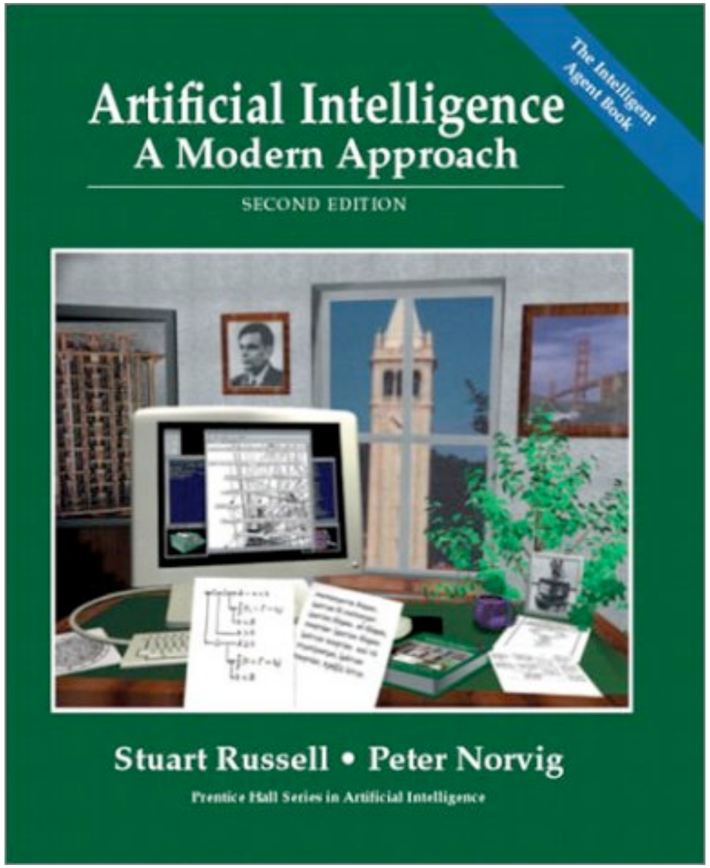


Bayesian Networks

Chapter 14

Section 1 – 2

Updating the Belief State

	toothache		¬toothache	
	pcatch	¬pcatch	pcatch	¬pcatch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- Let D now observe toothache with probability 0.8 (e.g., “the patient says so”)
- How should D update its belief state?

Updating the Belief State

	toothache		¬toothache	
	pcatch	¬pcatch	pcatch	¬pcatch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- Let E be the evidence such that $P(toothache|E) = 0.8$
- We want to compute $P(c \wedge t \wedge pc|E) = P(c \wedge pc|t,E)\ P(t|E)$
- Since E is not directly related to the cavity or the probe catch, we consider that c and pc are independent of E given t, hence: $P(c \wedge pc|t,E) = P(c \wedge pc|t)$

Updating the Belief State

	Toothache		¬Toothache	
	PCatch	¬PCatch	PCatch	¬PCatch
Cavity	~~0.108~~ 0.432	~~0.012~~ 0.048	~~0.072~~ 0.018	~~0.008~~ 0.002
¬Cavity	~~0.016~~ 0.064	~~0.064~~ 0.256	~~0.144~~ 0.036	~~0.576~~ 0.144

- Let E be the evidence such that P(Toothache|E) = 0.8
- t∧pc|E) = P(c∧pc|t,E) P(t|E)
- Since E is not directly related to the cavity or the probe catch, we … e independent of E given t, he … t)

To get these 4 probabilities we normalize their sum to 0.8

To get these 4 probabilities we normalize their sum to 0.2

Issues

- If a state is described by n propositions, then a belief space contains 2^n states (possibly, some have probability 0)
- → Modeling difficulty: many numbers must be entered in the first place
- → Computational issue: memory size and time

	toothache		¬toothache	
	pcatch	¬pcatch	pcatch	¬pcatch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- Toothache and pcatch are independent given cavity (or ¬cavity), but this relation is hidden in the numbers ! [Verify this]
- Bayesian networks explicitly represent independence among propositions to reduce the number of probabilities defining a belief state

Bayesian networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

- Syntax:
  - a set of nodes, one per variable
  - a directed, acyclic graph (link ≈ "directly influences")
  - a conditional distribution for each node given its parents:

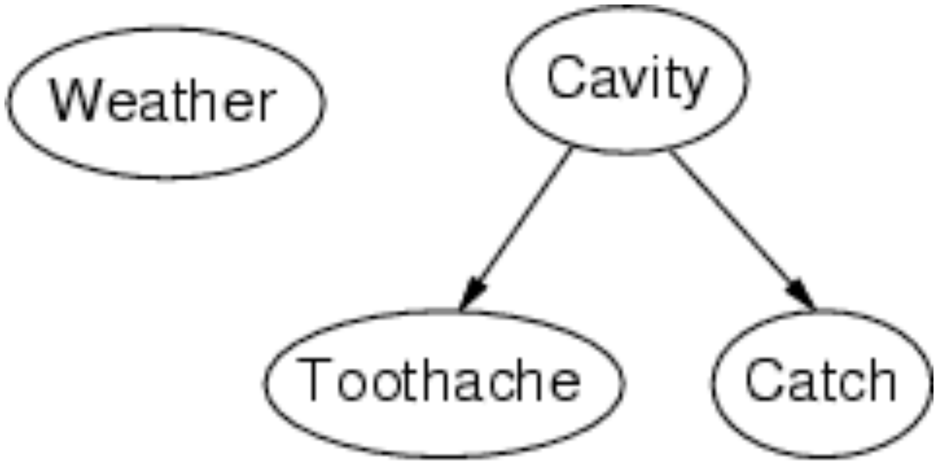
$$\mathbf{P}(X_i \mid \text{Parents}(X_i))$$

- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values

Following slides by Hwee Tou Ng (Singapore)

Example (1)

- Topology of network encodes conditional independence assertions:



