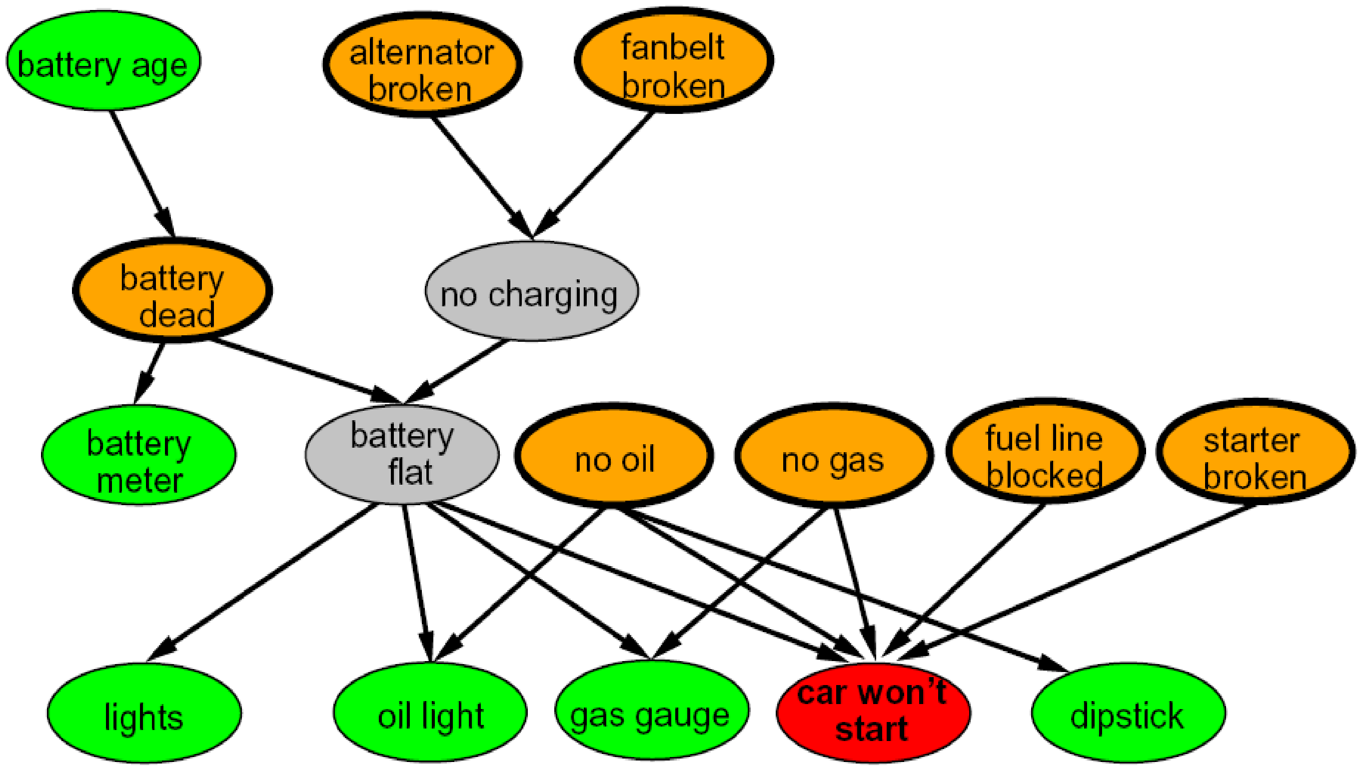
CSCI 3202: Intro to Artificial Intelligence

Introduction to Bayesian Networks

Rhonda Hoenigman
Department of Computer Science

Exam grades by Monday

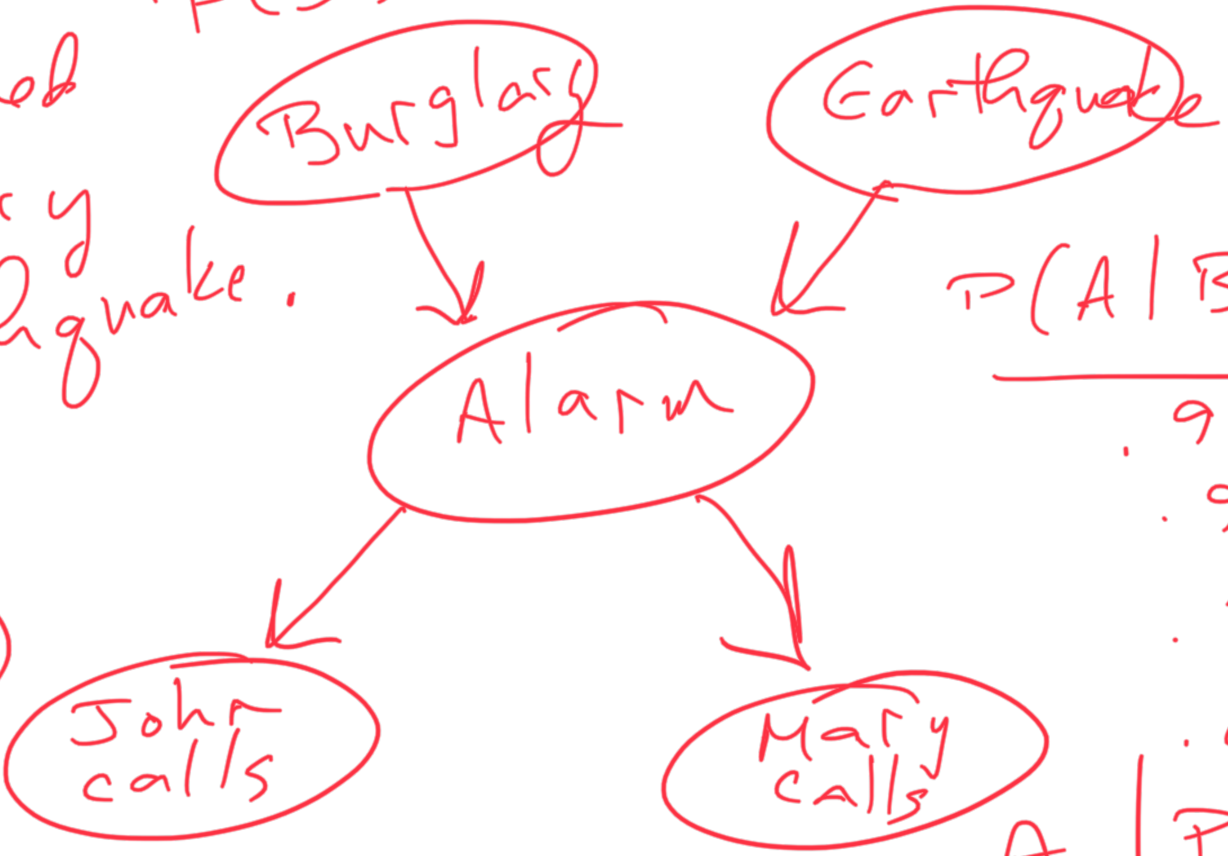


Bayes Nets

Alarm is conditioned on burglary and earthquake.

$P(B) = .001$

$P(E) = .002$



$P(A \mid B, E)$	B	E
.95	T	T
.94	T	F
.29	F	T
.001	F	F

A	$P(J \mid A)$
T	.90
F	.05

A	$P(M \mid A)$
T	.70
F	.01

Bayes Nets

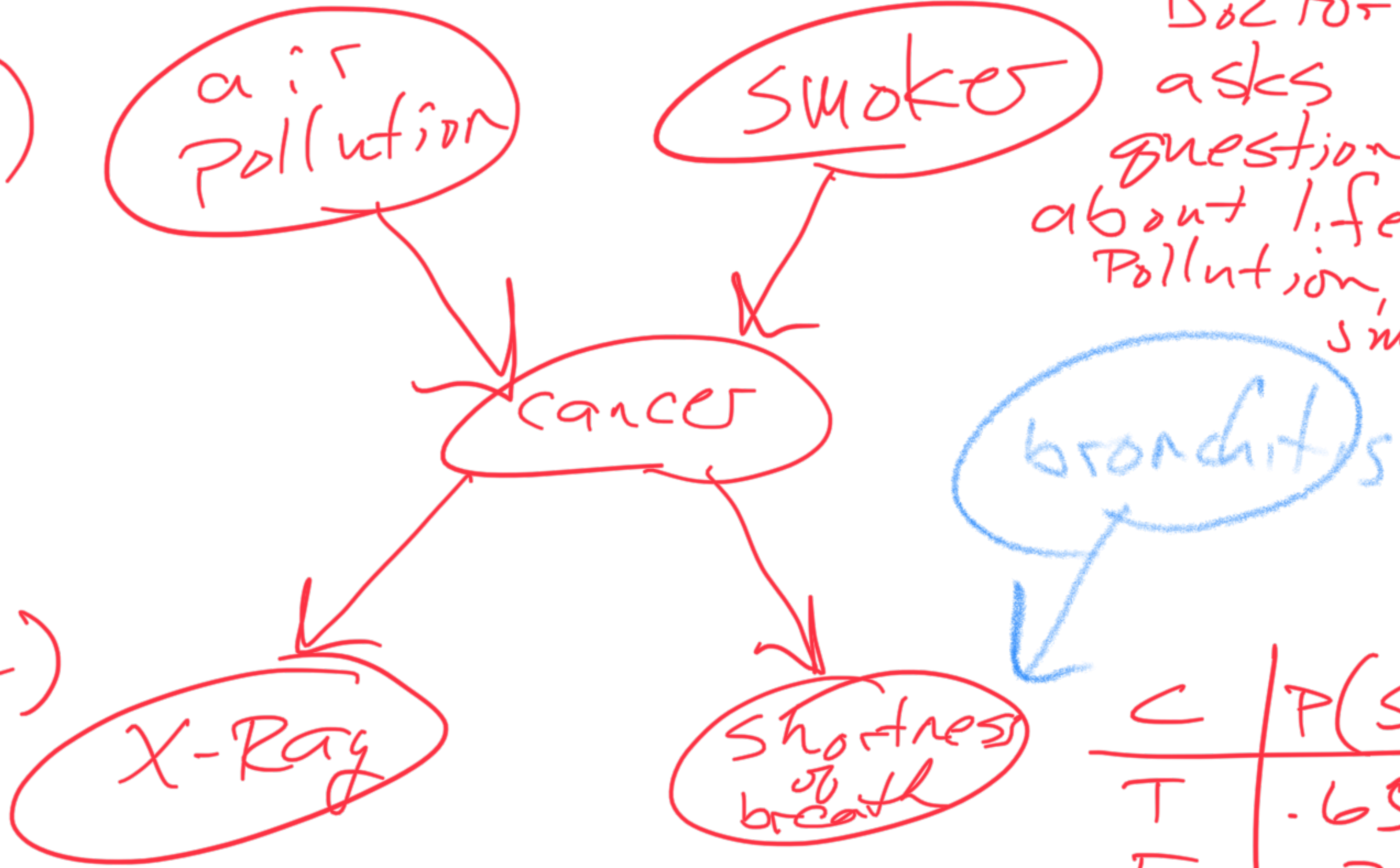
← Person goes to doctor complaining of shortness of breath.

