Bayesian Networks

CSE 4308/5360: Artificial Intelligence I

University of Texas at Arlington

Motivation for Bayesian Networks

- An important task for probabilistic systems is inference.
- In probability, inference is the task of computing:

  $P(A_1, \ldots, A_k \mid B_1, \ldots, B_m)$

  where $A_1, \ldots, A_k, B_1, \ldots., B_m$ are any random variables.
  - Note that m can be zero, in which case we simply want to compute $P(A_1, \ldots, A_k)$.
- So far we have seen one way to solve the inference problem: ???

Motivation for Bayesian Networks

- An important task for probabilistic systems is inference.
- In probability, inference is the task of computing:

  $P(A_1, \ldots, A_k \mid B_1, \ldots, B_m)$

  where $A_1, \ldots, A_k, B_1, \ldots, B_m$ are any random variables.
  - Note that m can be zero, in which case we simply want to compute $P(A_1, \ldots, A_k)$.
- So far we have seen one way to solve the inference problem: Inference by enumeration (using a joint distribution table).
- However, inference by enumeration has three limitations:
  - Too slow: time exponential to k+m.
  - Too much memory needed: space exponential to k+m.
  - Too much training data and effort are needed to compute the entries in the joint distribution table.

Motivation for Bayesian Networks

- Bayesian networks offer a different way to represent joint probability distributions.
- They require space linear to the number of variables, as opposed to exponential.
  - This means fewer numbers need to be stored, so less memory is needed.
  - This also means that fewer numbers need to be computed, so less effort is needed to compute those numbers and specify the probability distribution.
- Also, in specific cases, Bayesian networks offer polynomial-time algorithms for inference, using dynamic programming.
  - In this course, we will not cover such polynomial time algorithms, but it is useful to know that they exist.
  - If you are curious, see the **variable elimination algorithm** in the textbook, Chapter 14.4.2.

Definition of Bayesian Networks

- A Bayesian network is a directed acyclic graph, that defines a joint probability distribution over N random variables.
- The Bayesian network contains N nodes, and each node corresponds to one of the N random variables.
- If there is a directed edge from node X to node Y, then we say that X is a *parent* of Y.
- Each node X has a conditional probability distribution P(X | Parents(X)) that describes the probability of any value of X given any combination of values for the parents of X.

An Example from the Textbook

Weather

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01

Cavity

P(CV)
0.2

Toothache

Cavity	P(TA)
T	0.6
F	0.1

Catch

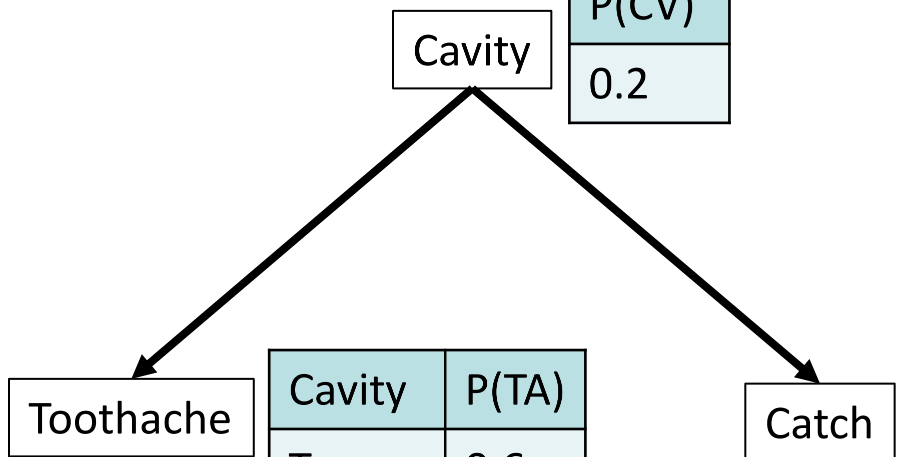
Cavity	P(CT)
T	0.9
F	0.2

- How many random variables do we have?