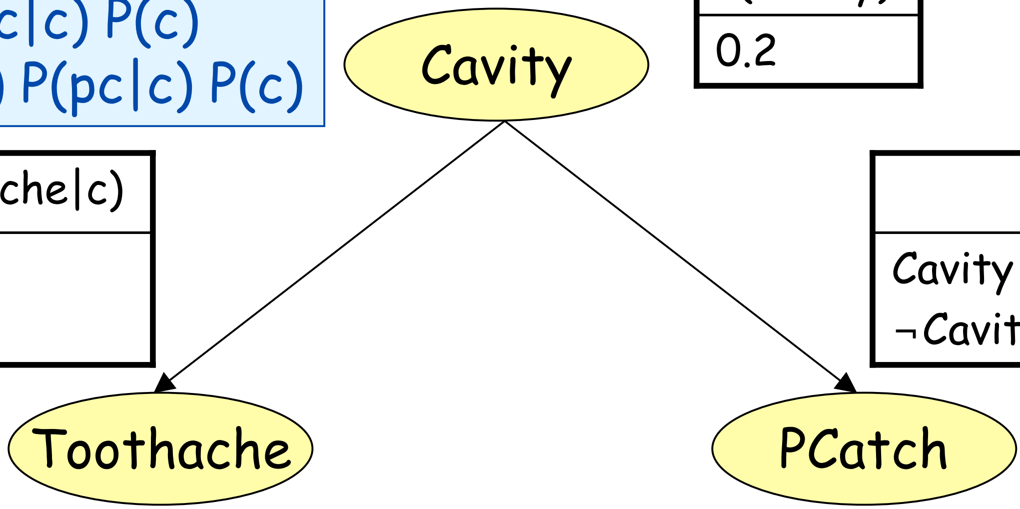
- *Weather* is independent of the other variables
- *Toothache* and *Catch* are conditionally independent given *Cavity*

Bayesian Network

- Notice that Cavity is the "cause" of both Toothache and PCatch, and represent the causality links explicitly
- Give the prior probability distribution of Cavity
- Give the conditional probability tables of Toothache and PCatch

$$\begin{aligned} P(c \wedge t \wedge pc) &= P(t \wedge pc|c)\,P(c) \\ &= P(t|c)\,P(pc|c)\,P(c) \end{aligned}$$

P(Cavity)
0.2



	P(Toothache|c)
Cavity	0.6
¬Cavity	0.1

	P(PCatch|c)
Cavity	0.9
¬Cavity	0.02

5 probabilities, instead of 7

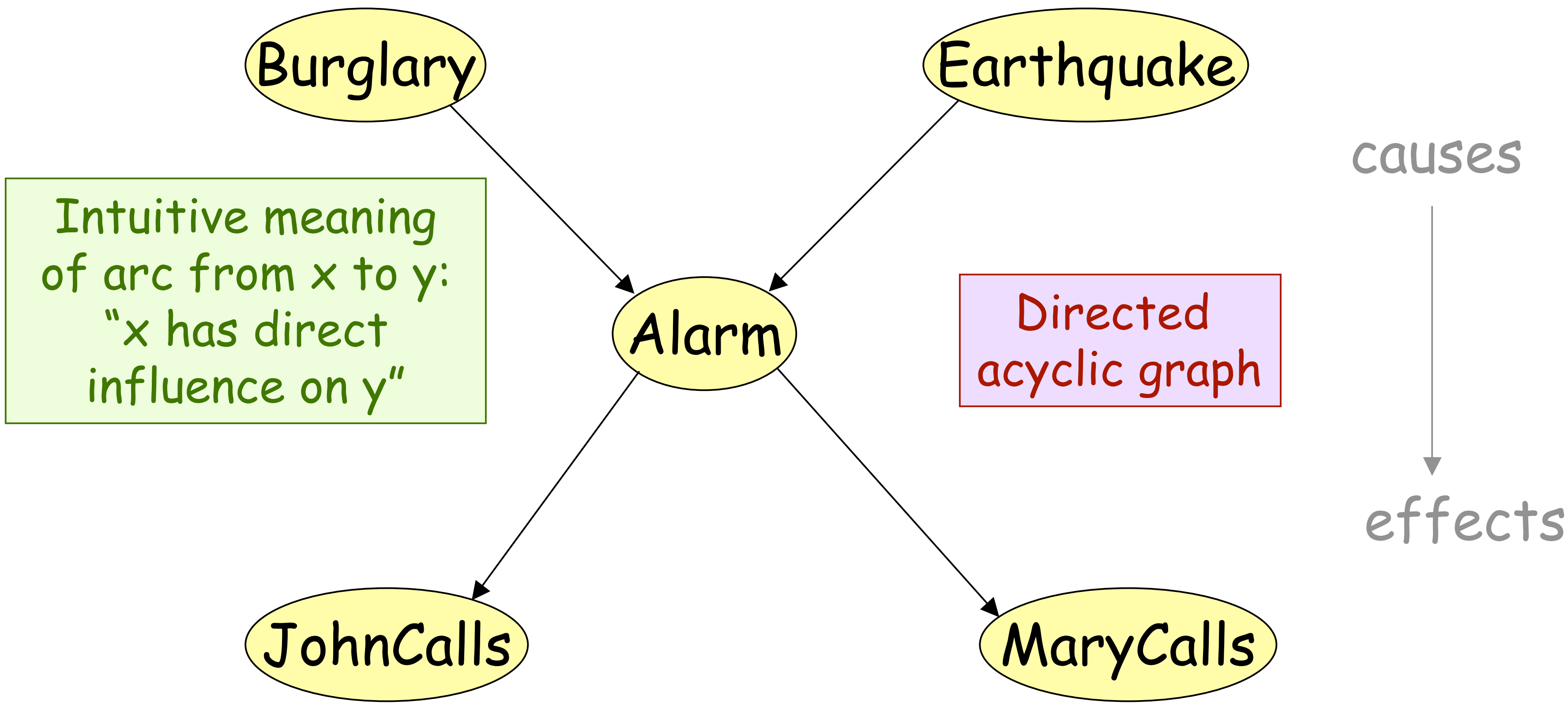
Example (2)

- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?

