CMSC 170

Introduction to Artificial Intelligence

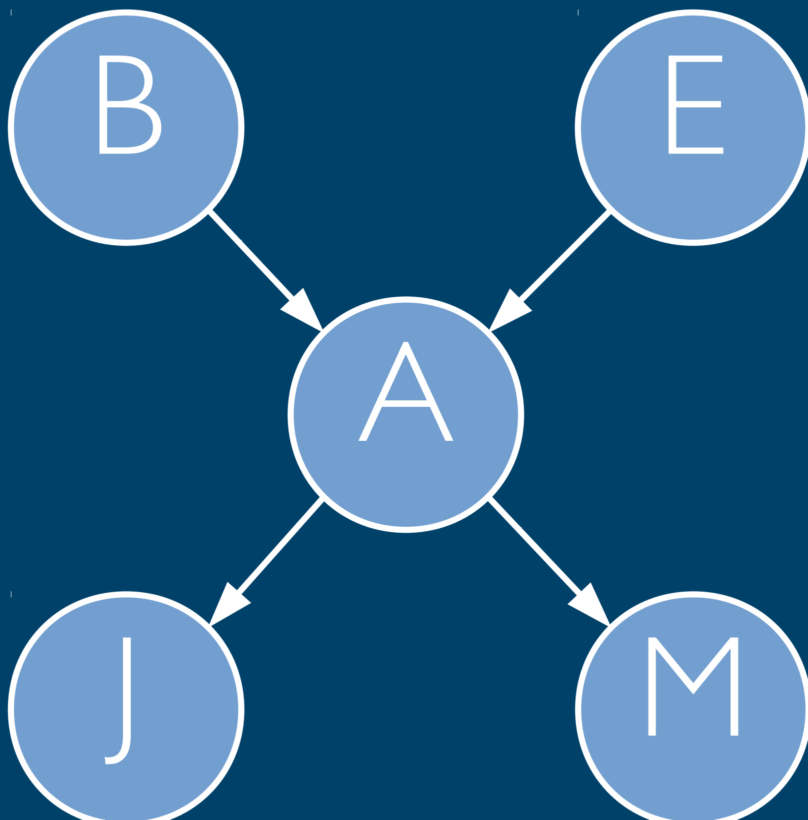
CNM Peralta

2nd Semester AY 2014-2015

PROBABILISTIC INFERENCE

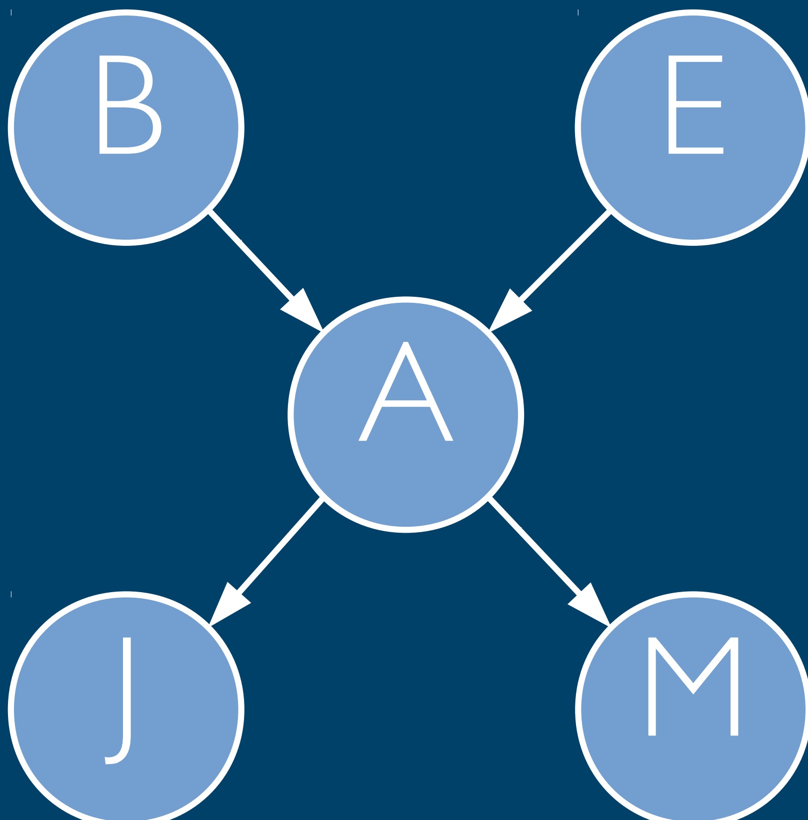
We will now use our knowledge of Bayes networks for **inference**.

EXAMPLE



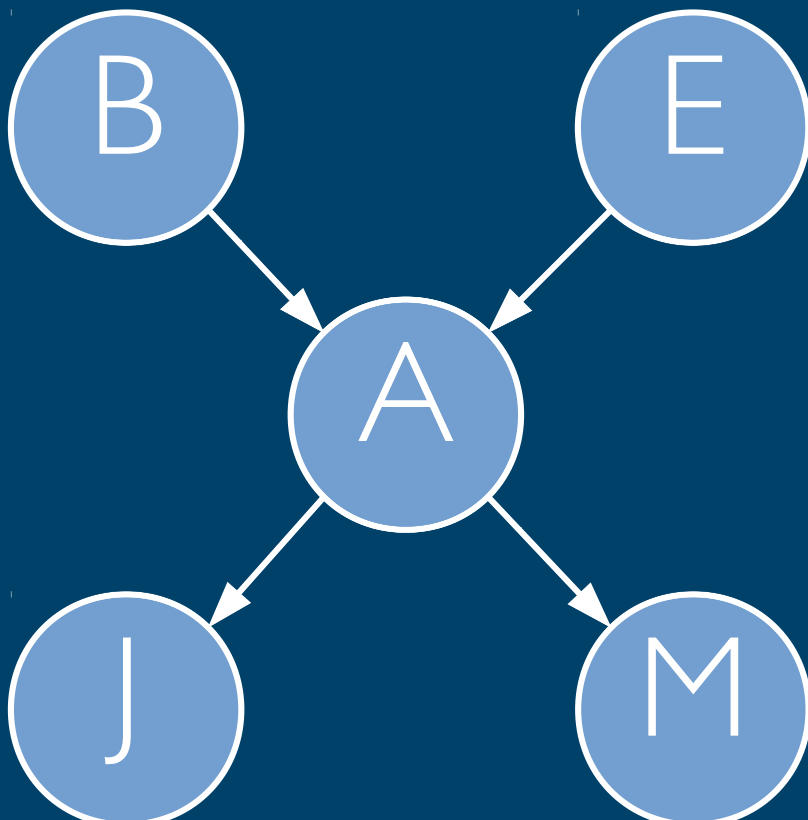
If there is a Burglary or Earthquake, the house Alarm will go off. If the house Alarm goes off, either John or Mary will call.