An Example from the Textbook

Weather

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01



P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

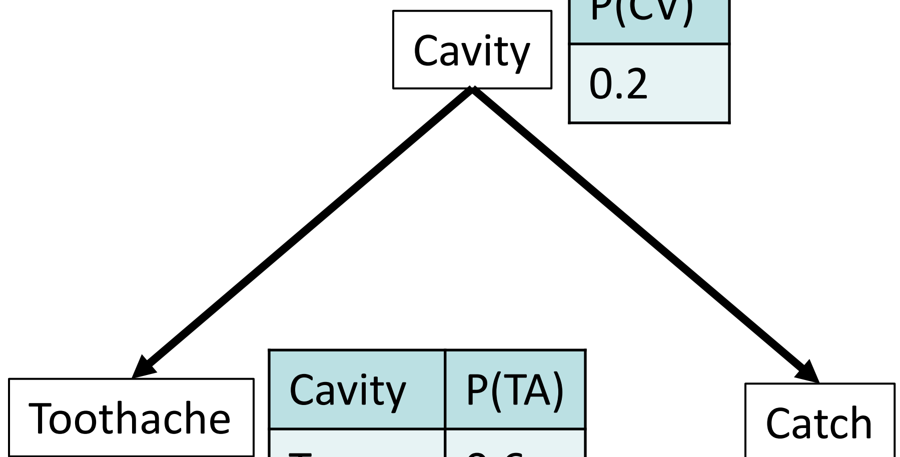
Cavity	P(CT)
T	0.9
F	0.2

- How many random variables do we have?
  - 4: Weather, Cavity, Toothache, Catch.
- Note that Weather can take 4 discrete values.
- The other three variables are boolean.

An Example from the Textbook

Weather

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01



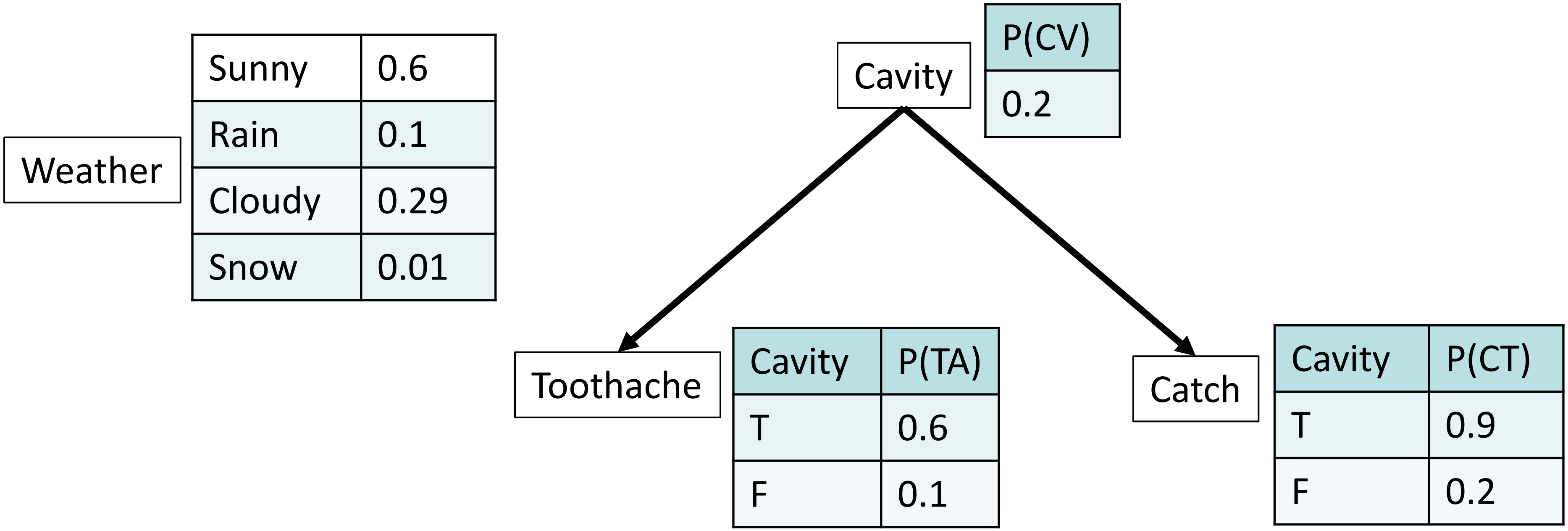
P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

Cavity	P(CT)
T	0.9
F	0.2

- What are the parents of Weather?
- What are the parents of Cavity?
- What are the parents of Toothache?
- What are the parents of Catch?