$P(P=L)=.9$

$P(S=T)=.3$

Doctor asks questions about lifestyle. Pollution, smokes

P	S	$P(c \mid P,S)$
H	T	.05
H	F	.02
L	T	.03
L	F	.001



C	$P(X=pos \mid c)$
T	.9
F	.2

C	$P(S \mid c)$
T	.65
F	.30

Bayes Nets – dealing with Uncertainty

Probabilistic reasoning framework for managing uncertain beliefs and knowledge.

a.k.a - belief network, decision network

In general:

- **Observed variables (evidence)**: Agent knows certain things about the state of the world (e.g. sensor readings or symptoms) X-Ray, Breathing, Alarm
- **Unobserved variables**: Agent needs to reason about other aspects (e.g. where an object is or what disease is present) Cancer
- **Model**: Agent knows something about how the known variables relate to the unknown variables smoking increases cancer likelihood.

Bayesian Networks

The point of Bayes nets is to **represent full joint probability distributions**, *and*

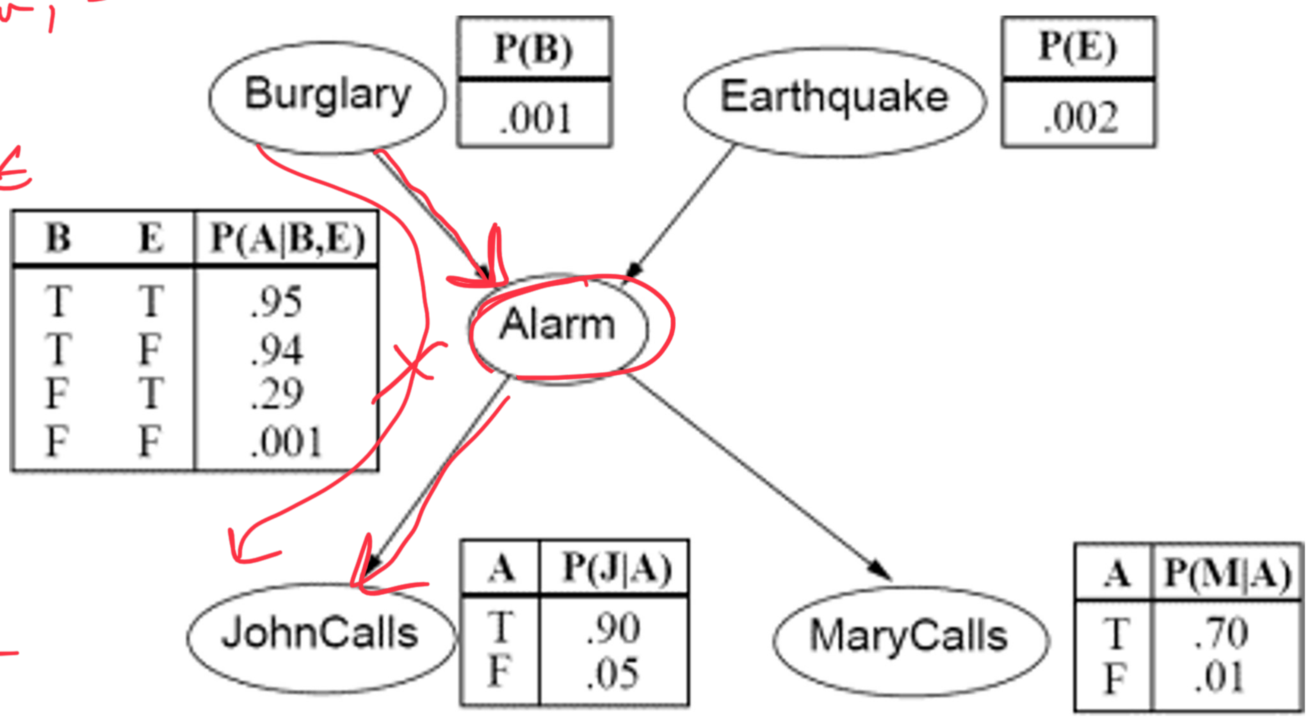
to encode an interrelated set of **conditional independence/probability statements**

- Directed, acyclic graph
- Consists of **nodes** (events), and

  e.g. Alarm, smoking
- Each edge is conditional dependency

  Alarm conditioned on B, E
- **conditional probability tables (CPTs)**, relating those events
- Describe **local** variable interactions

  Alarm → J
- Chain together local interactions to estimate **global**, **indirect** interactions   B → J, B → A, → J



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

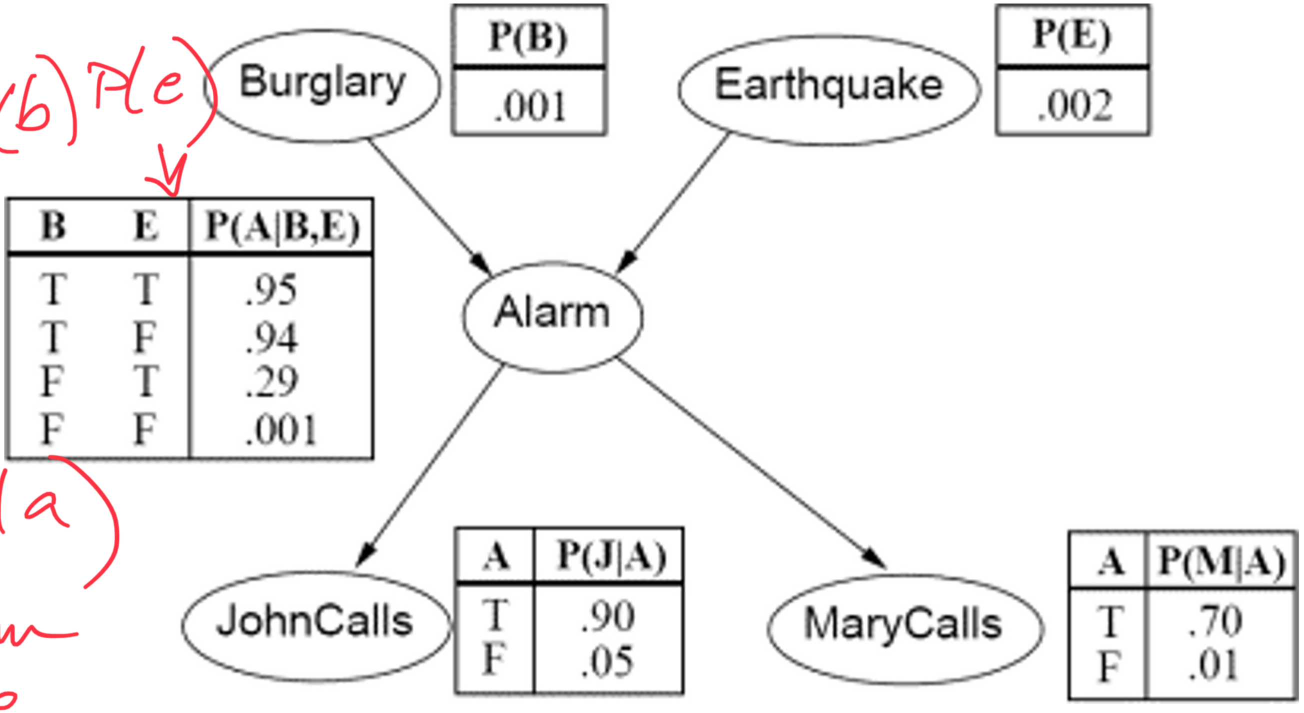
A	P(M|A)
T	.70
F	.01

Bayesian Networks

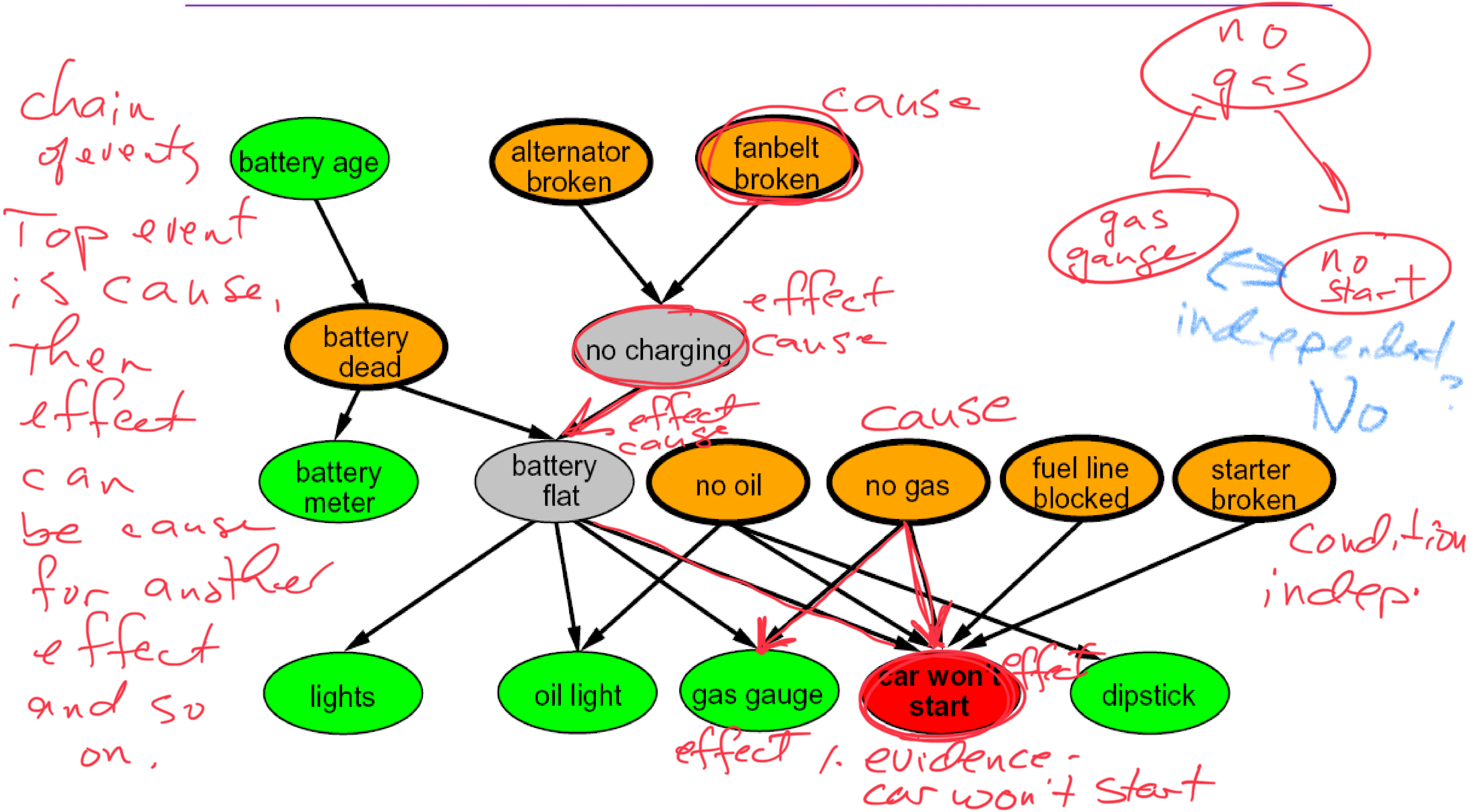
The point of Bayes nets is to **represent full joint probability distributions**, *and*

to encode an interrelated set of **conditional independence/probability statements**

- P(a | b, e) ~ given in CPT
- P(a) = $\sum_b \sum_e P(a|b,e)P(b)P(e)$
- P(b | a) CPT gives P(a|b) Bayes Rule to P(b|a)
- P(J | b) ~ sum over Alarm hidden variable