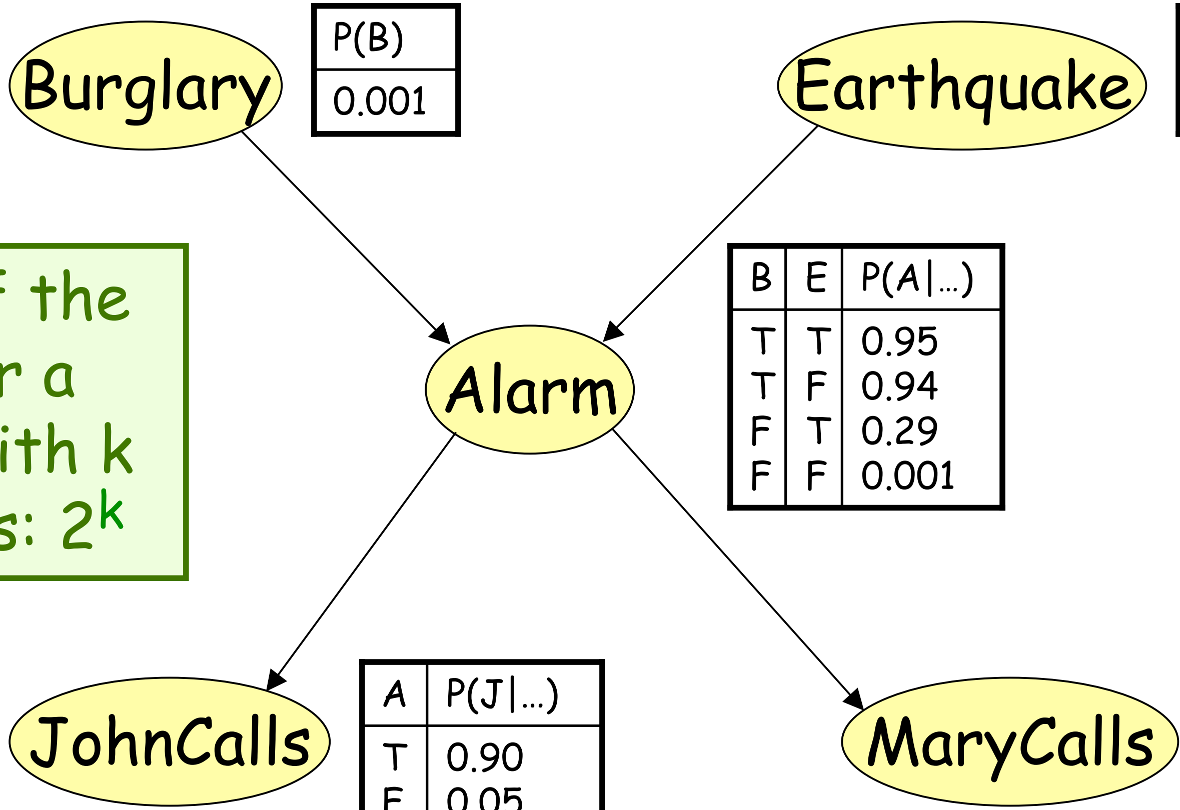
- Variables: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

- Network topology reflects "causal" knowledge:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call

A More Complex BN



A More Complex BN



P(B)
0.001

P(E)
0.002

B	E	P(A|…)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

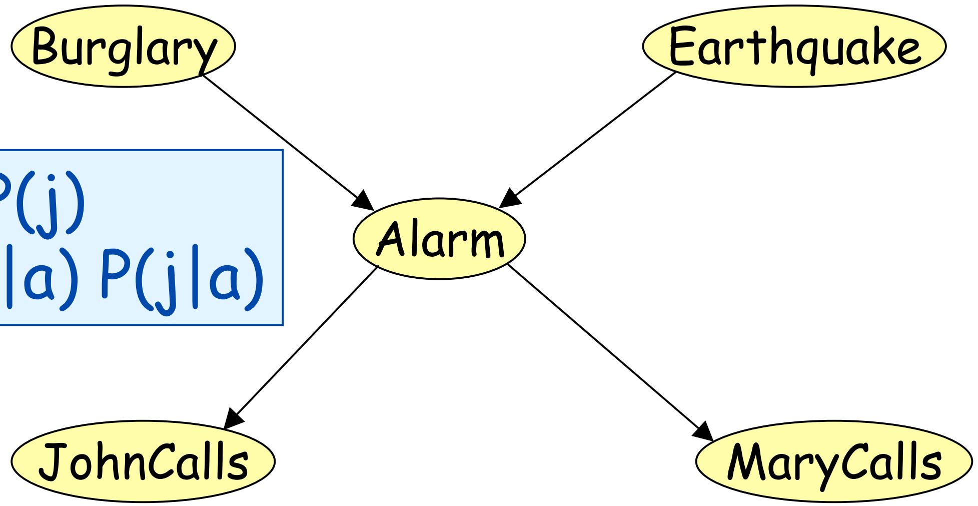
A	P(J|…)
T	0.90
F	0.05

A	P(M|…)
T	0.70
F	0.01

Size of the CPT for a node with k parents: 2^k

10 probabilities, instead of 31

What does the BN encode?

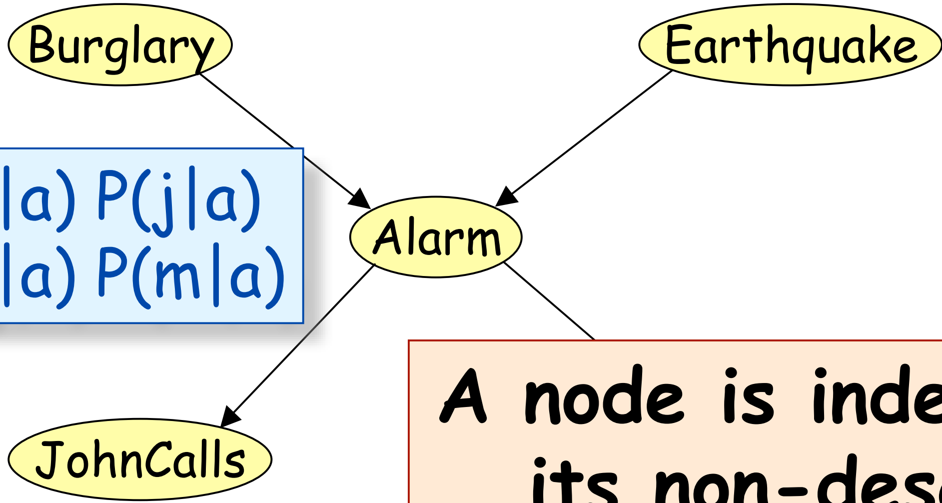


$P(b \wedge j) \neq P(b)\,P(j)$
$P(b \wedge j | a) = P(b|a)\,P(j|a)$

Each of the beliefs JohnCalls and MaryCalls is independent of Burglary and Earthquake given Alarm or $\neg$Alarm

For example, John does not observe any burglaries directly

What does the BN encode?