An Example from the Textbook



Weather	
Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01

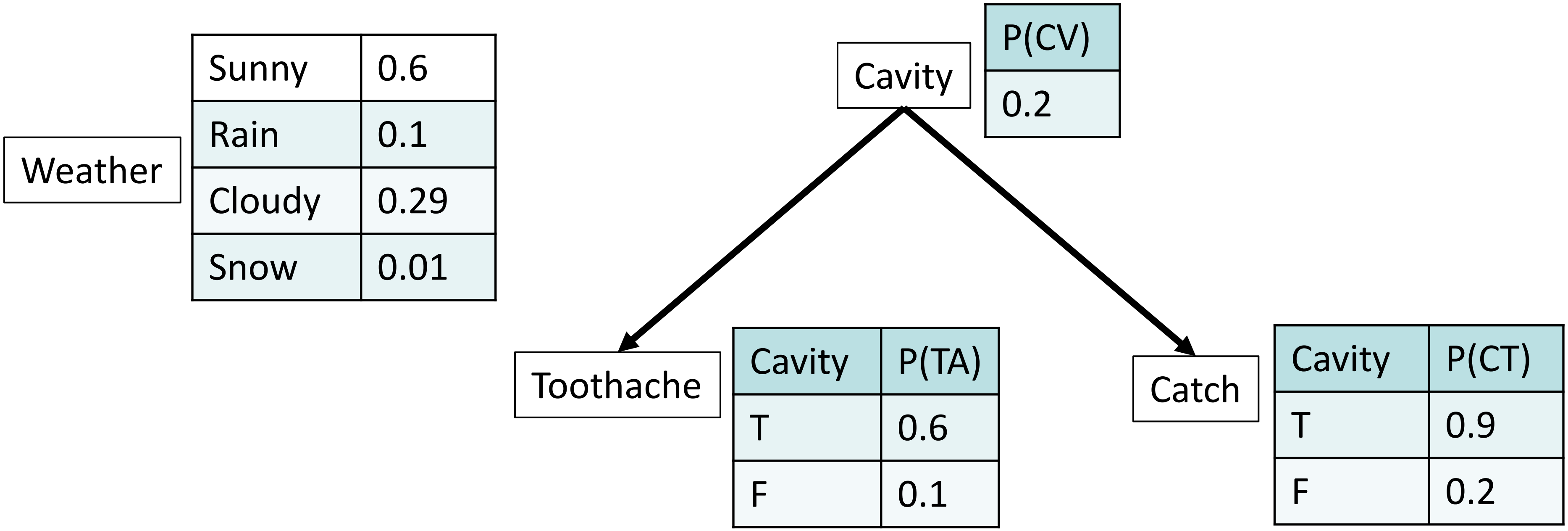
P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

Cavity	P(CT)
T	0.9
F	0.2

- What are the parents of Weather? None.
- What are the parents of Cavity? None.
- What are the parents of Toothache? Cavity.
- What are the parents of Catch? Cavity.

An Example from the Textbook



Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01

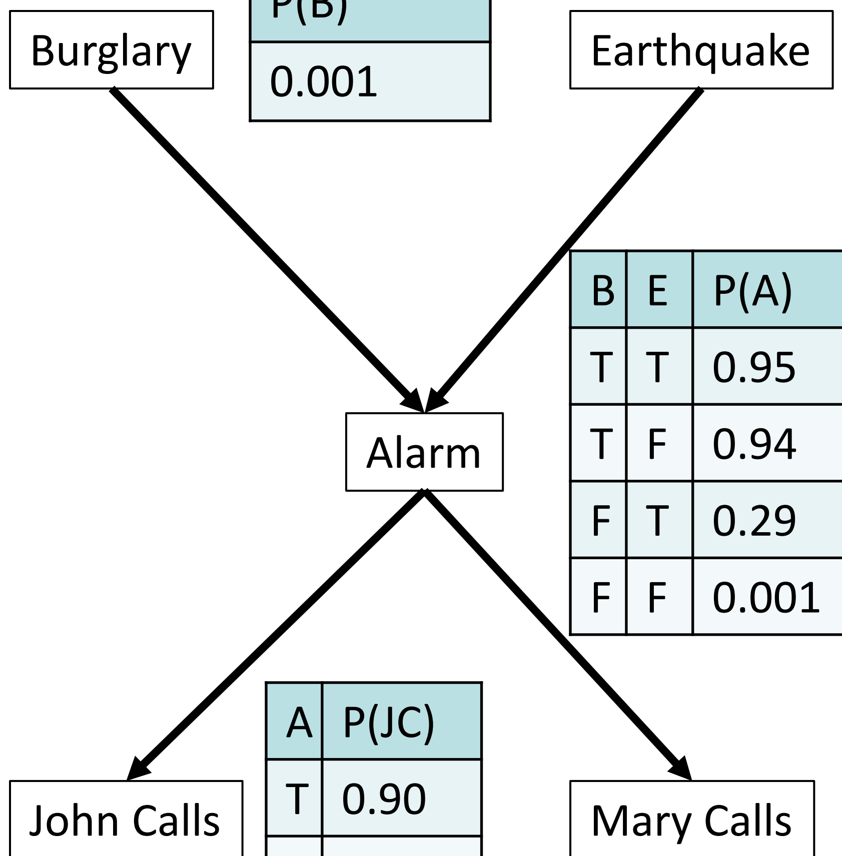
P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

Cavity	P(CT)
T	0.9
F	0.2

- What does this network mean?
  - Weather is independent of the other three variables.
  - Cavities can cause both toothaches and catches.
  - Toothaches and catches are conditionally independent given the value for cavity.

