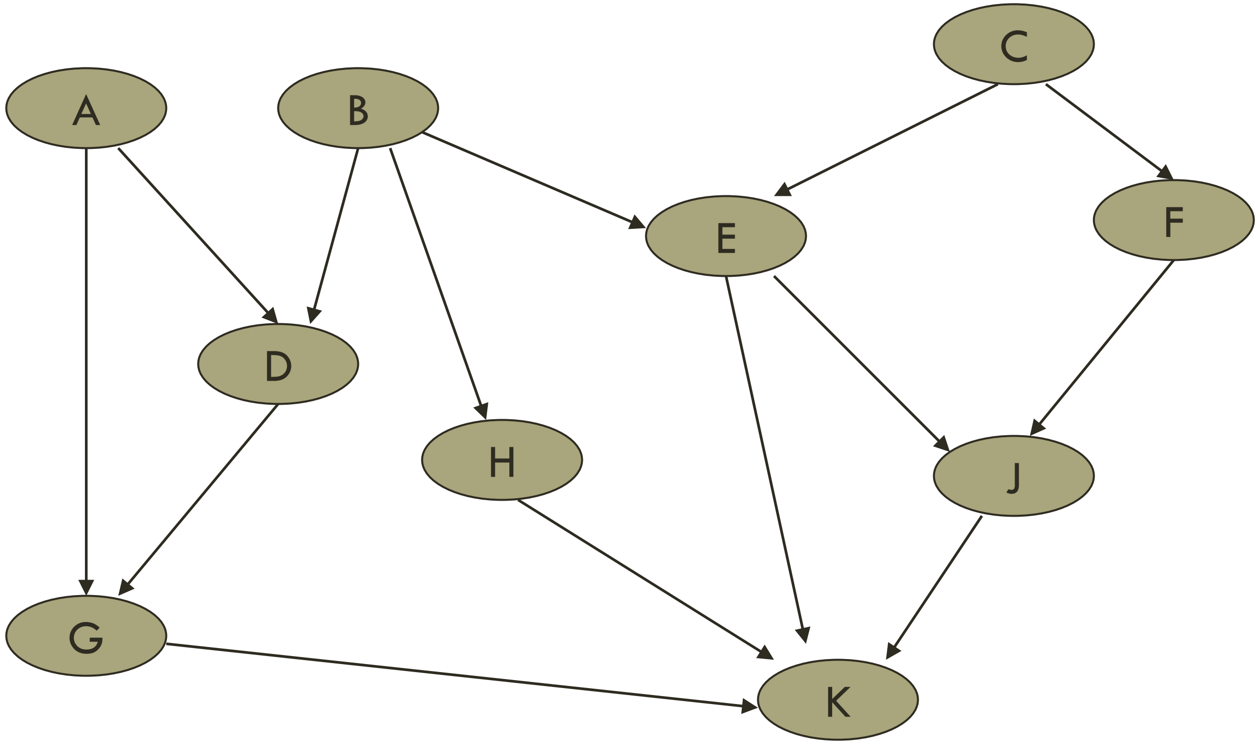
A and B are independent given D? (F)

E and J are independent given K? (F)

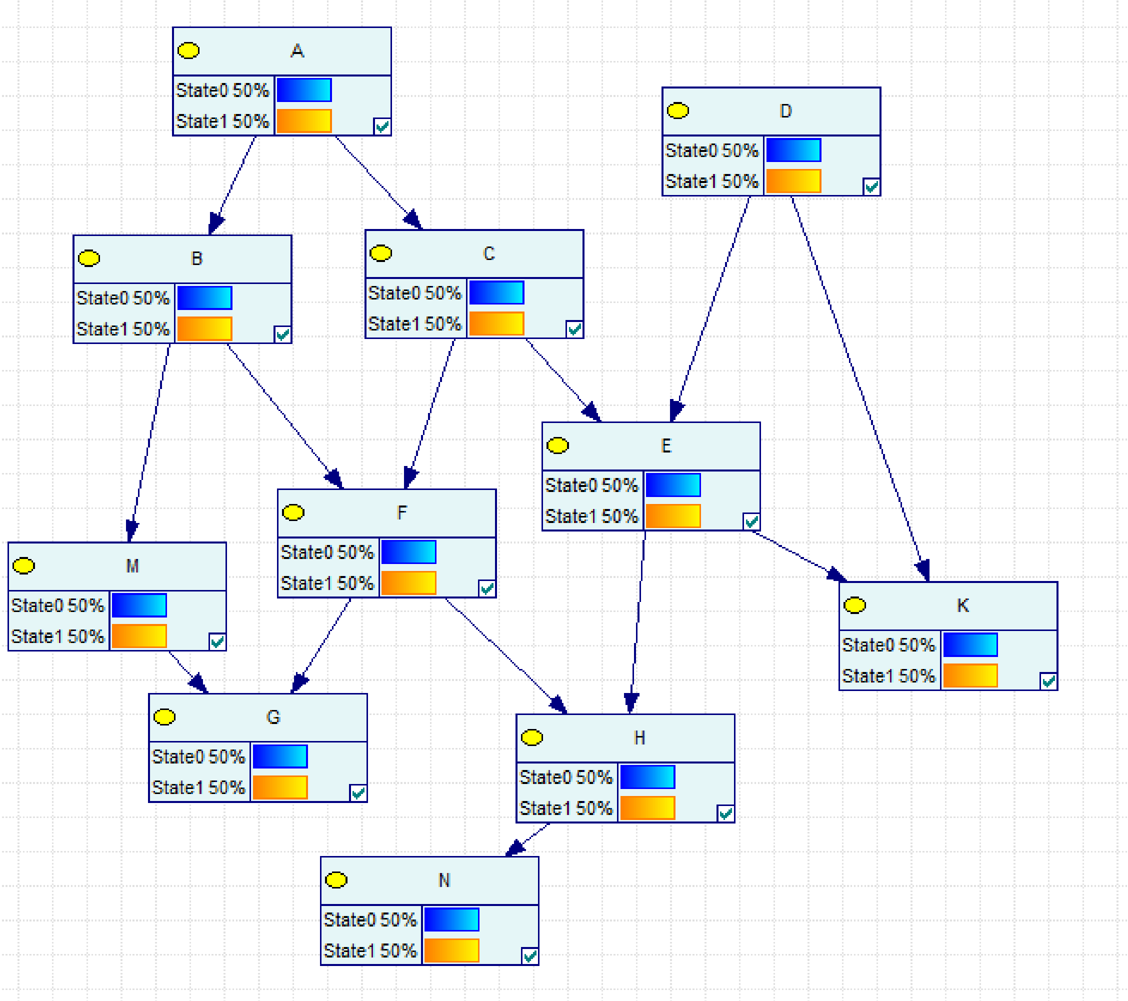
G and B are independent given D? (F)

E and H are independent given B? (T)

E and H are independent given B and K? (F)



# INFERENCE MECHANISM USING GENIE



# INFERENCE MECHANISM USING GENIE (CONT'D)

When you enter evidence on a node (say E), GeNIe doesn't hide/change the probability of nodes which are independent of the node E.

For instance, in this example, M and D are independent. Hence entering evidence on M doesn't change the probability of D.