EXAMPLE



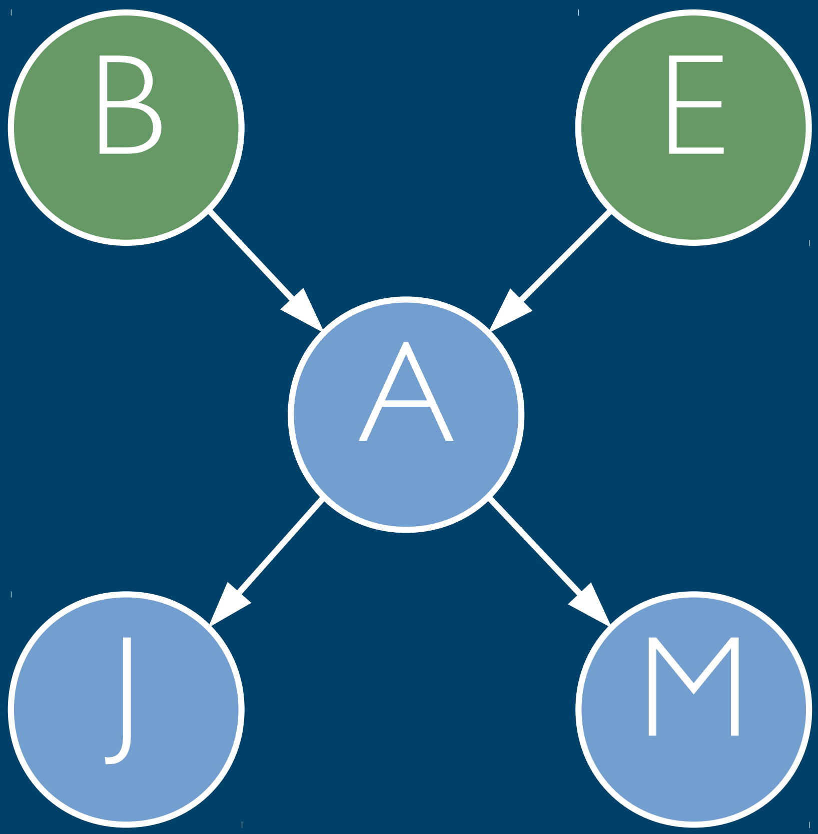
This can be treated as an input-output problem.

EXAMPLE



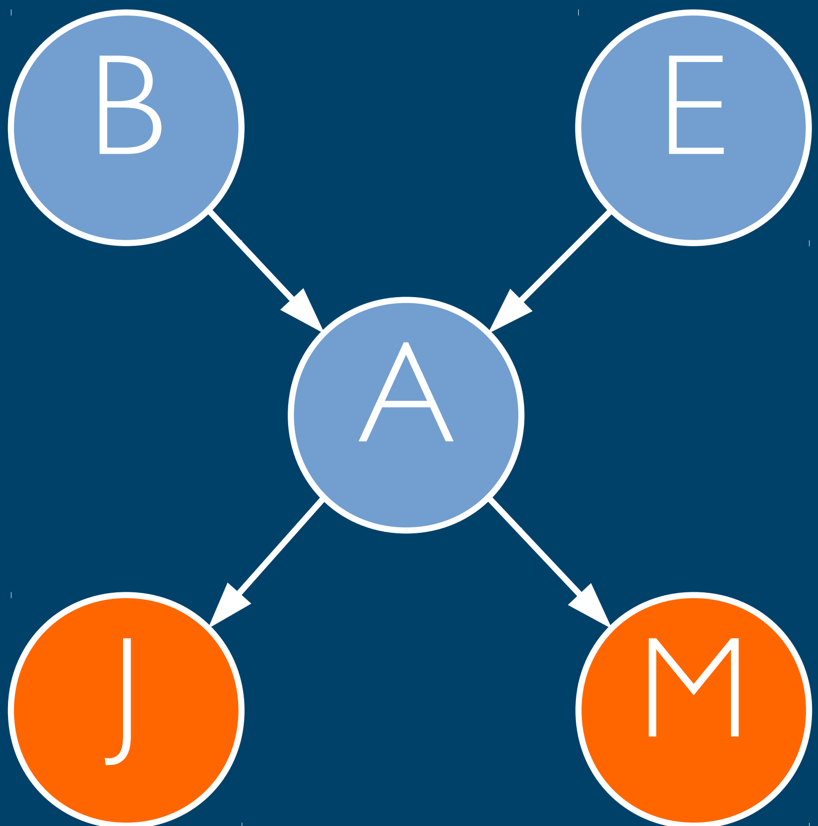
Given the inputs, what are the outputs?

EXAMPLE



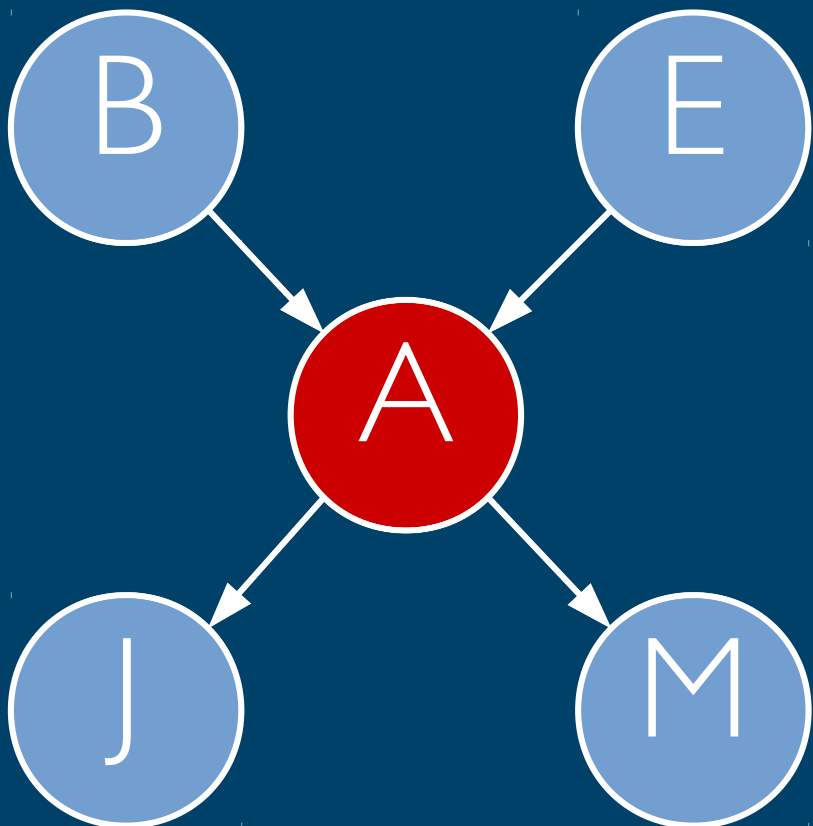
The inputs are called the ***evidence***.