Another Textbook Example



P(B)
0.001

P(E)
0.002

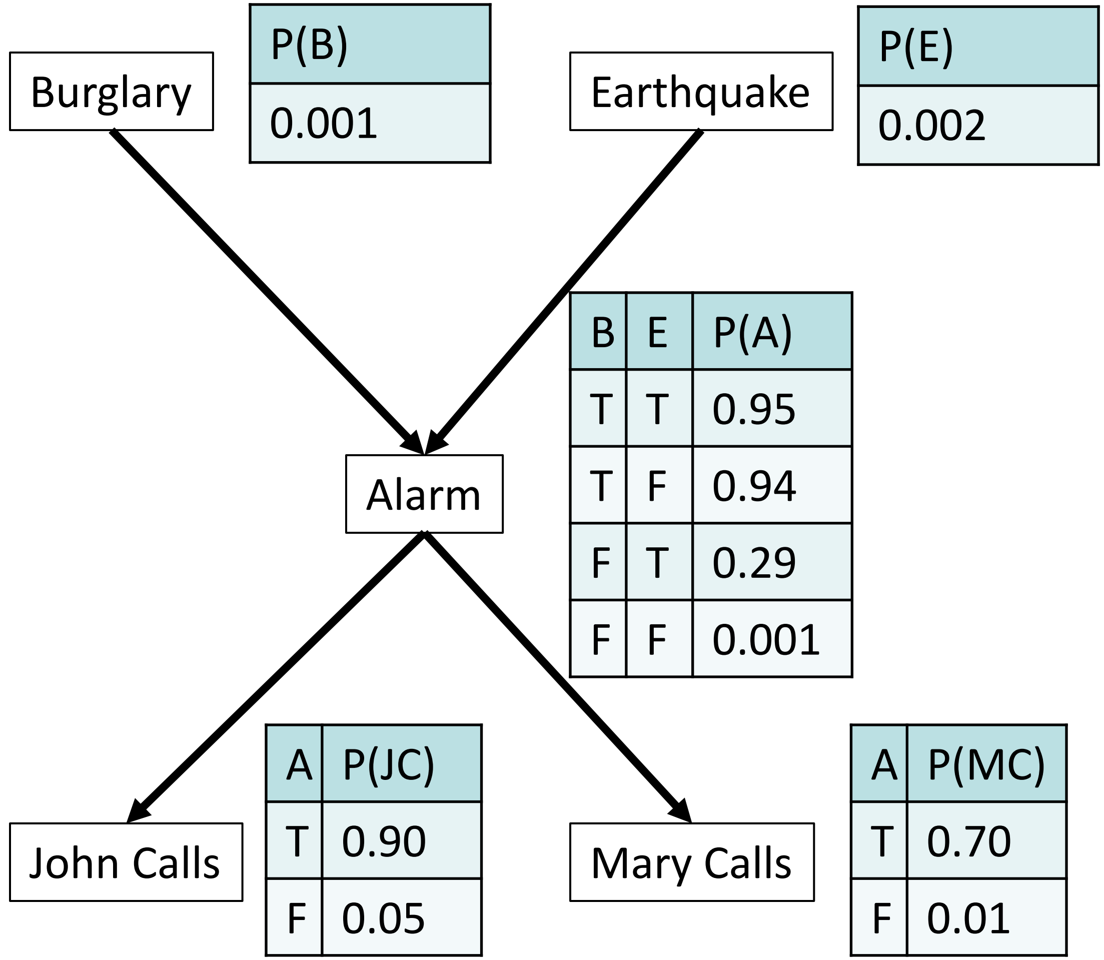
B	E	P(A)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

- How many random variables do we have?