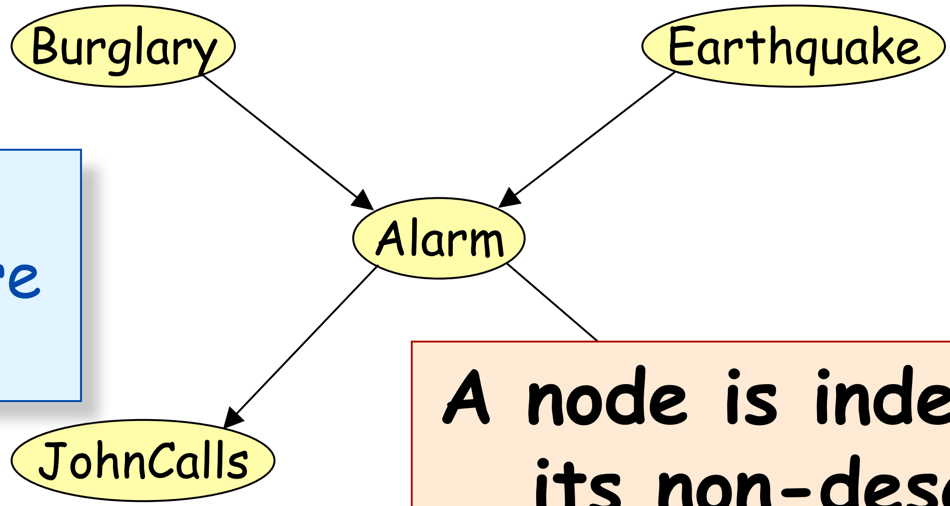
$P(b \wedge j \mid a) = P(b \mid a)\, P(j \mid a)$
$P(j \wedge m \mid a) = P(j \mid a)\, P(m \mid a)$

A node is independent of its non-descendants given its parents

The beliefs JohnCalls and MaryCalls are independent given Alarm or ¬Alarm

For instance, the reasons why John and Mary may not call if there is an alarm are unrelated

What does the BN encode?



Burglary and Earthquake are independent

A node is independent of its non-descendants given its parents

The beliefs JohnCalls and MaryCalls are independent given Alarm or ¬Alarm

For instance, the reasons why John and Mary may not call if there is an alarm are unrelated

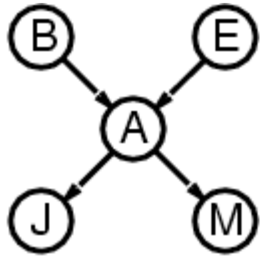
Locally Structured World

- A world is locally structured (or sparse) if each of its components interacts directly with relatively few other components
- In a sparse world, the CPTs are small and the BN contains much fewer probabilities than the full joint distribution
- If the # of entries in each CPT is bounded by a constant, i.e., O(1), then the # of probabilities in a BN is linear in n – the # of propositions – instead of 2^n for the joint distribution

Semantics

The full joint distribution is defined as the product of the local conditional distributions:

$$\mathbf{P}(X_1, \ldots, X_n) = \pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$$