Graphical Model Notation – more complicated example



Independence and conditional independence

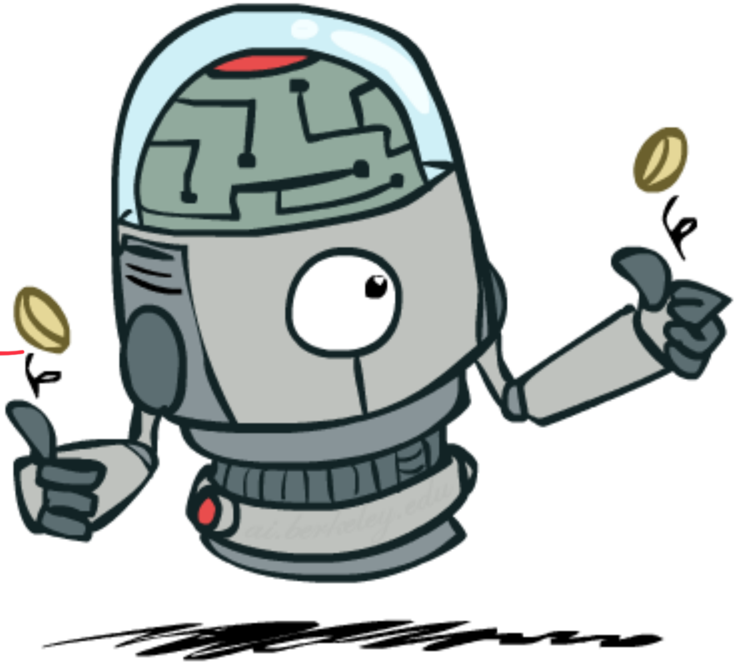
- Two variables are *independent* if:

$$\forall x, y : P(x, y) = P(x)P(y)$$

  - This says that their joint distribution *factors* into a product two simpler distributions
  - Another form:

$$\forall x, y : P(x|y) = P(x)$$

Knowing y doesn't change belief about x

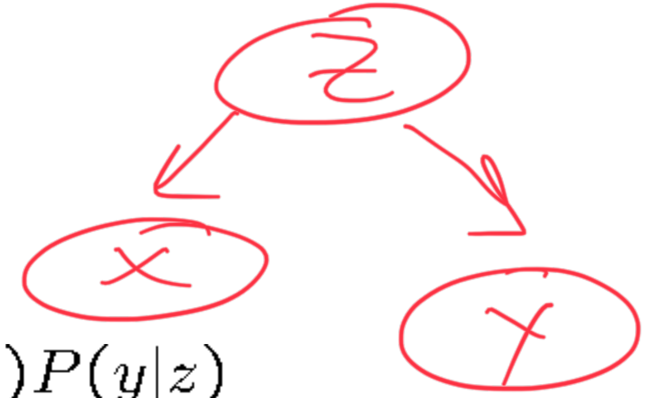


- Independence is a simplifying *modeling assumption*
  - *Empirical* joint distributions: at best "close" to independent
  - What could we assume for {Weather, Traffic, Cavity, Toothache}?

Conditional Independence

- Unconditional (absolute) independence very rare (why?)

- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.



- X is conditionally independent of Y given Z

if and only if: $\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$

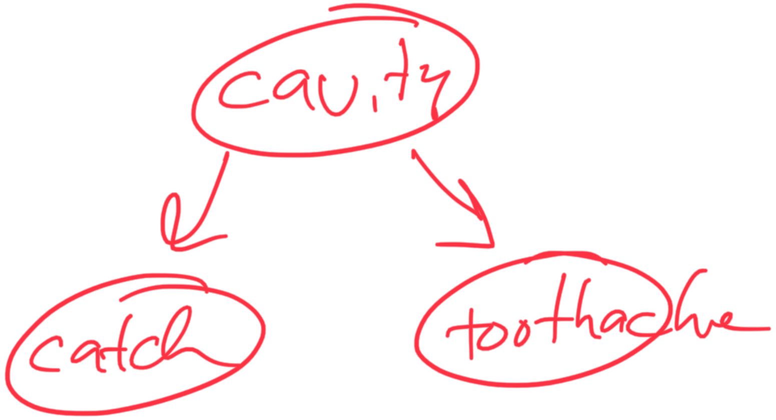
J calls and Mary calls are indep once alarm is observed.

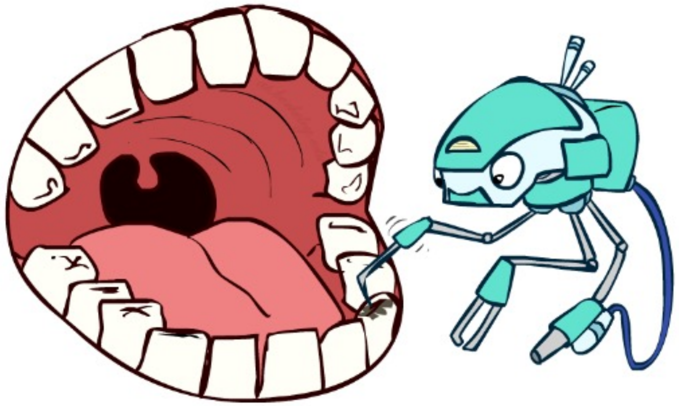
or, equivalently, if and only if $\forall x, y, z : P(x|z, y) = P(x|z)$

Knowing that Mary called doesn't change our belief about John calling once we've observed Alarm.

Conditional Independence

- P(Toothache, Cavity, Catch)
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +toothache, +cavity) = P(+catch | +cavity)
- The same independence holds if I don't have a cavity:
  - P(+catch | +toothache, -cavity) = P(+catch| -cavity)
- Catch is *conditionally independent* of Toothache given Cavity:
  - P(Catch | Toothache, Cavity) = P(Catch | Cavity)





- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)
  - One can be derived from the other easily

Conditional Independence examples

$P(\text{rain}, \text{traffic}, \text{umbrella})$

u and t are indep given rain.

$u \perp\!\!\!\perp t \mid r$

notation of conditional indep

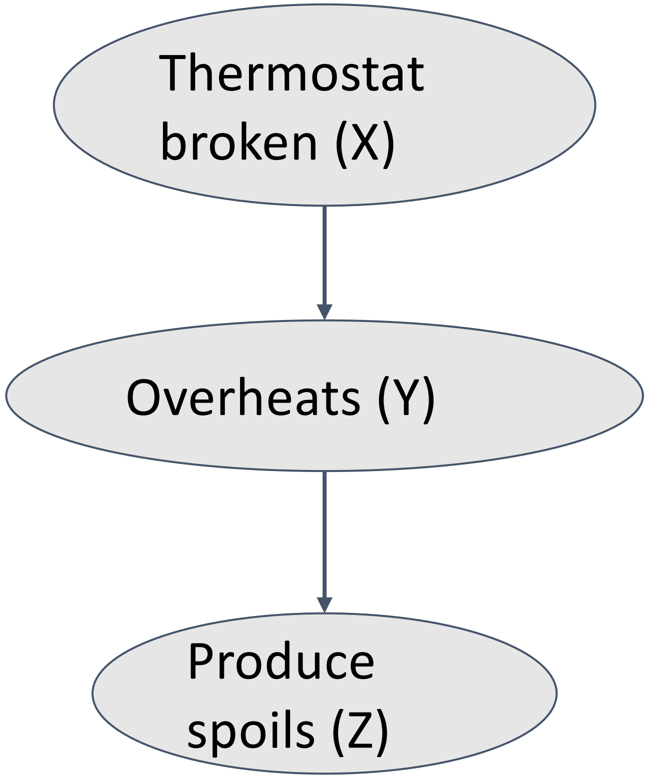
Bayesian Networks: Canonical Cases

Important Bayes net question: Are two nodes independent *given* certain evidence?

- If yes -- can prove using algebra
- If no -- can prove using a counterexample

Example: Are X and Z necessarily independent?

No!



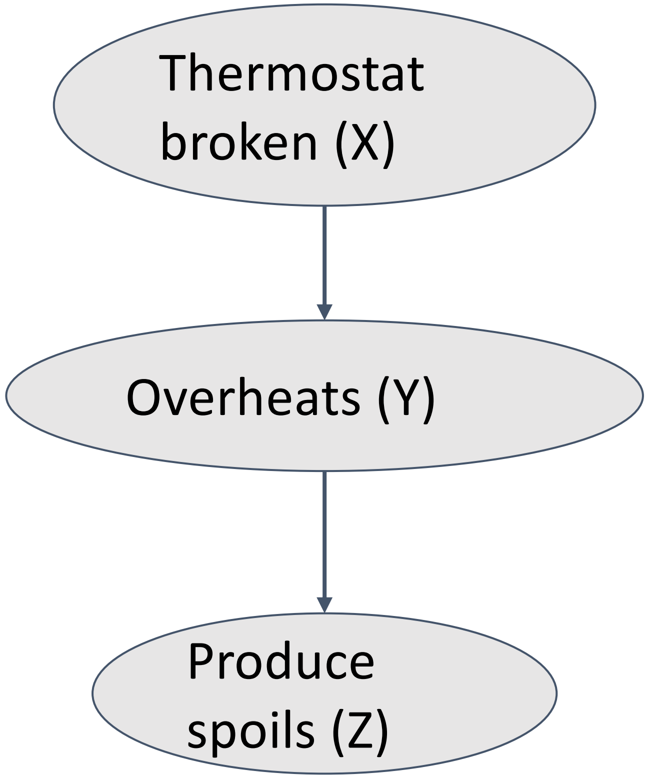
Bayesian Networks: Canonical Cases

Important Bayes net question: Are two nodes independent *given* certain evidence?

- If yes -- can prove using algebra
- If no -- can prove using a counterexample

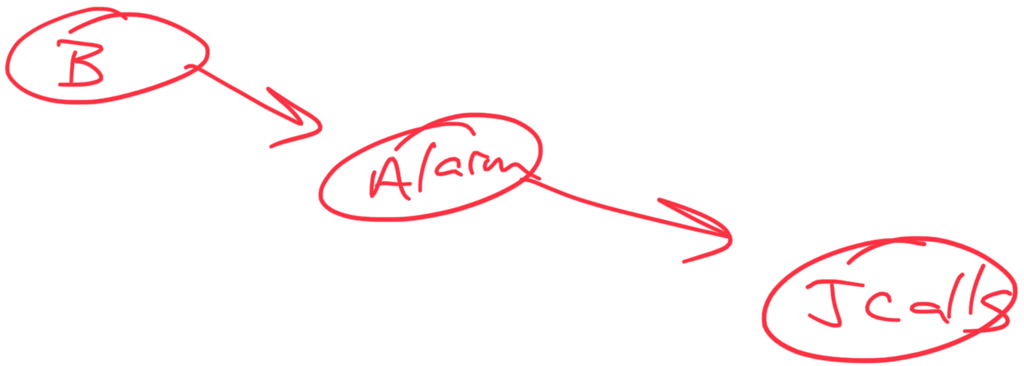
causal chain

Example: Are X and Z necessarily independent?



No!