Another Textbook Example



P(B)
0.001

P(E)
0.002

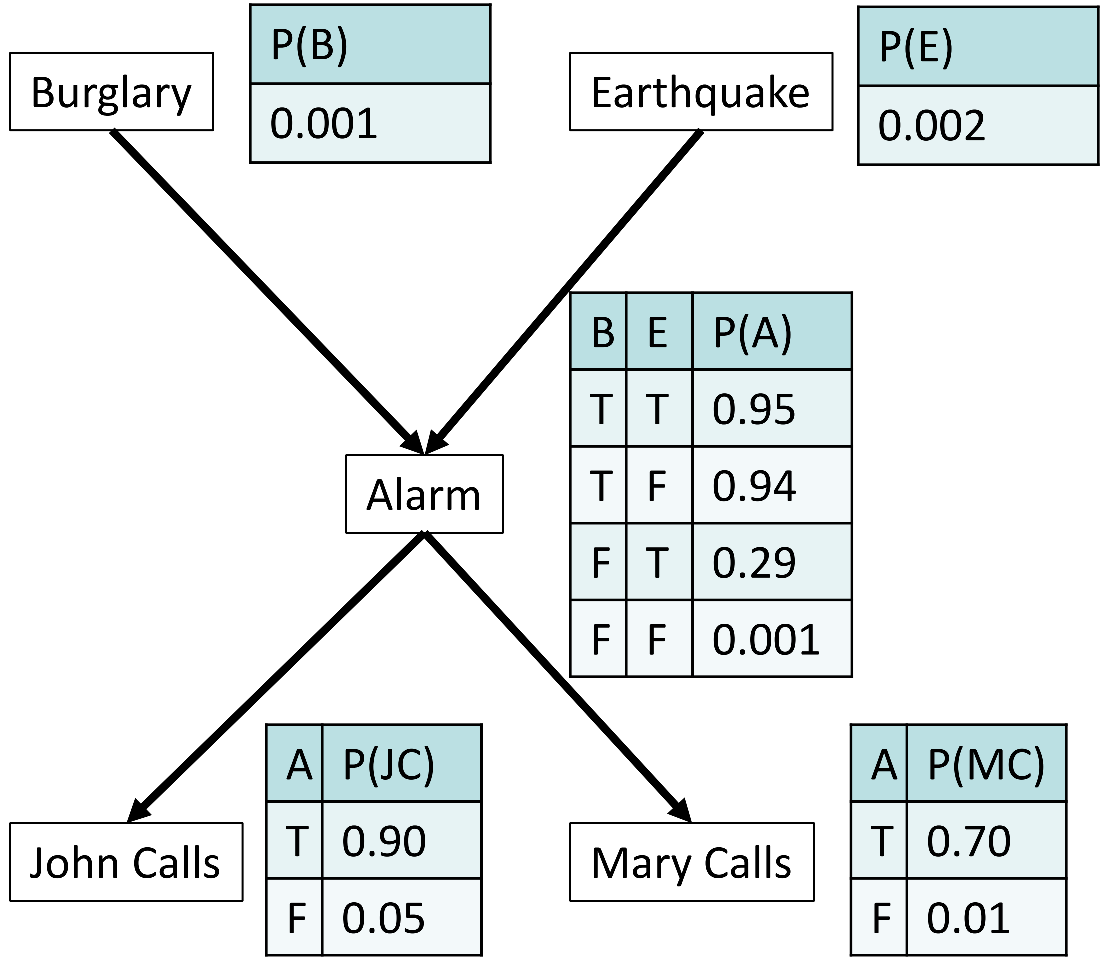
B	E	P(A)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

- How many random variables do we have?
- 5: Burglary, Earthquake, Alarm, JohnCalls, MaryCalls.
  - All boolean.