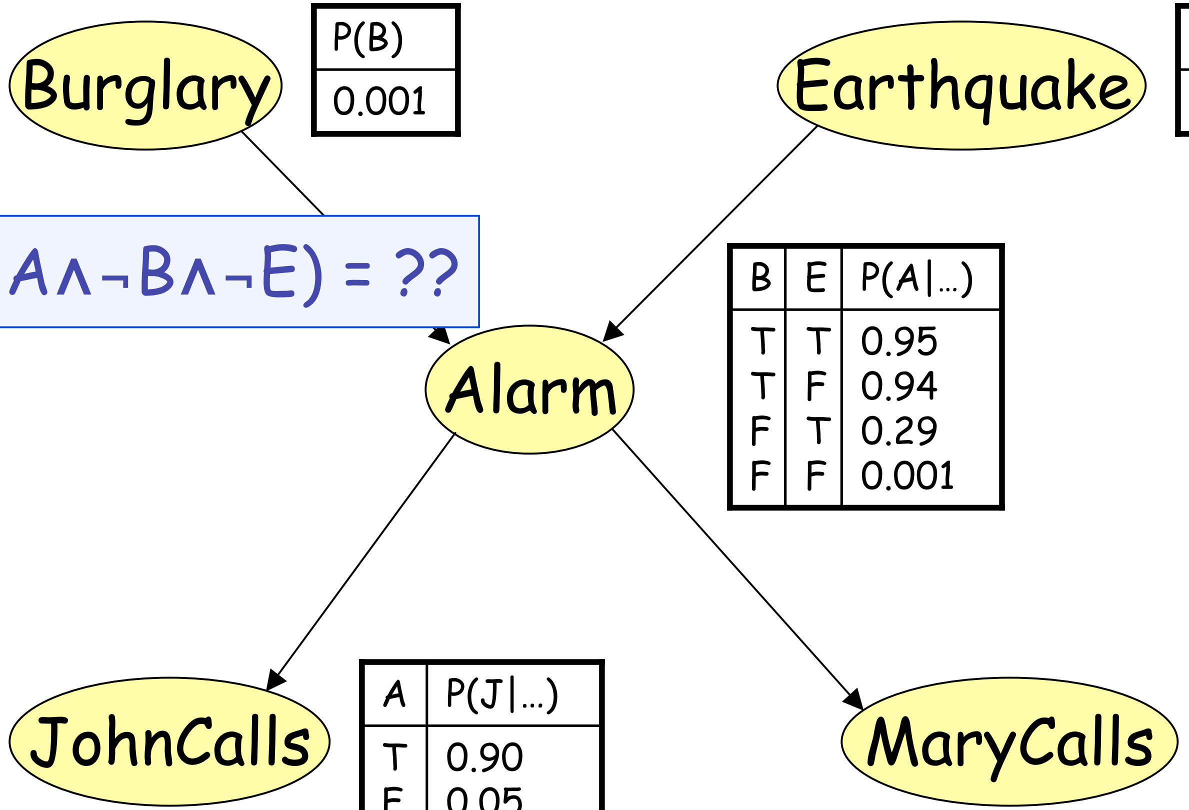


e.g., $\mathbf{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

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

Calculation of Joint Probability

$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) = ??$



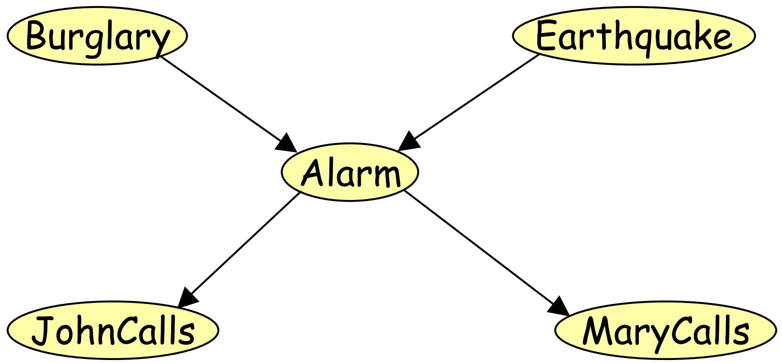
P(B)
0.001

P(E)
0.002

B	E	P(A|…)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

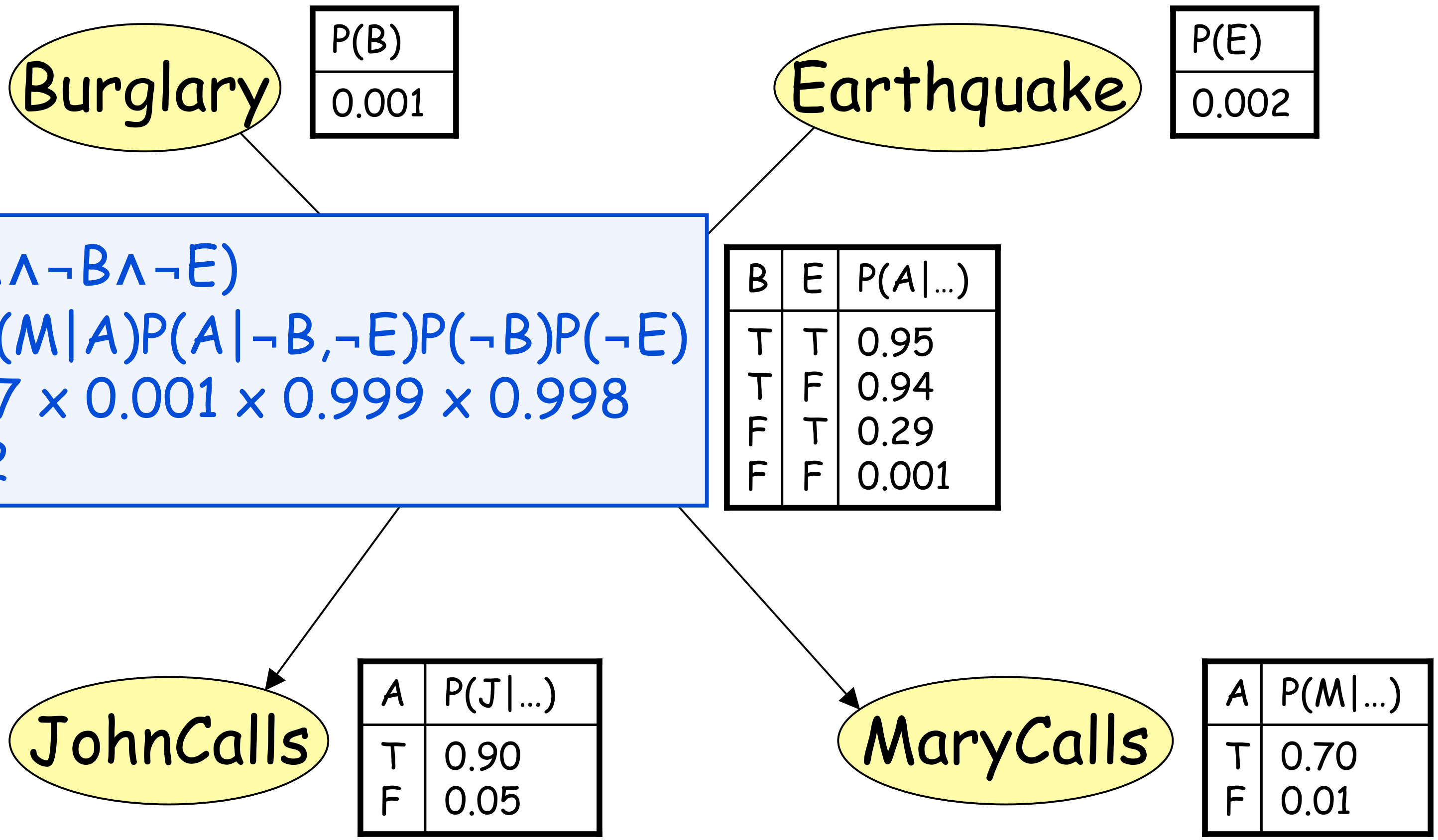
A	P(J|…)
T	0.90
F	0.05

A	P(M|…)
T	0.70
F	0.01



- $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$
  $= P(J \wedge M | A, \neg B, \neg E) \times P(A \wedge \neg B \wedge \neg E)$
  $= P(J | A, \neg B, \neg E) \times P(M | A, \neg B, \neg E) \times P(A \wedge \neg B \wedge \neg E)$
  (J and M are independent given A)
- $P(J | A, \neg B, \neg E) = P(J | A)$
  (J and $\neg B \wedge \neg E$ are independent given A)
- $P(M | A, \neg B, \neg E) = P(M | A)$
- $P(A \wedge \neg B \wedge \neg E) = P(A | \neg B, \neg E) \times P(\neg B | \neg E) \times P(\neg E)$
  $= P(A | \neg B, \neg E) \times P(\neg B) \times P(\neg E)$
  ($\neg B$ and $\neg E$ are independent)
- $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) = P(J | A) P(M | A) P(A | \neg B, \neg E) P(\neg B) P(\neg E)$