EXAMPLE



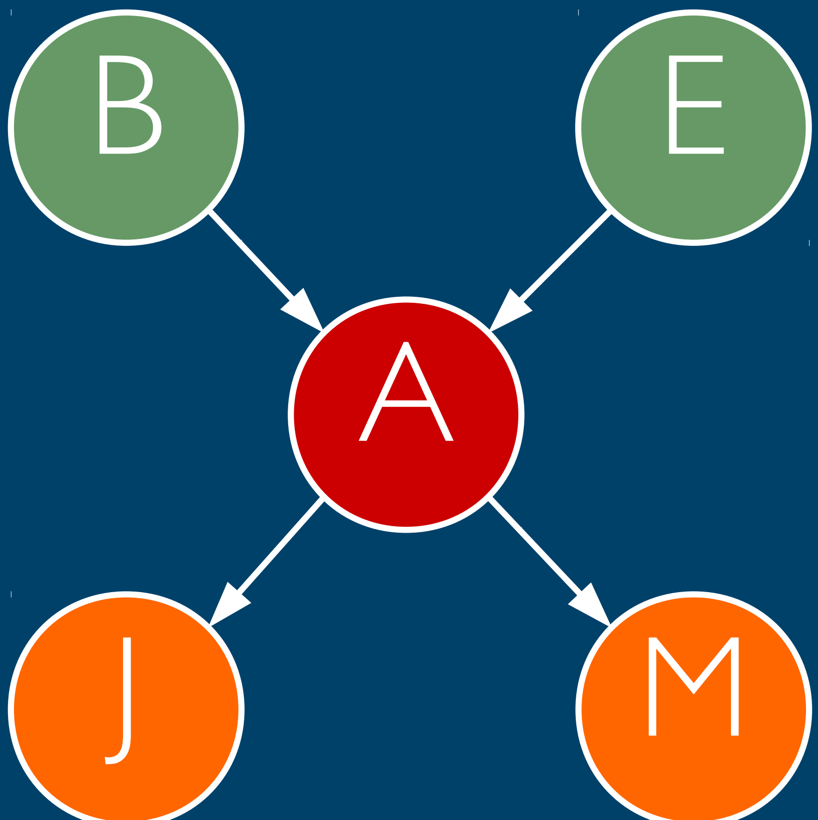
The outputs are called the *query*.

EXAMPLE



Anything that is neither input nor output is *hidden*.

EXAMPLE

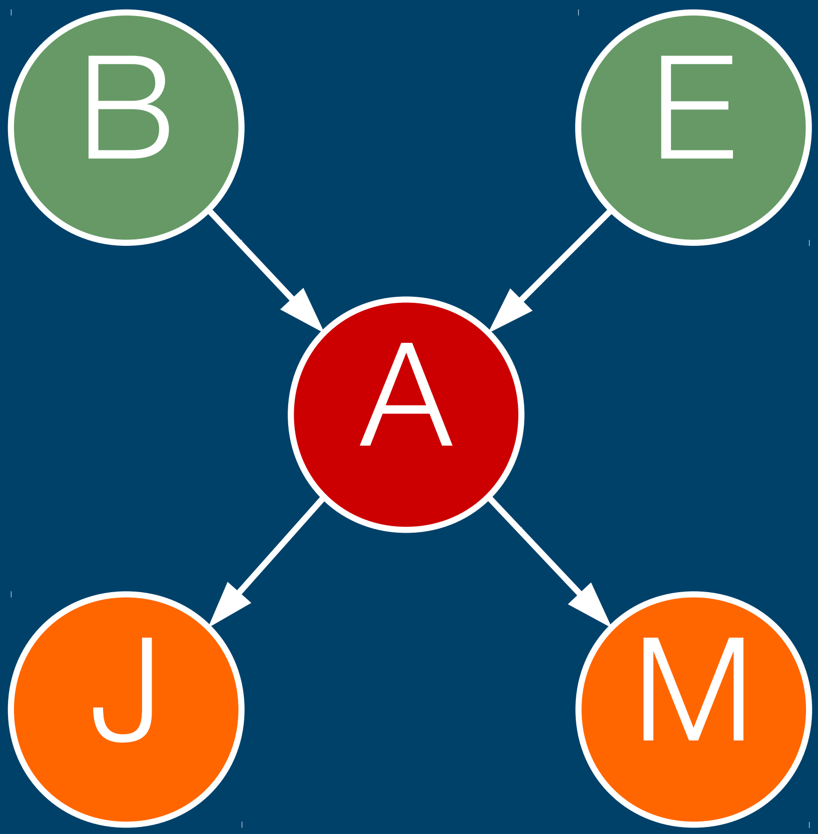


In this example, the **evidence** is that a **burglary** and **earthquake** occurred, and the **query** is whether **John** and **Mary will call**.