Another Textbook Example



P(B)
0.001

P(E)
0.002

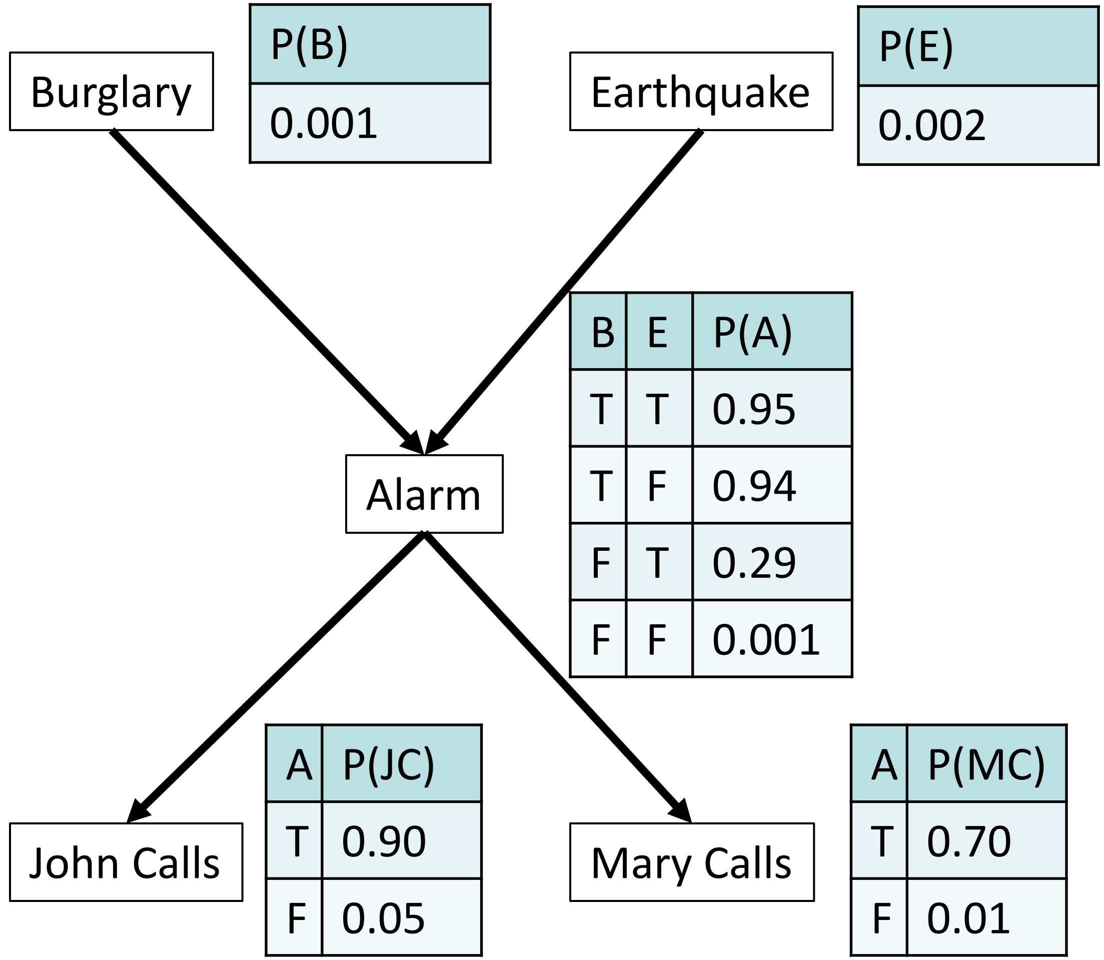
B	E	P(A)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

- What are the parents of Burglary?
- What are the parents of Earthquake?
- What are the parents of Alarm?
- What are the parents of JohnCalls?
- What are the parents of MaryCalls?

Another Textbook Example



P(B)
0.001

P(E)
0.002

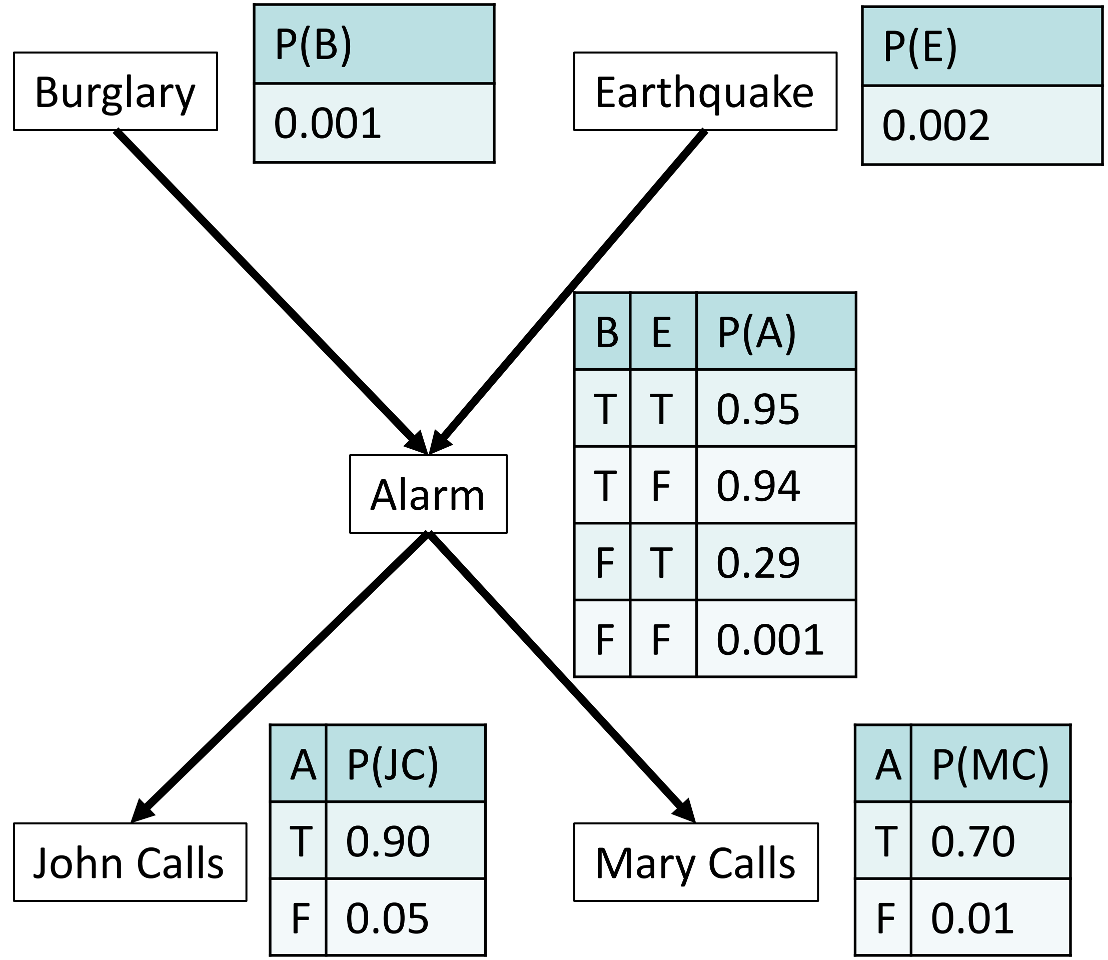
B	E	P(A)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