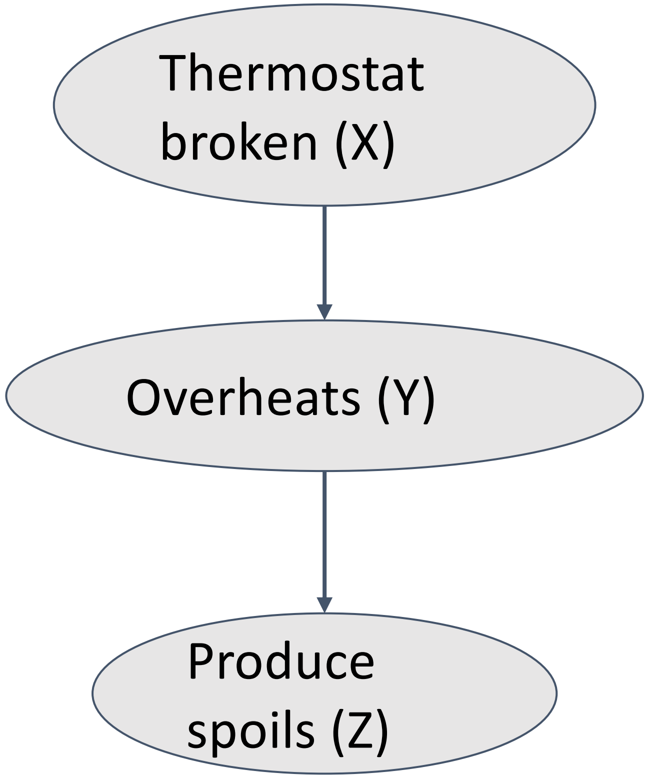
- X certainly influences Y, which influences Z
- Also, knowledge of Z influences beliefs about X (through Y)



Bayesian Networks: Canonical Cases

Example, rebooted: What about X and Z, ***given*** Y?

This is a canonical case is called a **causal chain**



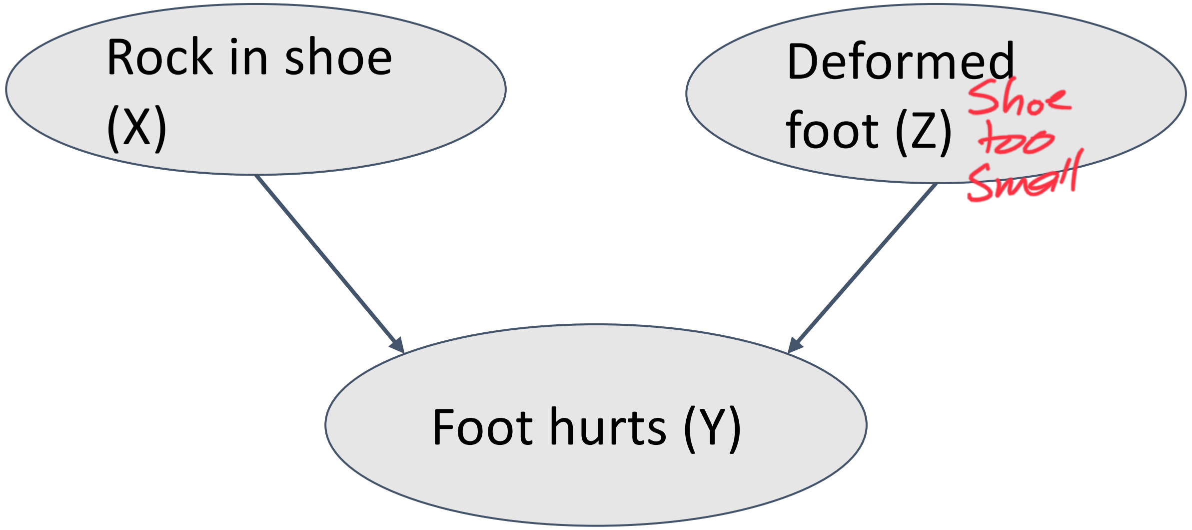
Bayesian Networks: Canonical Cases

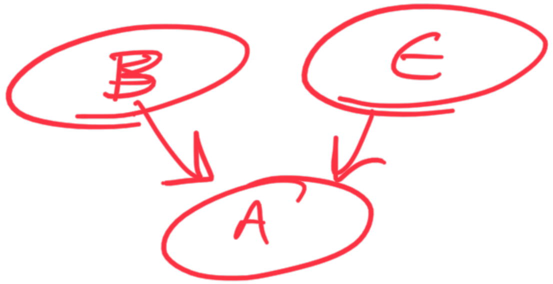
Common effect is the ~~third~~ next canonical case.

→ One effect, two possible causes

Example: Are X and Z independent? yes

Are X and Z independent ***given*** Y? not conditional indep.

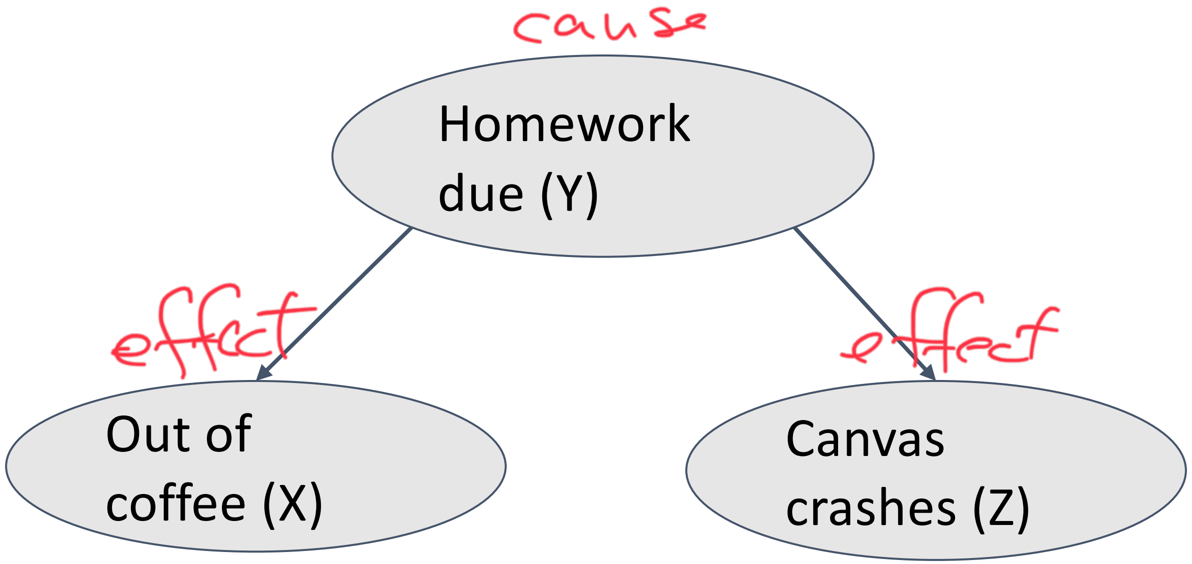




Bayesian Networks: Canonical Cases

Common cause is another canonical case.

→ Two effects, from the same cause



Example: Are X and Z independent?

Not necessarily

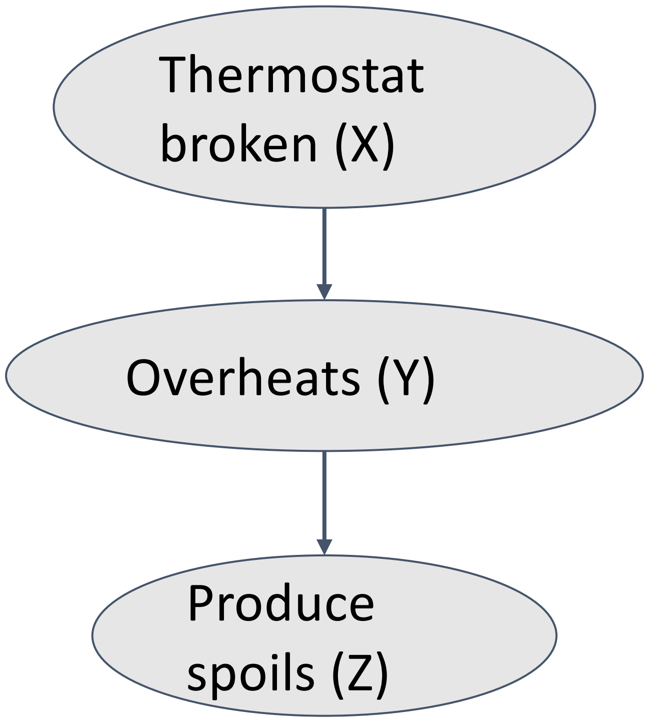
$P(Y|X,Z)$

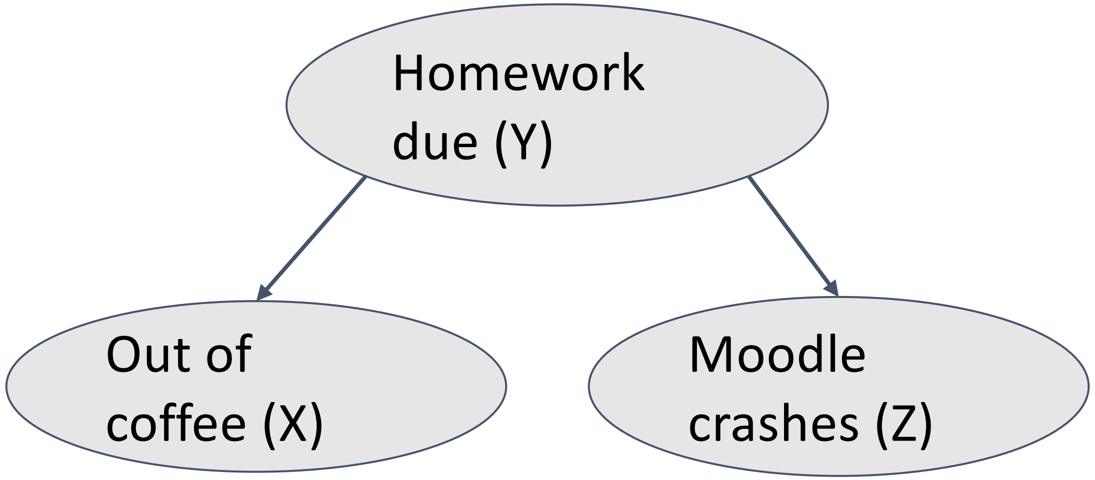
Are X and Z independent ***given*** Y?

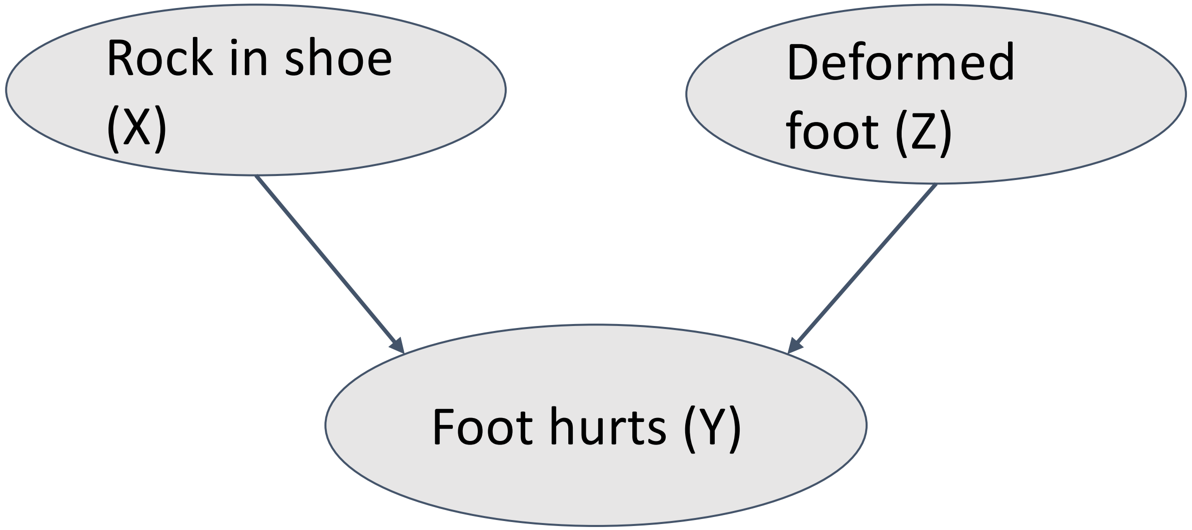
$X \perp\!\!\!\perp Z \mid Y$

Causal vs Diagnostic Modeling

Causal:





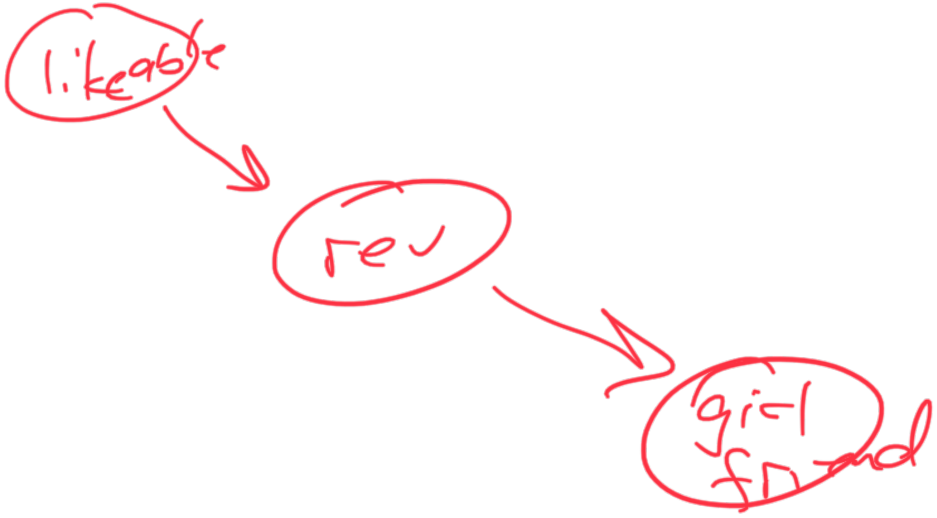


Diagnostic: observing an effect leads to competition between possible causes

→ *diagnose* which is most likely

Causal vs Diagnostic Modeling - Activity

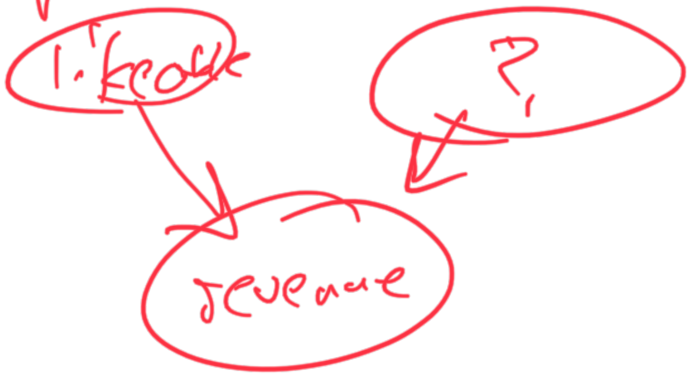
Causal: Come up w/ causal chain



Diagnostic: observing an effect leads to competition between possible causes

→ *diagnose* which is most likely

Diagnostic example.



Bayesian Networks: "Explaining Away"

Suppose we know that the alarm has gone off.

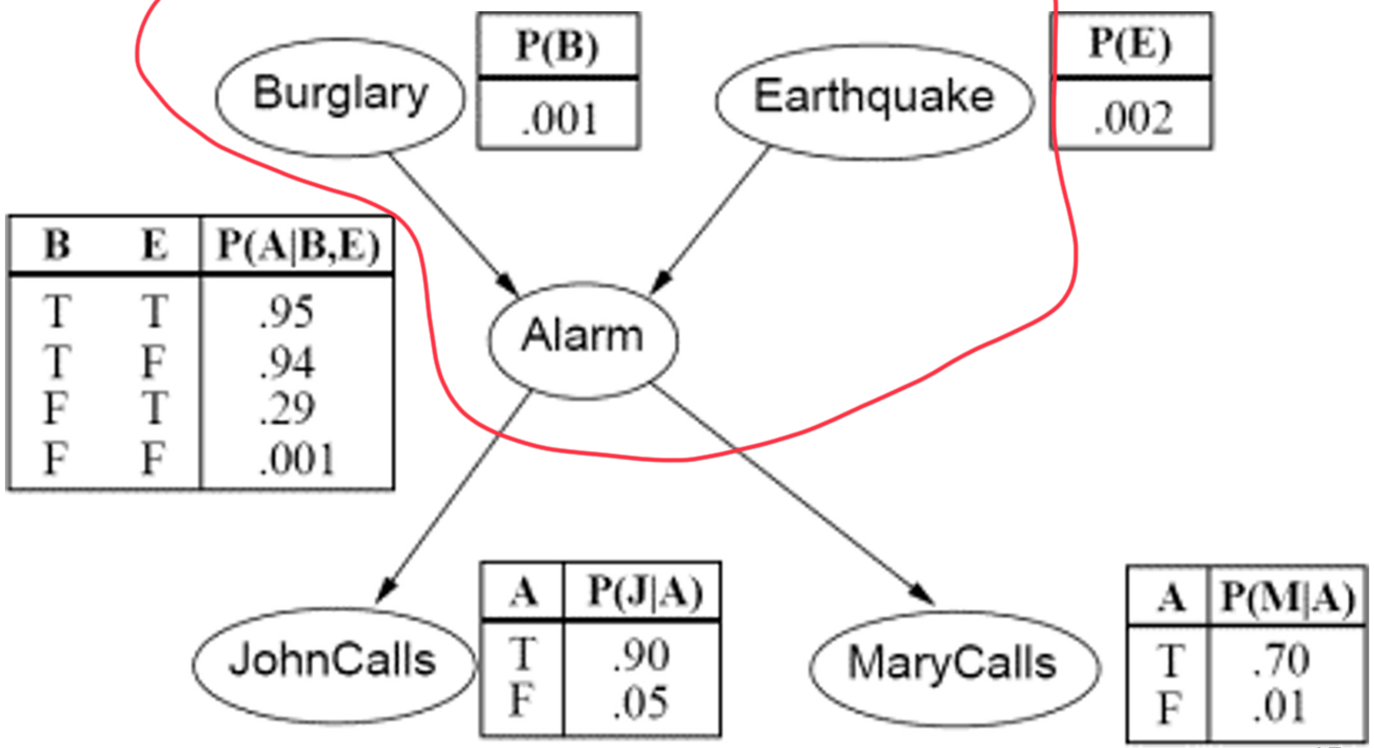
Suppose we find out later that we have been robbed

A = true

B = true

Explains away earthquake as cause of the alarm.

$P(e \mid a, b) < P(e)$



P(B)
.001

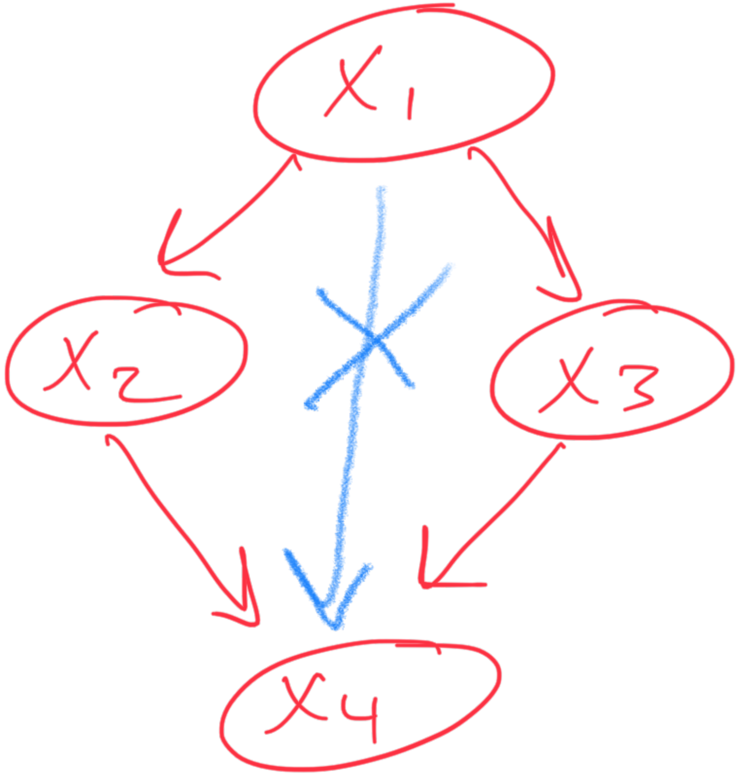
P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Bayesian Networks – basic structure



$$P(X_1, X_2, X_3, X_4) = P(X_4 \mid \text{parents}(X_4)) \times P(X_3 \mid \text{parents}(X_3)) \times P(X_2 \mid \text{parents}(X_2)) \times P(X_1)$$

Bayesian Networks

Bayes nets implicitly encode joint distributions as a product of the local conditional distributions:

$P(x_1, x_2, \dots, x_n) = ?$

$$
\begin{aligned}
P(x_1, x_2, \dots, x_n) &= P(x_n \mid x_{n-1}, x_{n-2}, \dots, x_2, x_1) P(x_{n-1}, x_{n-2}, \dots, x_2, x_1) \\
&= P(x_n \mid x_{n-1}, x_{n-2}, \dots, x_2, x_1) P(x_{n-1} \mid x_{n-2}, \dots, x_2, x_1) P(x_{n-2}, \dots, x_2, x_1) \\
&= \dots \\
&= P(x_n \mid x_{n-1}, x_{n-2}, \dots, x_2, x_1) P(x_{n-1} \mid x_{n-2}, \dots, x_2, x_1) \dots P(x_3 \mid x_2, x_1) P(x_2 \mid x_1) P(x_1) \\
&= \prod_{i=1}^{n} P(x_i \mid x_{i-1}, x_{i-2}, \dots, x_2, x_1)
\end{aligned}
$$

Bayesian Networks – example

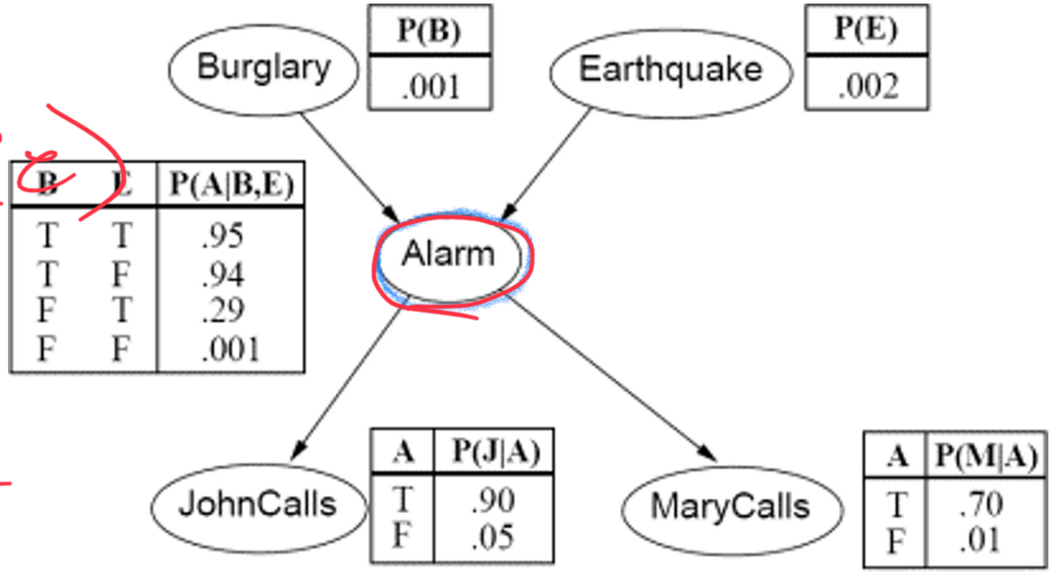
Bayes nets implicitly encode joint distributions as a product of the local conditional distributions:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))$$