The output is a complete, joint probability distribution overe the query variables, called the

posterior distribution given the evidence

EXAMPLE



Output:

$$P(J, M \mid B, E)$$

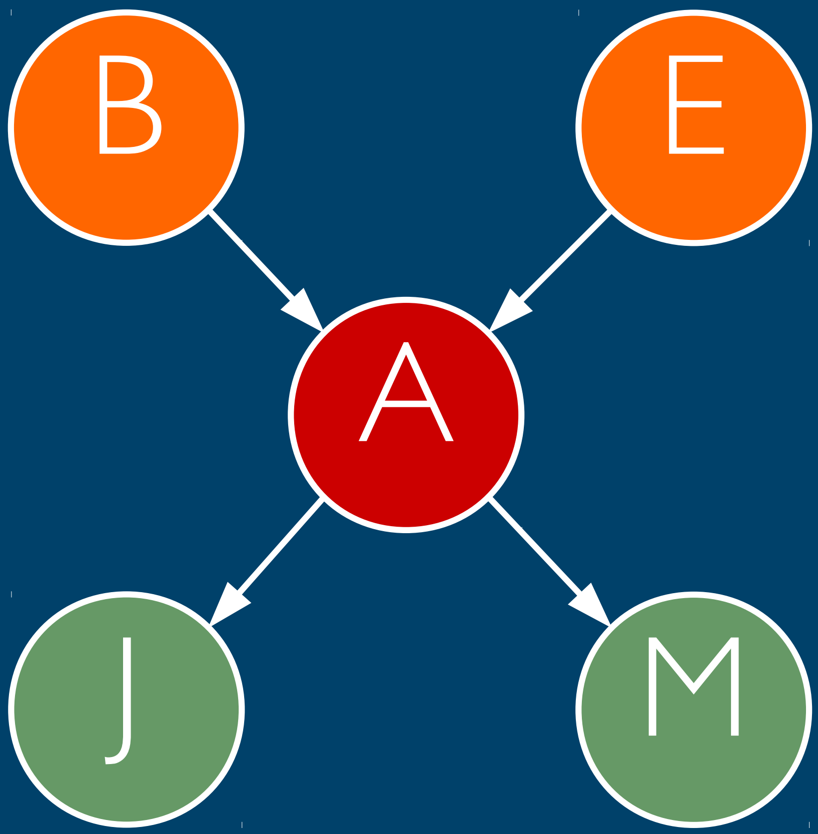
Probability of one or more query variables...

...given the evidence.

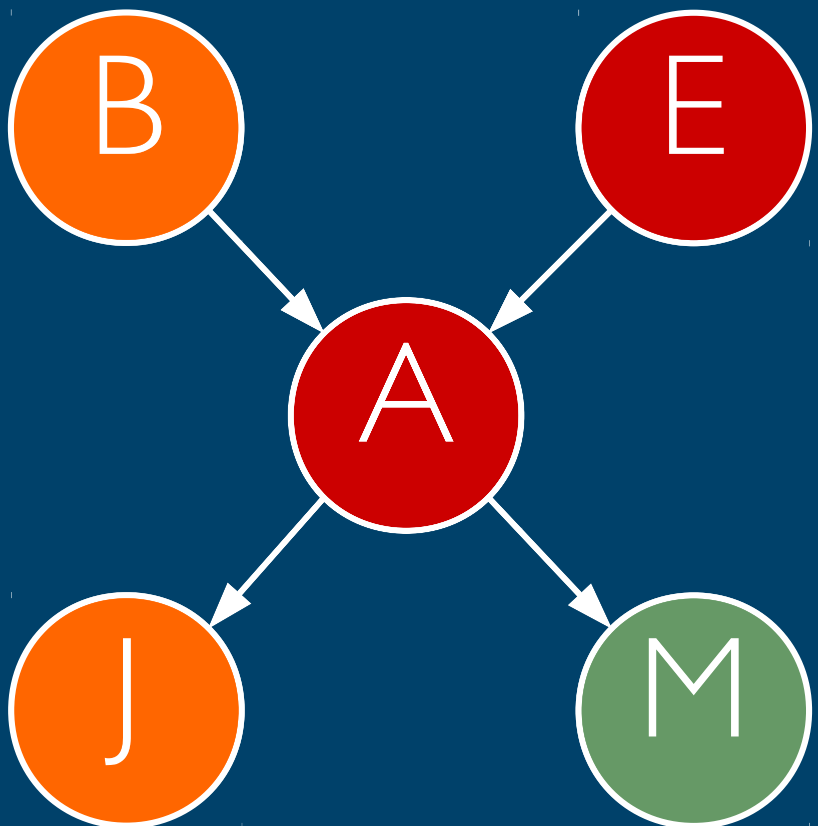
It is also possible to ask what the **most likely explanation** is, by instead computing:

$$max(P(J=q_1, M=q_2 \mid B = T, E = T)$$

What combination of values of J and M is most likely to occur given both B and E are true?

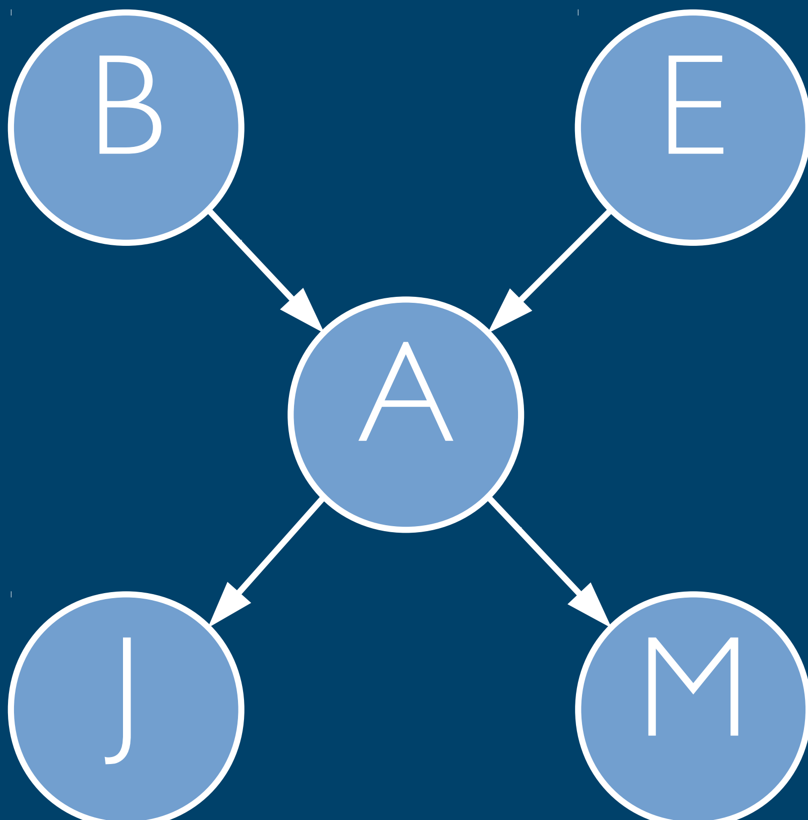


The flow of input-output can actually be reversed...