- What are the parents of Burglary? None.
- What are the parents of Earthquake? None.
- What are the parents of Alarm? Burglary and Earthquake.
- What are the parents of JohnCalls? Alarm.
- What are the parents of MaryCalls? Alarm.

Another Textbook Example



P(B)
0.001

P(E)
0.002

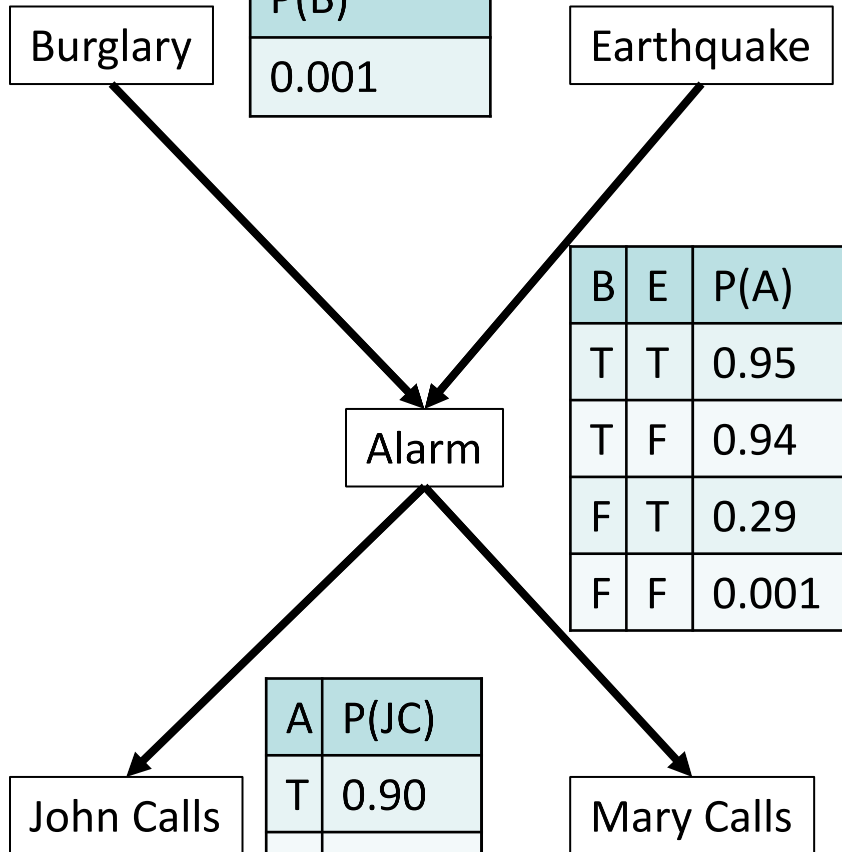
B	E	P(A)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

- What does this network mean?