$P(a|b,e)$

P(a) = $\sum_b \sum_e \underbrace{P(a|b,e)} P(b) P(e)$

$= (.95 \times .001 \times .002) +$
$(.94 \times .001 \times .998) +$
$(.29 \times .999 \times .002) +$
$(.001 \times .999 \times .998) = .0025$



Bayesian Networks – example

Bayes nets implicitly encode joint distributions as a product of the local conditional distributions:

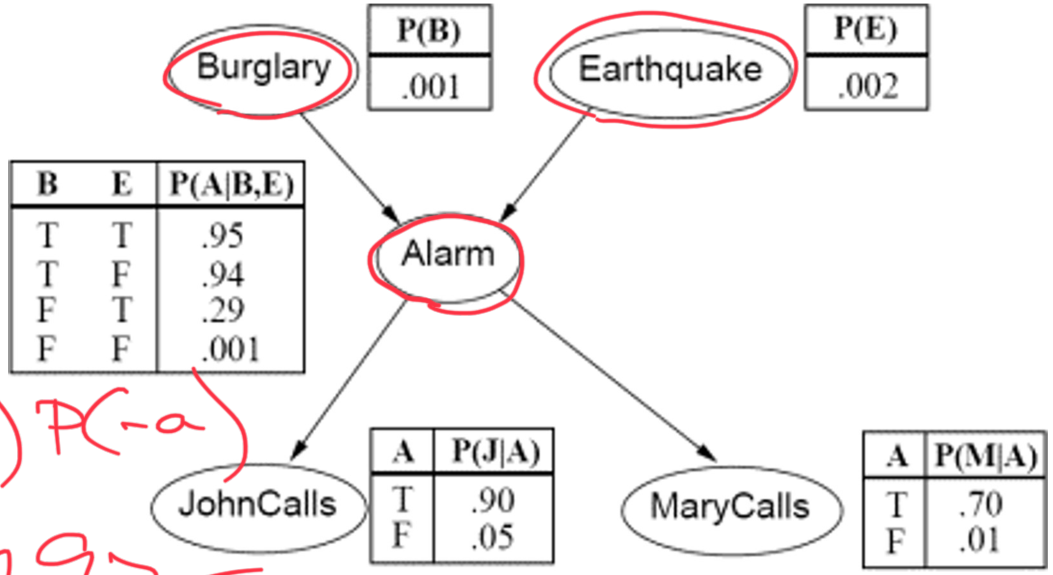
$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))$$

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

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

$= .9 \times .0025 + .05 \times .9975$

$= .052$



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

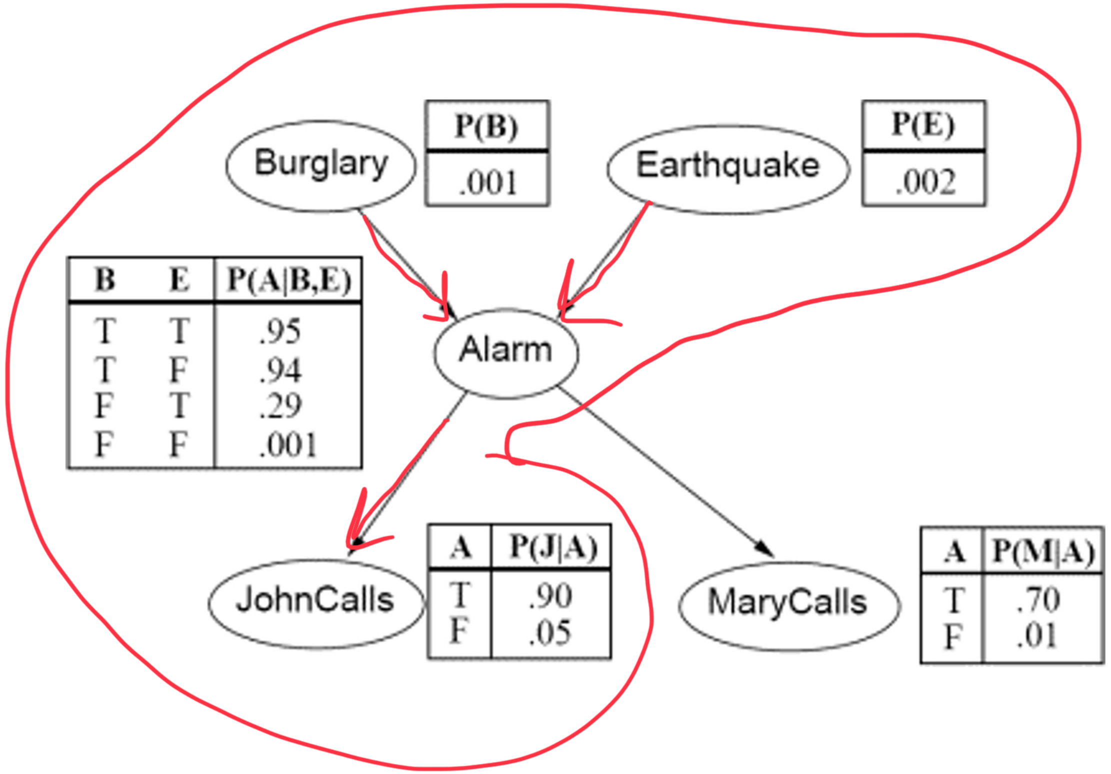
A	P(M|A)
T	.70
F	.01

Bayesian Networks – example

Bayes nets implicitly encode joint distributions as a product of the local conditional distributions:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))$$

P(b | j) =



P(B)
.001

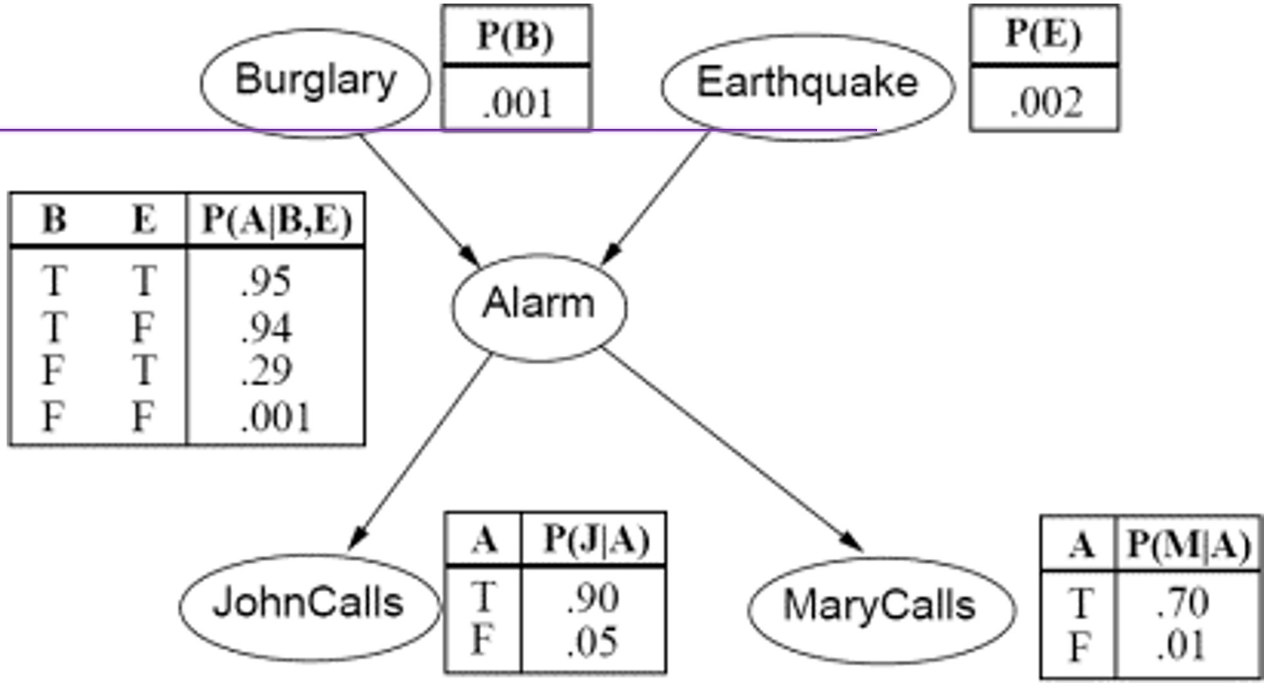
P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Bayesian Networks



Example: *S'pose* we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

→ $P(+b \mid +j, +m) = ?$

Query variables: what we want the **posterior** probability of, **given** some **evidence**

$X = B$

Evidence variables: the variables we are given an assignment of (the **data**)