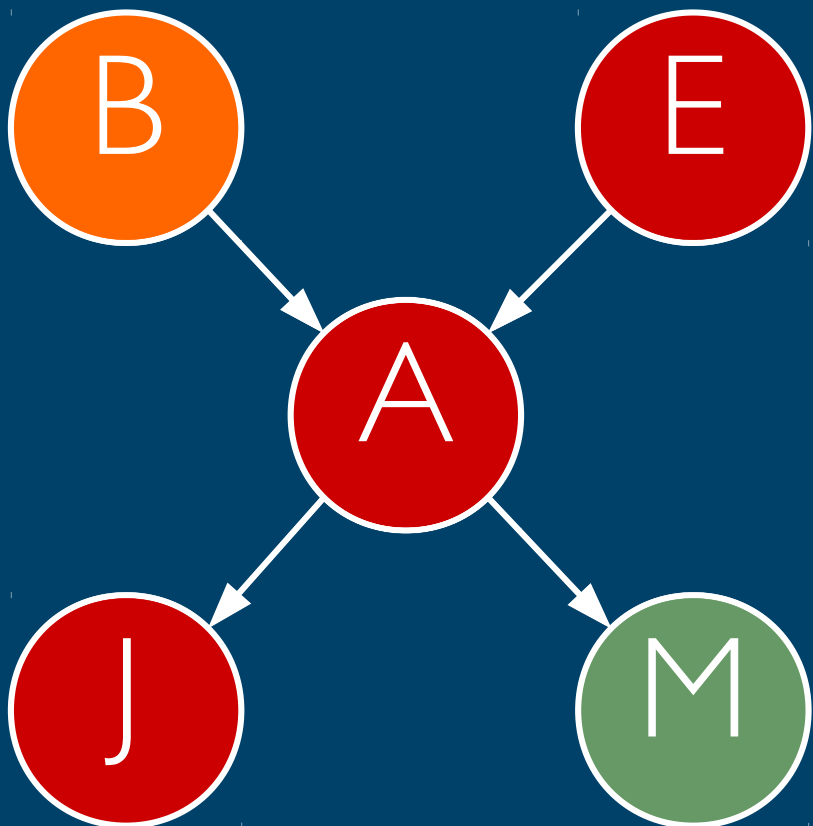
...or be in any direction, any combination.

EXAMPLE



If Mary calls, and we want to know if there was a burglary, which of the nodes are considered evidence, query, and hidden?

ANSWER



M is the evidence, B is the query, and J, A, and E are hidden.

Enumeration

A method of probabilistic inference that goes through all possibilities and adds them up to get an answer.

Conditional Probability

$$P(Q|E)=\frac{P(Q,E)}{P(E)}$$

EXAMPLE

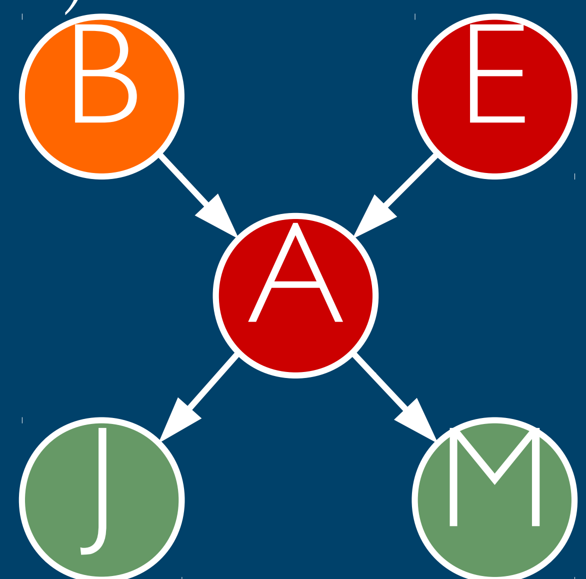
Given that John and Mary called, what is the probability of a burglary?

$$P\left(+B \middle| +J, +M\right) = \frac{P\left(+B, +J, +M\right)}{P\left(+J, +M\right)}$$

EXAMPLE

Expand the expression using the flow of arrows from the Bayes network.

$$P(+B,+J,+M)$$
$$=\sum_E \sum_A P(+B,+J,+M,E,A)$$
$$=P(+B)P(E)P(A|+B,E)$$
$$P(+J|A)P(+M|A)$$



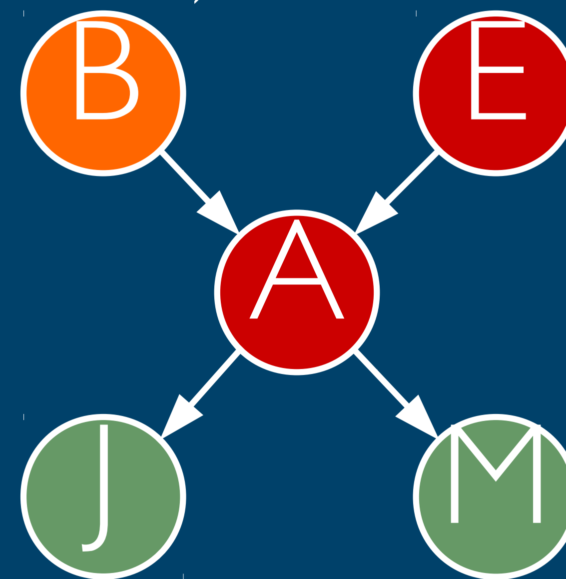
Let's call this expression $F(E, A)$.

EXAMPLE

Expand the expression using the flow of arrows from the Bayes network.

$$P(+B,+J,+M)$$
$$=F(+E,+A)+F(+E,\neg A)$$
$$+F(\neg E,+A)+F(\neg E,\neg A)$$

$$F(E,A) = P(+B)P(E) + P(A|+B,E) + P(+J|A) + P(+M|A)$$



B	P(B)
+B	0.001
¬B	0.999

E	P(E)
+E	0.002
¬E	0.998

A	J	P(J|A)
+A	+J	0.9
+A	¬J	0.1
¬A	+J	0.05
¬A	¬J	0.95

B	E	A	P(A|B, E)
+B	+E	+A	0.95
+B	+E	¬A	0.05
+B	¬E	+A	0.94
+B	¬E	¬A	0.06
¬B	+E	+A	0.29
¬B	+E	¬A	0.71
¬B	¬E	+A	0.001
¬B	¬E	¬A	0.999

A	M	P(M|A)
+A	+M	0.7
+A	¬M	0.3
¬A	+M	0.01
¬A	¬M	0.99

EXAMPLE

$$\begin{aligned} F(+E,+A) &= P(+B)P(+E)P(+A|+B,+E) \\ &\quad P(+J|+A)P(+M|+A) \\ &= 0.001\times 0.002\times 0.95\times 0.9\times 0.7 \end{aligned}$$