Calculation of Joint Probability



P(B)
0.001

P(E)
0.002

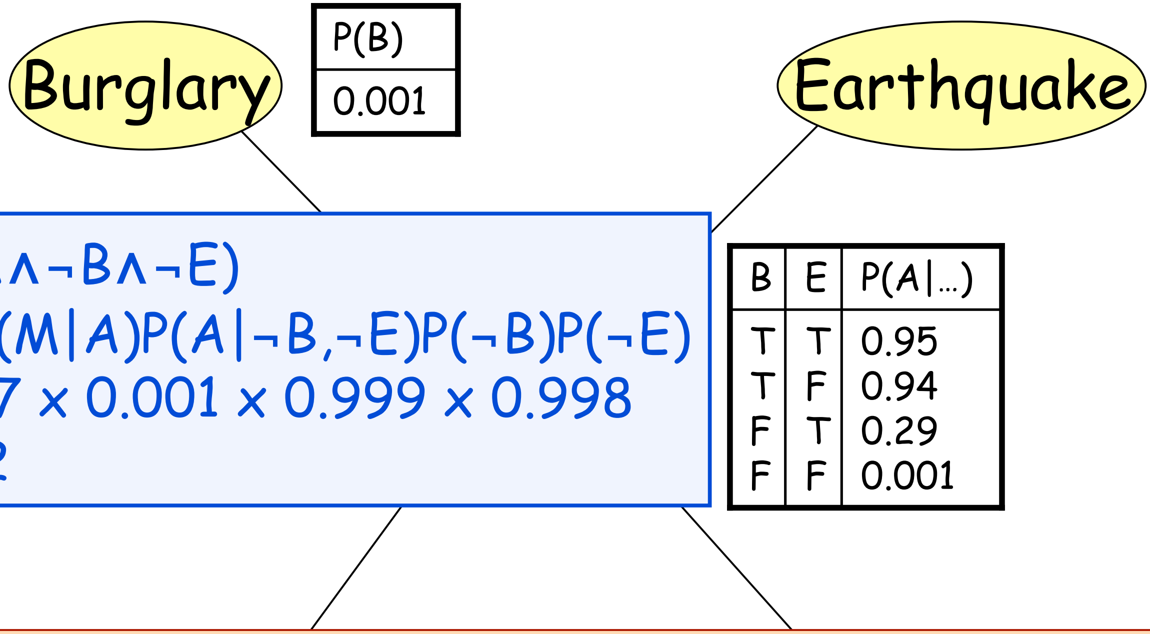
B	E	P(A|…)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|…)
T	0.90
F	0.05

A	P(M|…)
T	0.70
F	0.01

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$$
$$= P(J|A)P(M|A)P(A|\neg B,\neg E)P(\neg B)P(\neg E)$$
$$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$$
$$= 0.00062$$

Calculation of Joint Probability



P(B)
0.001

P(E)
0.002

B	E	P(A|…)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

$P(J \land M \land A \land \neg B \land \neg E)$
$= P(J|A)P(M|A)P(A|\neg B,\neg E)P(\neg B)P(\neg E)$
$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$
$= 0.00062$

$$P(x_1 \land x_2 \land \ldots \land x_n) = \Pi_{i=1,\ldots,n} P(x_i|\text{parents}(X_i))$$

→ full joint distribution table

Calculation of Joint Probability

Burglary

P(B)
0.001

$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$
$= P(J|A)P(M|A)P(A|\neg B,\neg E)$
$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$
$= 0.00062$

Since a BN defines the full joint distribution of a set of propositions, it represents a belief space

T	F	0.94
F	T	0.29
F	F	0.001

$$P(x_1 \wedge x_2 \wedge \ldots \wedge x_n) = \Pi_{i=1,\ldots,n} P(x_i|\text{parents}(X_i))$$

→ full joint distribution table

Constructing Bayesian networks

- 1. Choose an ordering of variables $X_1, \dots, X_n$
- 2. For i = 1 to n
  - add X_i to the network
  - select parents from $X_1, \dots, X_{i-1}$ such that

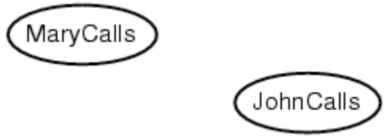
$$\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1, \dots X_{i-1})$$

This choice of parents guarantees:

$$\begin{aligned}\mathbf{P}(X_1, \dots, X_n) &= \pi_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \dots, X_{i-1}) \text{ (chain rule)} \\ &= \pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i)) \text{ (by construction)}\end{aligned}$$

Example

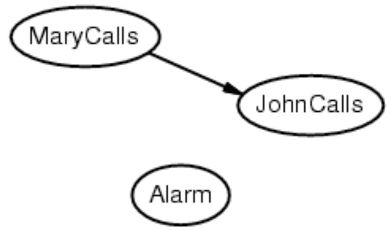
- Suppose we choose the ordering *M, J, A, B, E*



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

Example

- Suppose we choose the ordering *M, J, A, B, E*

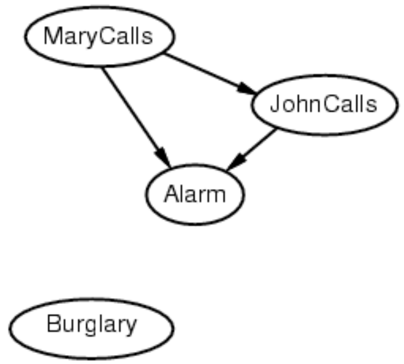


$\mathbf{P}(J \mid M) = \mathbf{P}(J)$? **No**

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$?