$\mathbf{E} = [+j, +m]$

Hidden variables: the non-evidence, non-query variables

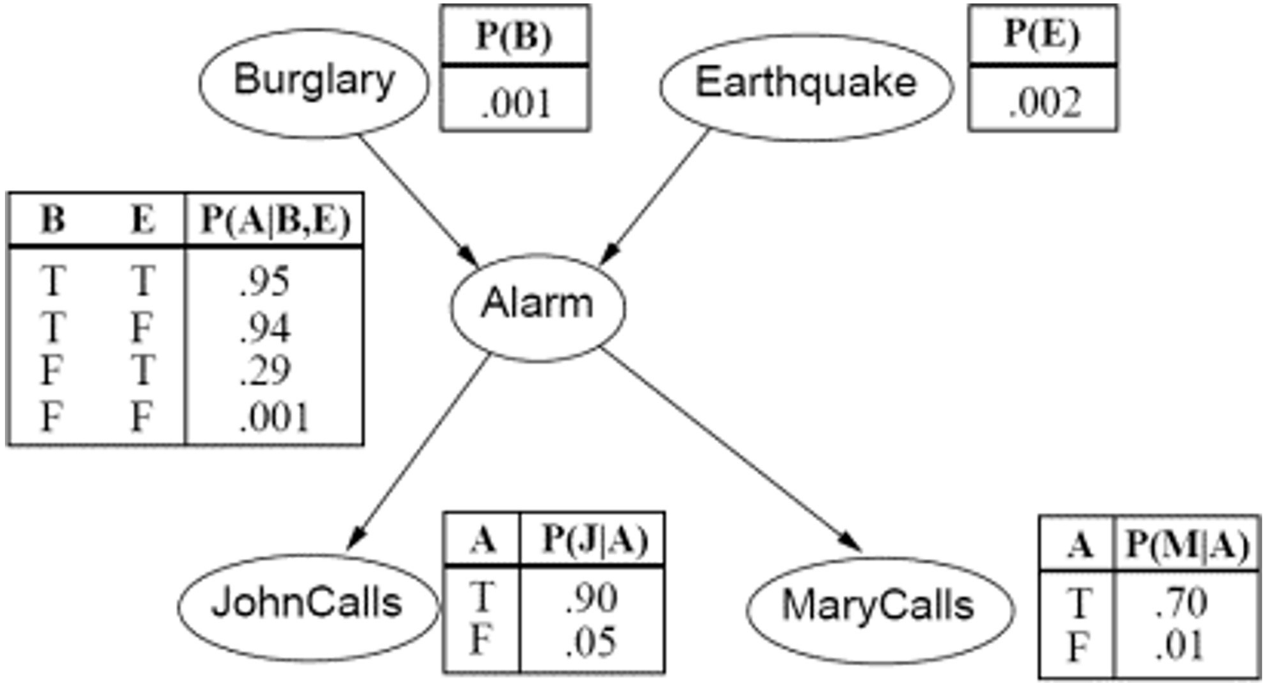
$\mathbf{y} = [E, A]$

Bayesian Networks

Example: *S'pose* we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

→ $P(+b \mid +j, +m) = ?$

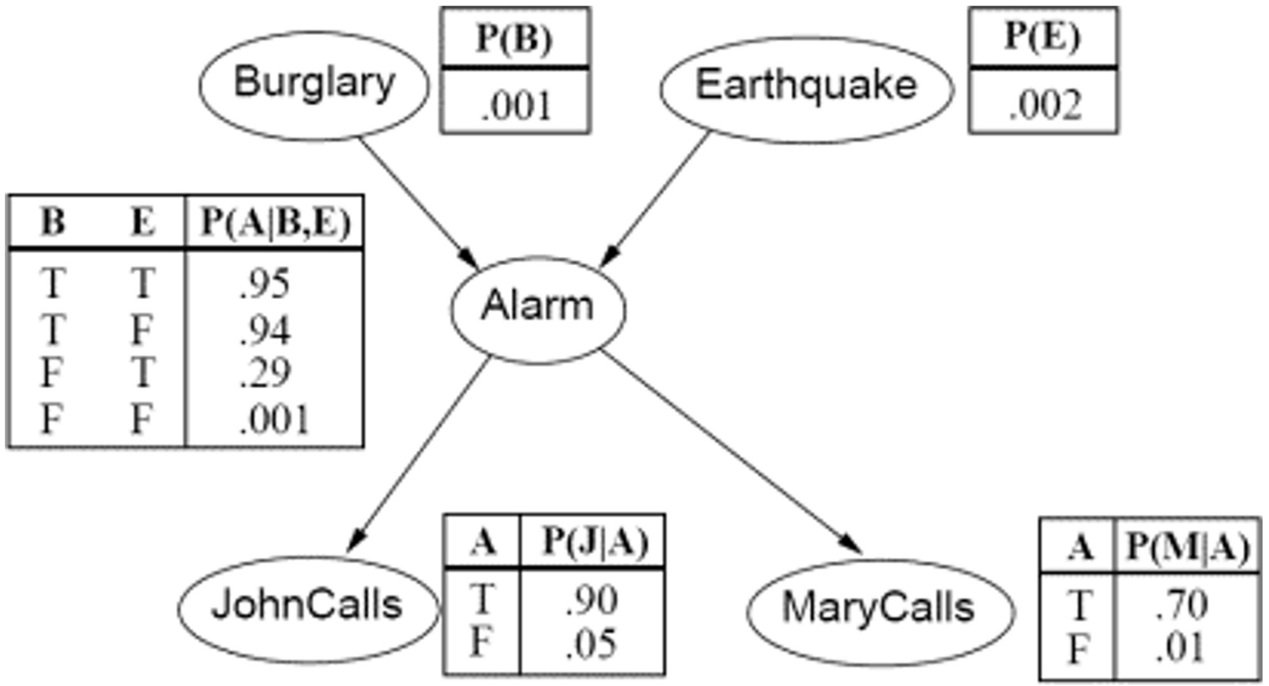
Calculation by **enumeration:**



Bayesian Networks

Example: S'pose we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

→ *P*(+b | +j, +m) = ?



Calculation by **enumeration:**

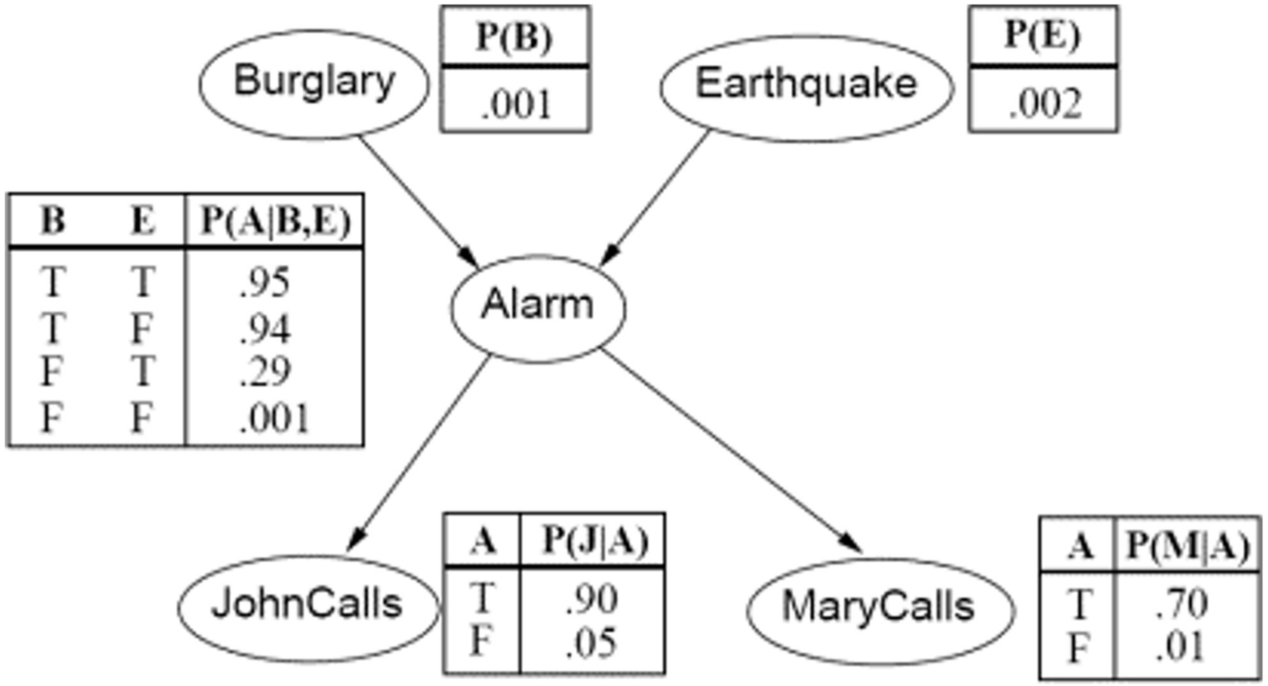
$$P(B \mid j,m) = \frac{P(B,j,m)}{P(j,m)} \quad \Rightarrow \quad \alpha = \frac{1}{P(j,m)} \quad \Rightarrow \quad P(B \mid j,m) = \alpha P(B,j,m)$$

We'll do our thing, then figure out the normalizing constant α later

Bayesian Networks

Example: Suppose we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

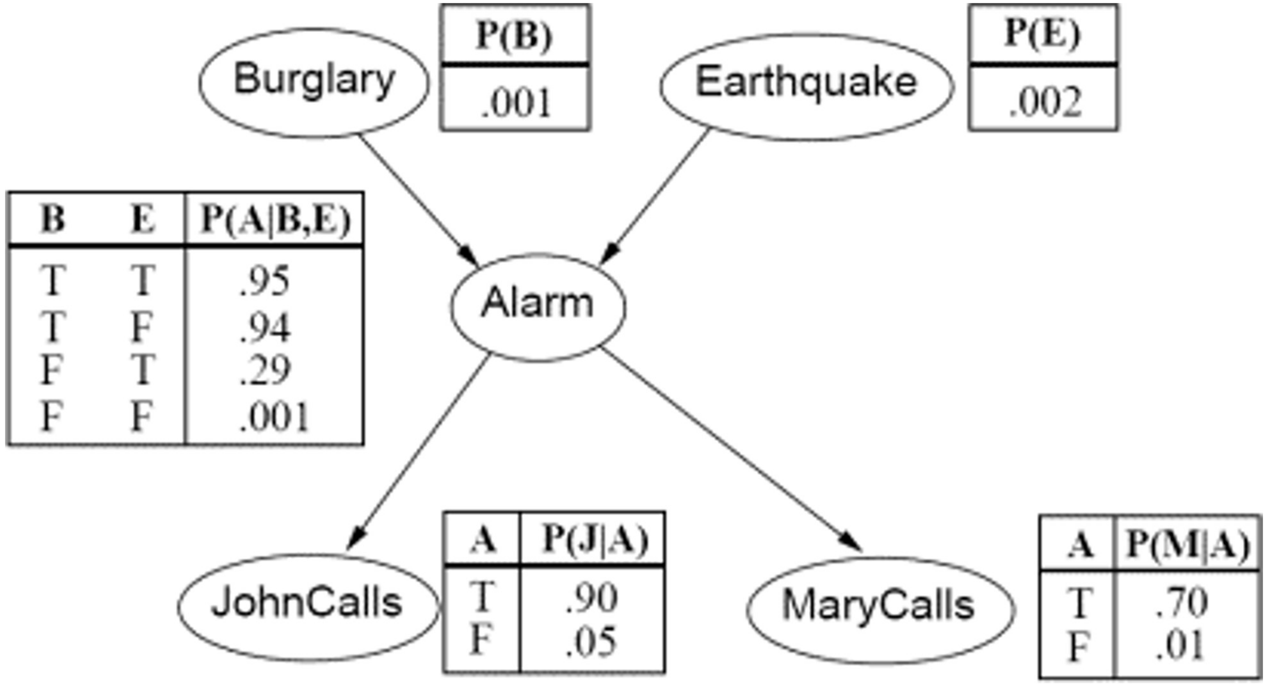
→ $P(+b \mid +j, +m) = ?$



Calculation by **enumeration:**

$$
\begin{aligned}
P(B \mid j, m) &= \alpha P(B, j, m) \\
&= \alpha \sum_a P(B, j, m \mid a) P(a) \\
&= \alpha \sum_a P(B, j, m, a)
\end{aligned}
$$

Bayesian Networks



Example: Suppose we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

→ *P*(+b | +j, +m) = ?