EXAMPLE

$$
\begin{aligned}
F(+E,+A) &= P(+B)P(+E)P(+A|+B,+E) \\
&\quad P(+J|+A)P(+M|+A) \\
&= 0.001\times 0.002\times 0.95\times 0.9\times 0.7 \\
&= 0.00001197
\end{aligned}
$$

EXAMPLE

$$
\begin{aligned}
F(+E,\neg A) &= P(+B)P(+E)P(\neg A|+B,+E) \\
&\quad P(+J|\neg A)P(+M|\neg A) \\
&= 0.001 \times 0.002 \times 0.05 \times 0.05 \times 0.01
\end{aligned}
$$

EXAMPLE

$$\begin{aligned} F(+E,\neg A) &= P(+B)P(+E)P(\neg A|+B,+E) \\ &\quad P(+J|\neg A)P(+M|\neg A) \\ &= 0.001\times 0.002\times 0.05\times 0.05\times 0.01 \\ &= 0.0000000005 \end{aligned}$$

EXAMPLE

$$
\begin{aligned}
F(\neg E, +A) &= P(+B)P(\neg E)P(+A|+B, \neg E) \\
&\quad P(+\mathrm{J}|+A)P(+M|+A) \\
&= 0.001 \times 0.998 \times 0.94 \times 0.9 \times 0.7
\end{aligned}
$$

EXAMPLE

$$
\begin{aligned}
F(\neg E, +A) &= P(+B)P(\neg E)P(+A|+B, \neg E) \\
&\quad P(+\mathrm{J}|+A)P(+M|+A) \\
&= 0.001 \times 0.998 \times 0.94 \times 0.9 \times 0.7 \\
&= 0.0005910156
\end{aligned}
$$

EXAMPLE

$$
\begin{aligned}
F(\neg E, \neg A) &= P(+B)P(\neg E)P(\neg A|+B, \neg E) \\
&\quad P(+J|\neg A)P(+M|\neg A) \\
&= 0.001 \times 0.998 \times 0.06 \times 0.05 \times 0.01
\end{aligned}
$$

EXAMPLE

$$
\begin{aligned}
F(\neg E, \neg A) &= P(+B)P(\neg E)P(\neg A|+B, \neg E) \\
&\quad P(+J|\neg A)P(+M|\neg A) \\
&= 0.001 \times 0.998 \times 0.06 \times 0.05 \times 0.01 \\
&= 0.0000002994
\end{aligned}
$$

EXAMPLE

$$P(+B,+J,+M) = \begin{aligned} & 0.000001197 + \\ & 0.0000000005 + \\ & 0.0005910156 + \\ & 0.0000002994 \end{aligned}$$

EXAMPLE

$$\begin{aligned} P(+B,+J,+M) &= 0.000001197+ \\ &\quad\ 0.00000000005+ \\ &\quad\ 0.0005910156+ \\ &\quad\ 0.0000002994 \\ &= 0.00059224259 \end{aligned}$$

EXAMPLE

$$P(+B \mid +J, +M) = \frac{0.00059224259}{P(+J, +M)}$$

Oops, that was just the numerator.

EXAMPLE

We still need to solve for the denominator:

$$P(+J,+M) = \sum_{B}\sum_{E}\sum_{A} P(B,+J,+M,E,A)$$

FAST FORWARD...

$$P(+B|+J,+M) = \frac{0.00059224259}{P(+J,+M)} = 0.284$$

Why is the probability low, even though both John and Mary called?

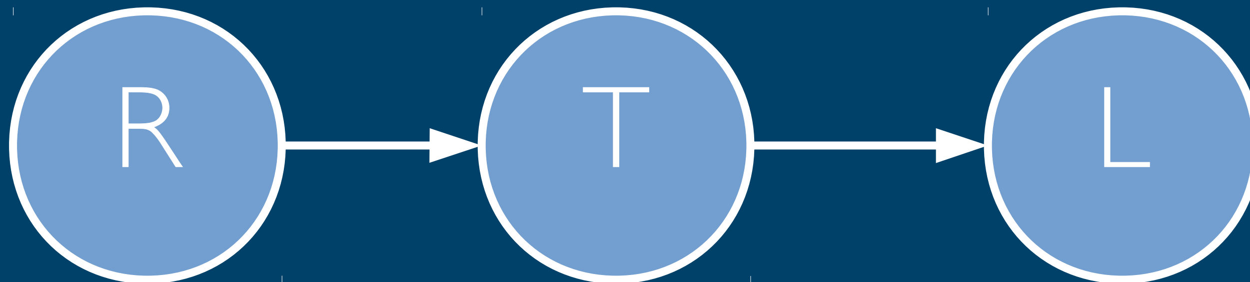
The problem with enumeration is that it takes too long, especially when there are too many nodes.

Variable Elimination

An alternative approach to probabilistic inference that **methodically shrinks the Bayes network** into a **single node**, with a probability distribution derived from eliminated variables.

Variable elimination is faster than enumeration in most practical networks; it is just a cycle of joining factors and marginalizing variables.

EXAMPLE



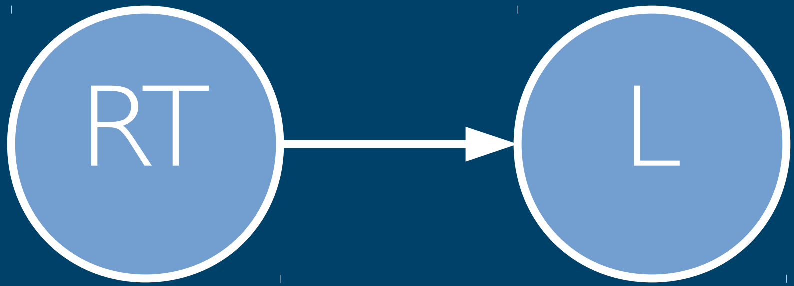
R	P(R)
+R	0.1
¬R	0.9

R	T	P(T | R)
+R	+T	0.8
+R	¬T	0.2
¬R	+T	0.1
¬R	¬T	0.9

T	L	P(L | T)
+T	+L	0.3
+T	¬L	0.7
¬T	+L	0.1
¬T	¬L	0.9

EXAMPLE

Join R and T.