Example

- Suppose we choose the ordering *M, J, A, B, E*



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$? **No**

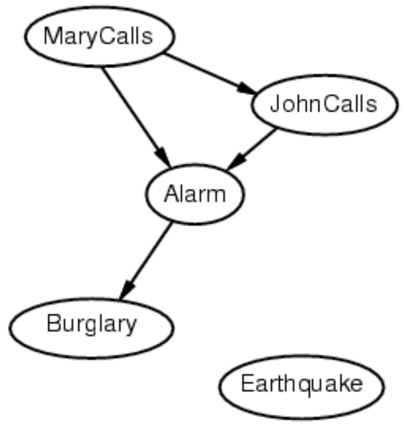
$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$?

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$?

Example

- Suppose we choose the ordering M, J, A, B, E



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$? **No**

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$? **Yes**

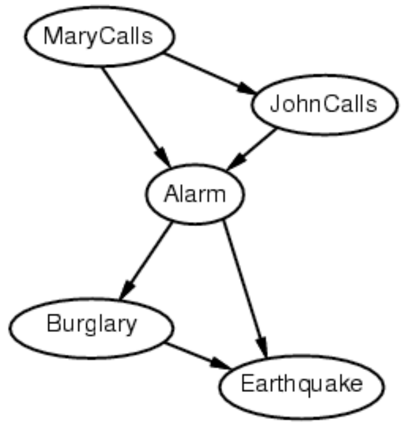
$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$? **No**

$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A)$?

$\mathbf{P}(E \mid B, A, J, M) = \mathbf{P}(E \mid A, B)$?

Example

- Suppose we choose the ordering M, J, A, B, E



$\boldsymbol{P}(J \mid M) = \boldsymbol{P}(J)$? **No**

$\boldsymbol{P}(A \mid J, M) = \boldsymbol{P}(A \mid J)$? $\boldsymbol{P}(A \mid J, M) = \boldsymbol{P}(A)$? **No**

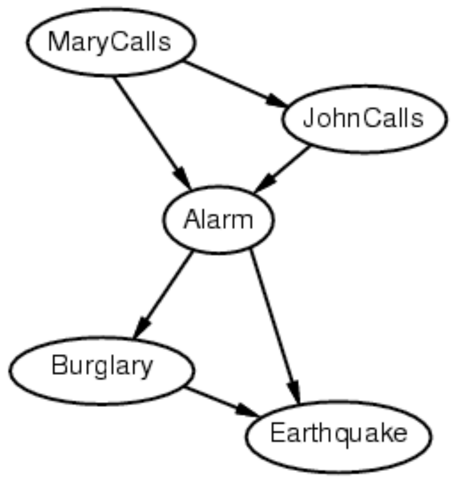
$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B \mid A)$? **Yes**

$\boldsymbol{P}(B \mid A, J, M) = \boldsymbol{P}(B)$? **No**

$\boldsymbol{P}(E \mid B, A ,J, M) = \boldsymbol{P}(E \mid A)$? **No**

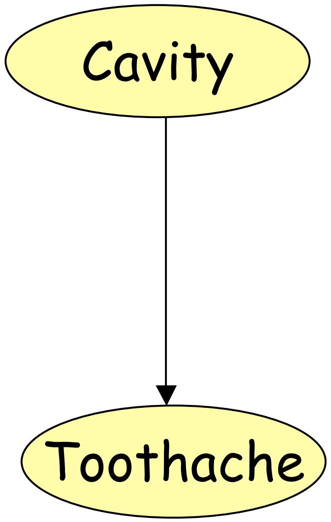
$\boldsymbol{P}(E \mid B, A, J, M) = \boldsymbol{P}(E \mid A, B)$? **Yes**

Example contd.



- Deciding conditional independence is hard in noncausal directions
- (Causal models and conditional independence seem hardwired for humans!)
- Network is less compact: 1 + 2 + 4 + 2 + 4 = 13 numbers needed

Querying the BN



P(C)
0.1

C	P(T|c)
T	0.4
F	0.01111

- The BN gives $P(t|c)$
- What about $P(c|t)$?
- $P(Cavity|t)$
  $= P(Cavity \wedge t)/P(t)$
  $= P(t|Cavity)\, P(Cavity) / P(t)$
  [Bayes' rule]
- $P(c|t) = \alpha\, P(t|c)\, P(c)$
- Querying a BN is just applying the trivial Bayes' rule on a larger scale