Calculation by **enumeration:**

$$P(B \mid j, m) = \alpha \sum_a P(B, j, m, a)$$

$$= \alpha \sum_e \sum_a P(B, j, m, a \mid e) P(e)$$

$$= \alpha \sum_e \sum_a P(B, j, m, a, e)$$

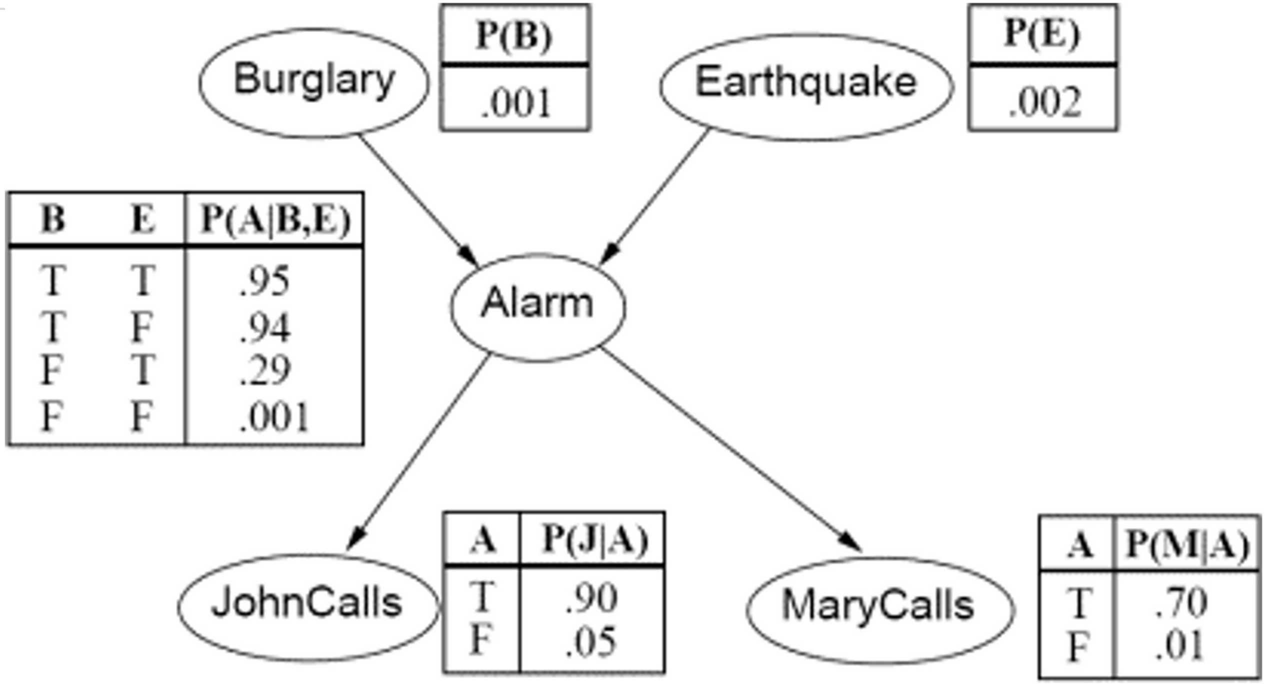
Bayesian Networks

<u>Example</u>: Suppose we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

→ $P(+b \mid +j, +m) = ?$

FINALLY we have:

$$P(B \mid j, m) = \alpha \sum_{e} \sum_{a} P(B, j, m, a, e)$$



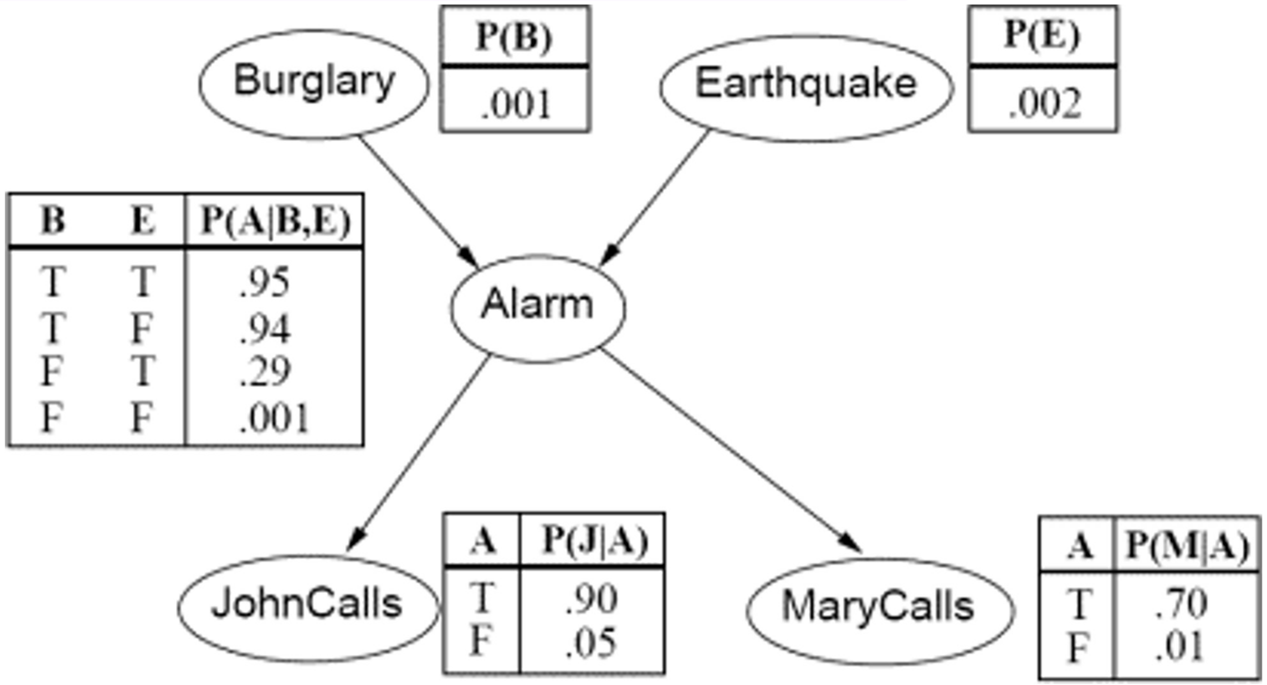
From the conditional independence of the Bayes net:

$$P(B \mid j, m) = \alpha \sum_{e} \sum_{a} \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))$$

Bayesian Networks

<u>Example</u>: Suppose we know the alarm has gone off, and both John and Mary have called to warn us. What is the probability that we have been burgled?

→ *P*(+b | +j, +m) = ?



So for this problem…

$$P(B \mid j, m) = \alpha \sum_e \sum_a \prod_{i=1}^{n} P(x_i \mid \text{parents}(X_i))$$

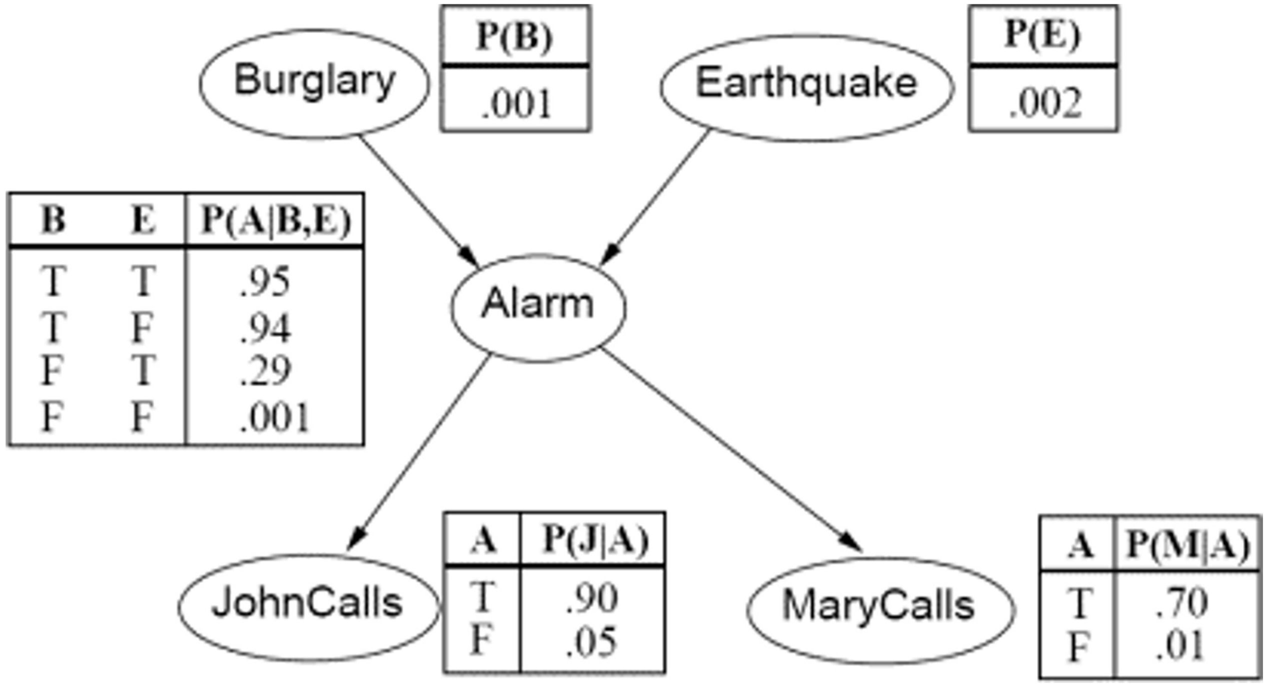
$$= \alpha \sum_e \sum_a P(B)P(e)P(a \mid B, e)P(j \mid a)P(m \mid a)$$

$$= \alpha P(B) \sum_e P(e) \sum_a P(a \mid B, e)P(j \mid a)P(m \mid a)$$

Bayesian Networks

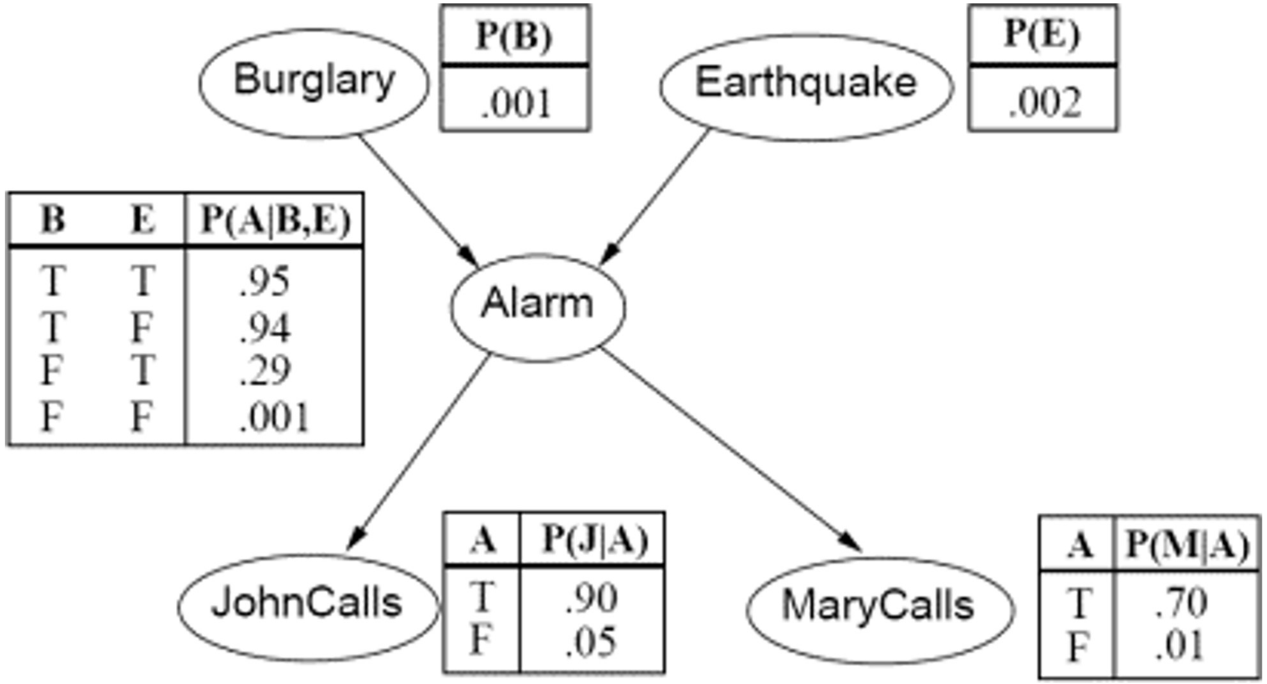
Example:

$$P(B \mid j, m) = \alpha P(B) \sum_e P(e) \sum_a P(a \mid B, e) P(j \mid a) P(m \mid a)$$



Bayesian Networks

$$P(B \mid j, m) = \alpha P(B) \sum_e P(e) \sum_a P(a \mid B, e) P(j \mid a) P(m \mid a)$$



Bayesian Networks

$$P(B \mid j, m) = \alpha P(B) \sum_{e} P(e) \sum_{a} P(a \mid B, e) P(j \mid a) P(m \mid a)$$