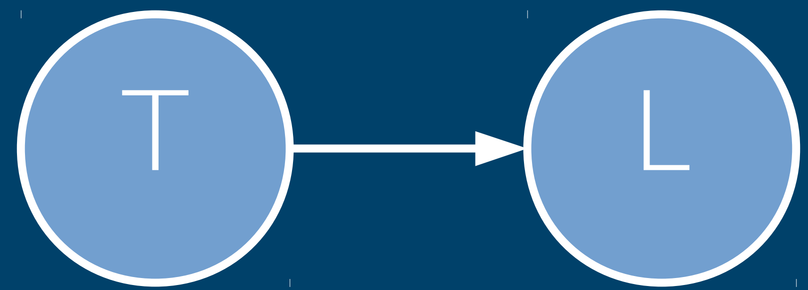
R	P(R)
+R	0.1
¬R	0.9

R	T	P(T | R)
+R	+T	0.8
+R	¬T	0.2
¬R	+T	0.1
¬R	¬T	0.9

R	T	P(R, T)
+R	+T	0.08
+R	¬T	0.02
¬R	+T	0.09
¬R	¬T	0.81

EXAMPLE

Marginalize R from RT.

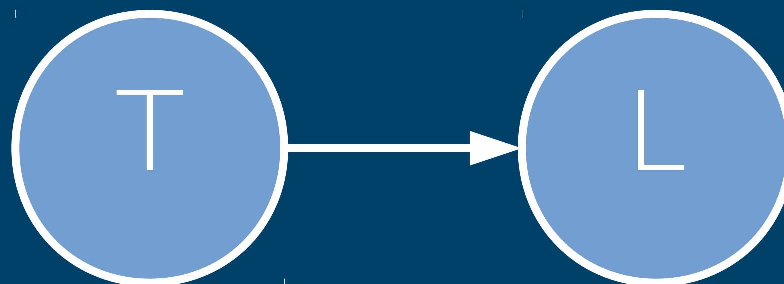


R	T	P(R, T)
+R	+T	0.08
+R	¬T	0.02
¬R	+T	0.09
¬R	¬T	0.81

T	P(T)
+T	0.17
¬T	0.83

EXAMPLE

We have eliminated R.



T	P(T)
+T	0.17
¬T	0.83

T	L	P(L | T)
+T	+L	0.3
+T	¬L	0.7
¬T	+L	0.1
¬T	¬L	0.9

EXAMPLE

Join T and L.

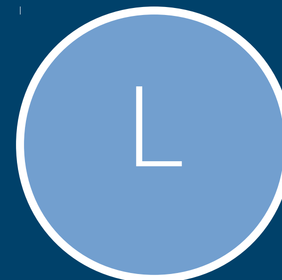
T	P(T)
+T	0.17
¬T	0.83

T	L	P(L | T)
+T	+L	0.3
+T	¬L	0.7
¬T	+L	0.1
¬T	¬L	0.9

T	L	P(TL)
+T	+L	0.051
+T	¬L	0.119
¬T	+L	0.083
¬T	¬L	0.747

EXAMPLE

Marginalize T from TL.



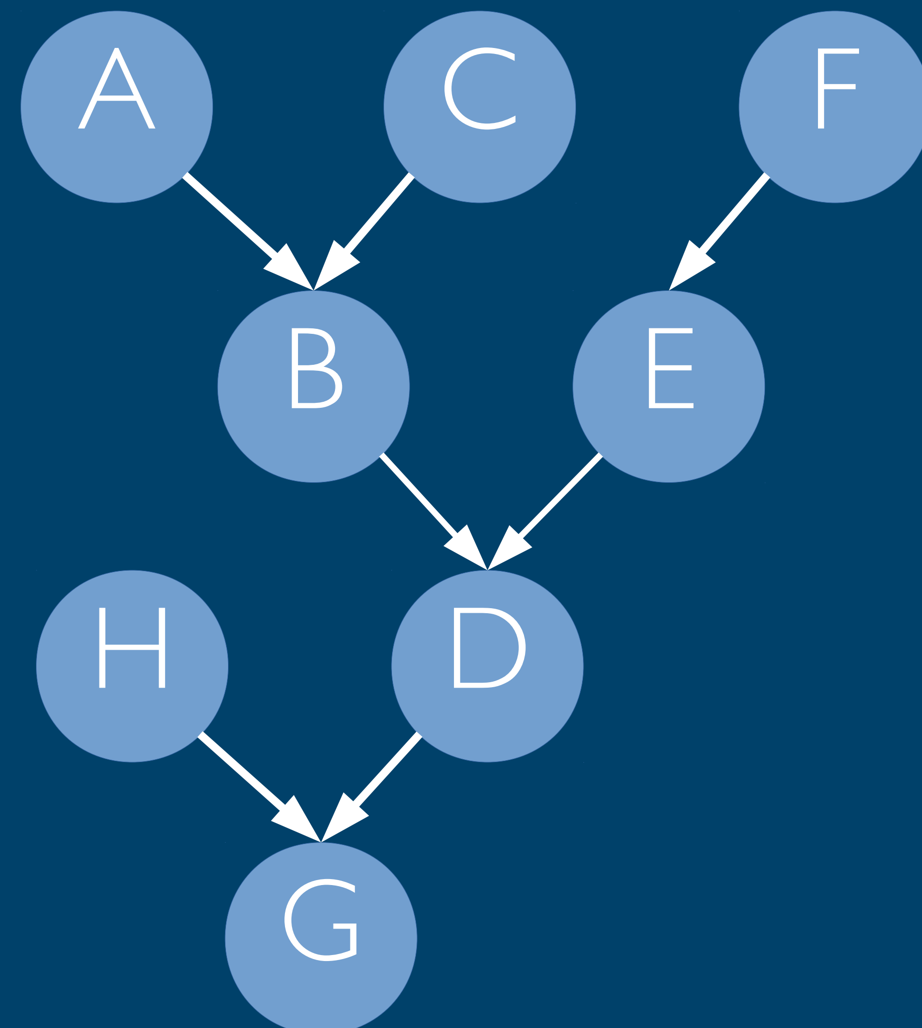
T	L	P(T, L)
+T	+L	0.051
+T	¬L	0.119
¬T	+L	0.083
¬T	¬L	0.747

L	P(L)
+L	0.134
¬L	0.866

The order in which joining factors and marginalizing variables is done will dictate if variable elimination is more efficient then enumeration; regardless, probabilistic inference is NP-hard.