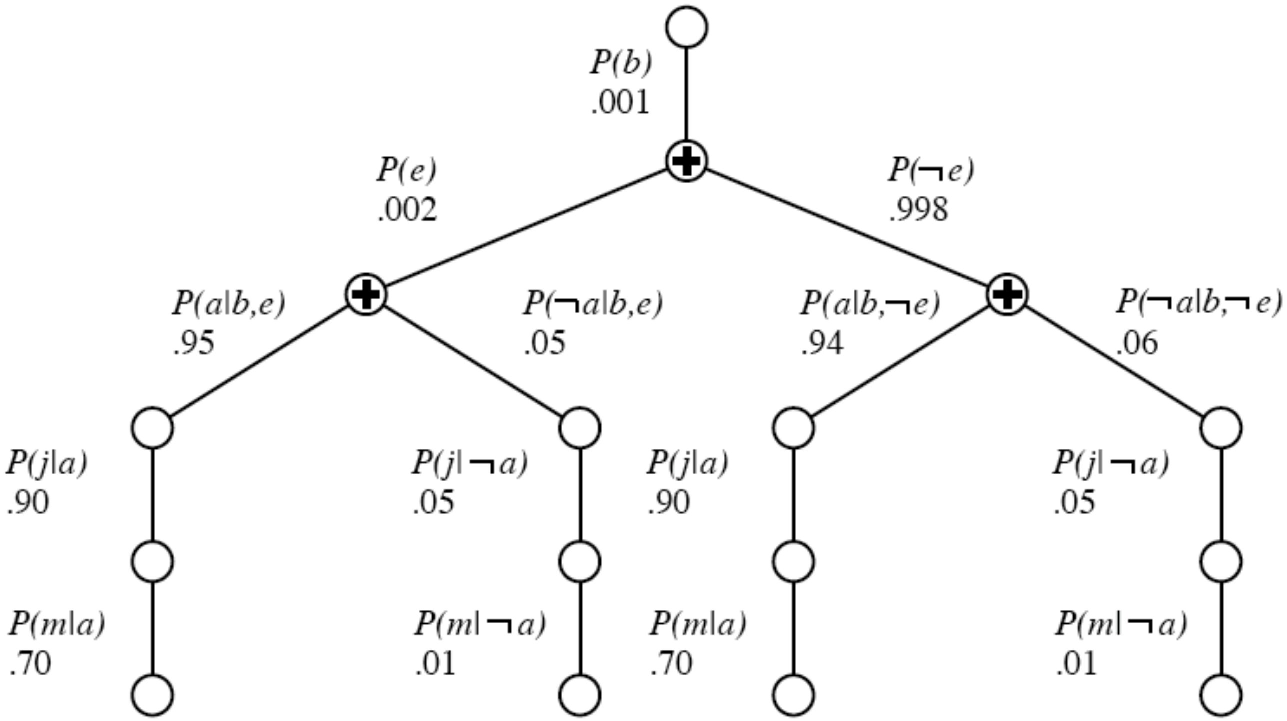
Slides:
Jean Claude Latombe

Querying the BN

- New evidence E indicates that JohnCalls with some probability p
- We would like to know the posterior probability of the other beliefs, e.g. P(Burglary|E)
- $P(B|E) = P(B \wedge J|E) + P(B \wedge \neg J|E)$
  $= P(B|J,E)\ P(J|E) + P(B\ |\neg J,E)\ P(\neg J|E)$
  $= P(B|J)\ P(J|E) + P(B|\neg J)\ P(\neg J|E)$
  $= p\ P(B|J) + (1\text{-}p)\ P(B|\neg J)$
- We need to compute P(B|J) and $P(B|\neg J)$

Querying the BN

- $P(b|J) = \alpha\, P(b \wedge J)$
  $= \alpha\, \Sigma_m \Sigma_a \Sigma_e P(b \wedge J \wedge m \wedge a \wedge e)$ [marginalization]
  $= \alpha\, \Sigma_m \Sigma_a \Sigma_e P(b)P(e)P(a|b,e)P(J|a)P(m|a)$ [BN]
  $= \alpha\, P(b)\Sigma_e P(e)\Sigma_a P(a|b,e)P(J|a)\Sigma_m P(m|a)$ [re-ordering]

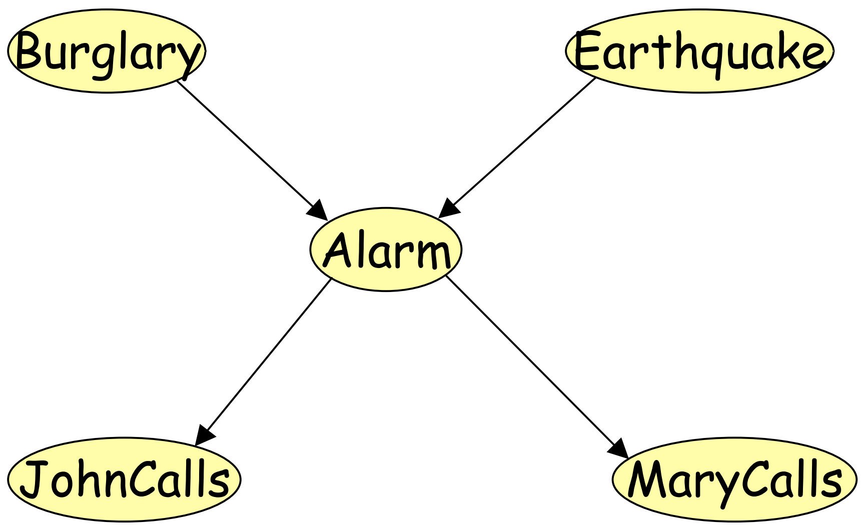


Querying the BN

- Depth-first evaluation of P(b|J) leads to computing each of the 4 following products twice:
  $P(J|A)\,P(M|A)$, $P(J|A)\,P(\neg M|A)$, $P(J|\neg A)\,P(M|\neg A)$, $P(J|\neg A)\,P(\neg M|\neg A)$
- Bottom-up (right-to-left) computation + caching – e.g., variable elimination algorithm (see R&N) – avoids such repetition
- For singly connected BN, the computation takes time linear in the total number of CPT entries (→ time linear in the # propositions if CPT's size is bounded)

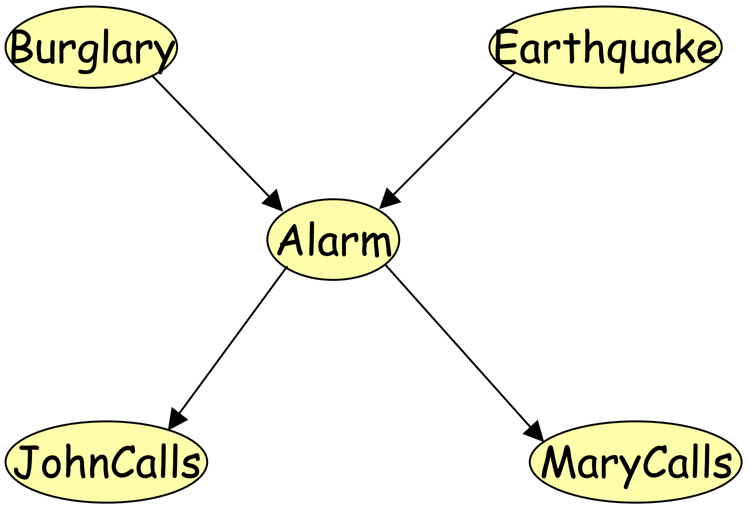
Singly Connected BN

A BN is singly connected if there is at most one undirected path between any two nodes



is singly connected

Comparison to Classical Logic



$Burglary \rightarrow Alarm$

$Earthquake \rightarrow Alarm$

$Alarm \rightarrow JohnCalls$

$Alarm \rightarrow MaryCalls$

If the agent observes
$\neg$JohnCalls,
it infers $\neg$Alarm,
$\neg$Burglary, and $\neg$Earthquake

If it observes JohnCalls, then it infers nothing

Summary

- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + CPTs = compact representation of joint distribution
- Generally easy for domain experts to construct