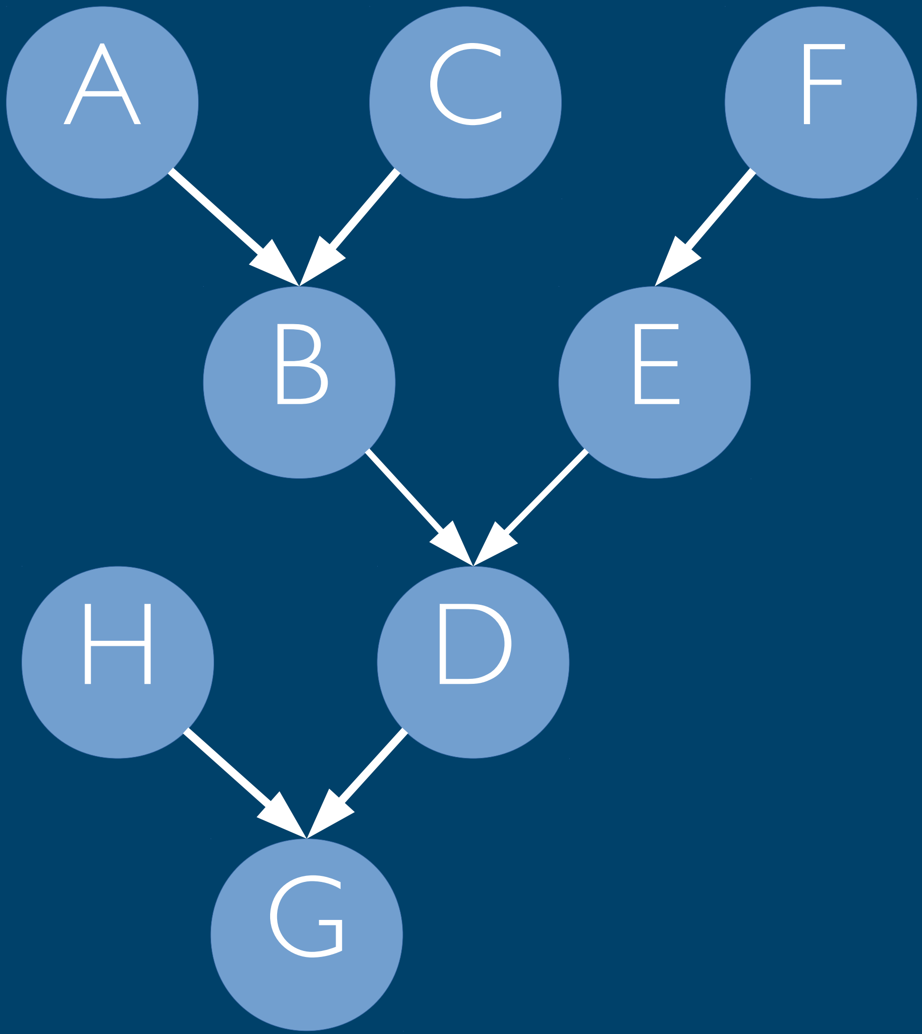
QUIZ (1/4)

Is…	Answer
$F \perp A$	
$F \perp A \mid D$	
$F \perp A \mid G$	
$F \perp A \mid H$	



QUIZ (1/4)

How many parameters are needed to represent this Bayes network?



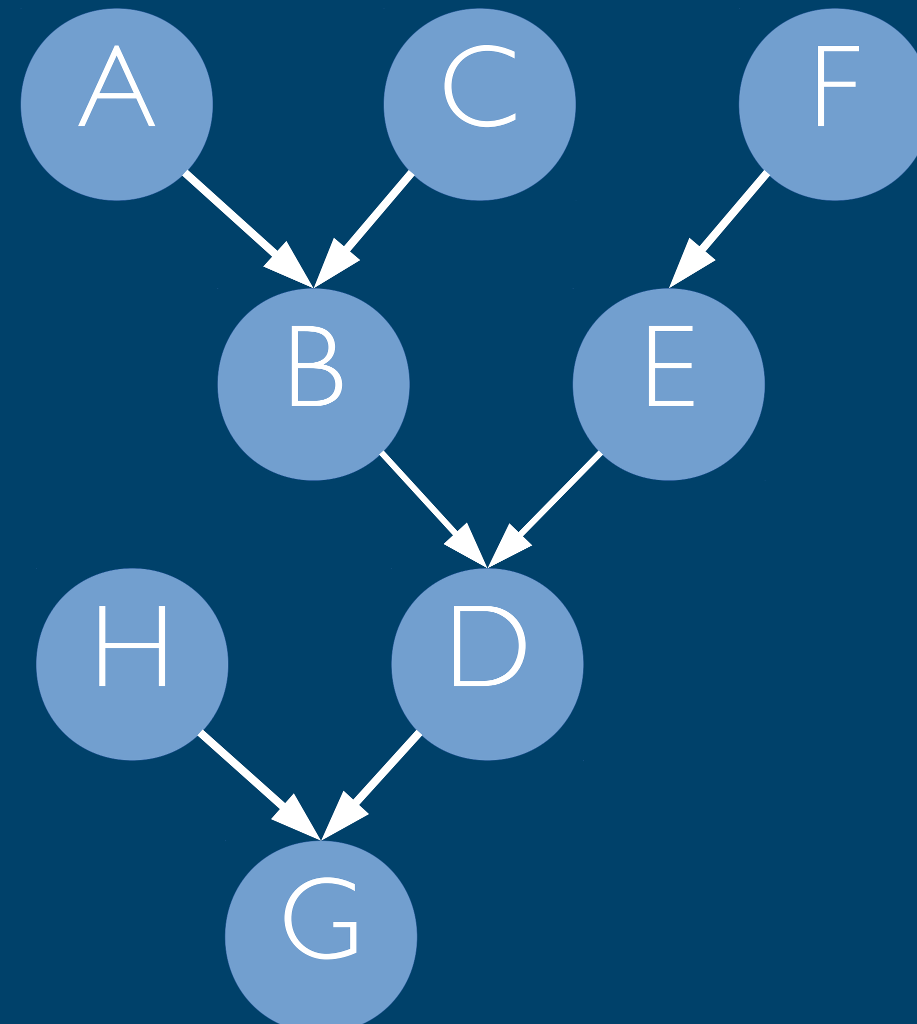
BONUS

Kung kayo ay isang bitwise operator, aling bitwise operator kayo, at bakit?

(Bitwise operators: and, or, exclusive or)

ANSWERS

Is…	Answer
$F \perp A$	Y
$F \perp A \mid D$	N
$F \perp A \mid G$	N
$F \perp A \mid H$	Y



QUIZ (1/4)

How many parameters are needed to represent this Bayes network?

18