Another Textbook Example



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(JC)
T	0.90
F	0.05

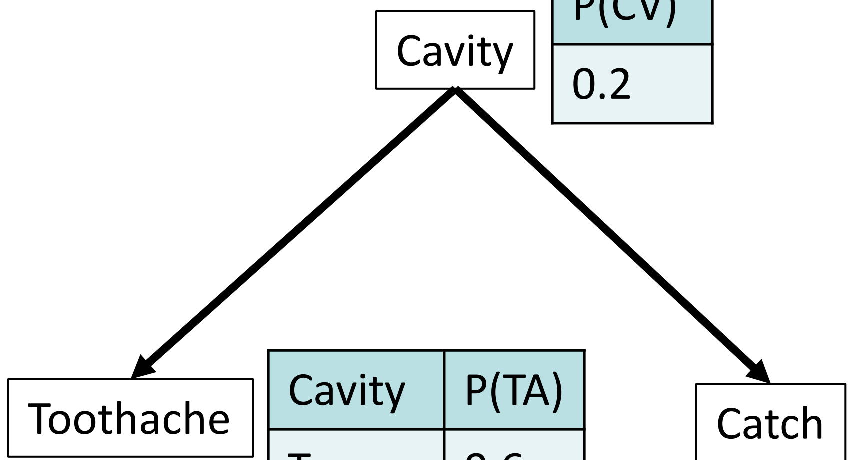
A	P(MC)
T	0.70
F	0.01

- What does this network mean?
  - Alarms can be caused by both burglaries and earthquakes.
  - Alarms can cause both John to call and Mary to call.
  - Whether John calls or not is conditionally independent of whether Mary calls or not, given the value of the Alarm variable.

Semantics

Weather

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01



P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

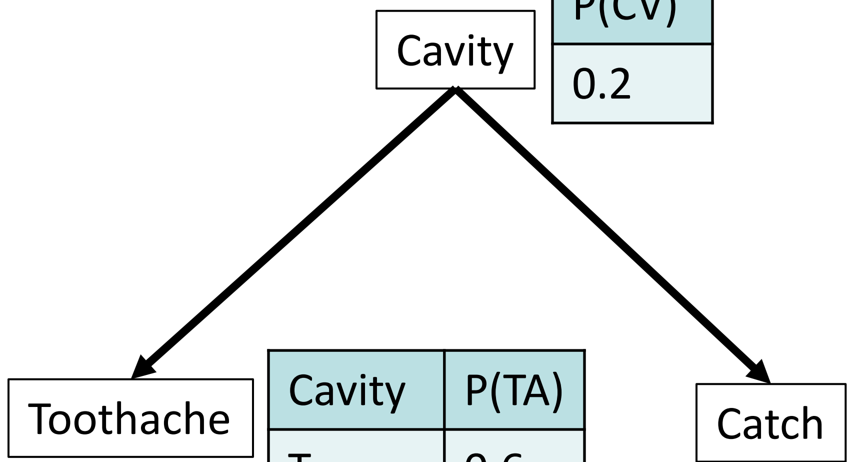
Cavity	P(CT)
T	0.9
F	0.2

- So far, we have described the structure of a Bayesian network, as a directed acyclic graph.
- We also need to define the meaning: what does this graph mean? What information does it provide.

Semantics

Weather

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01



P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

Cavity	P(CT)
T	0.9
F	0.2

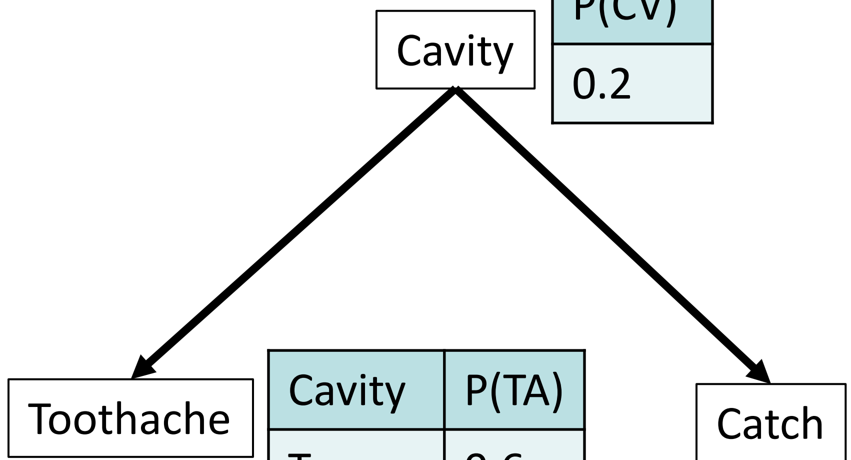
- A Bayesian network defines the joint probability distribution of the variables represented by its nodes.
- If X_1, …, X_n are the n variables of the network, then:

$$P(X_1, \dots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

Semantics

Weather

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01



P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

Cavity	P(CT)
T	0.9
F	0.2

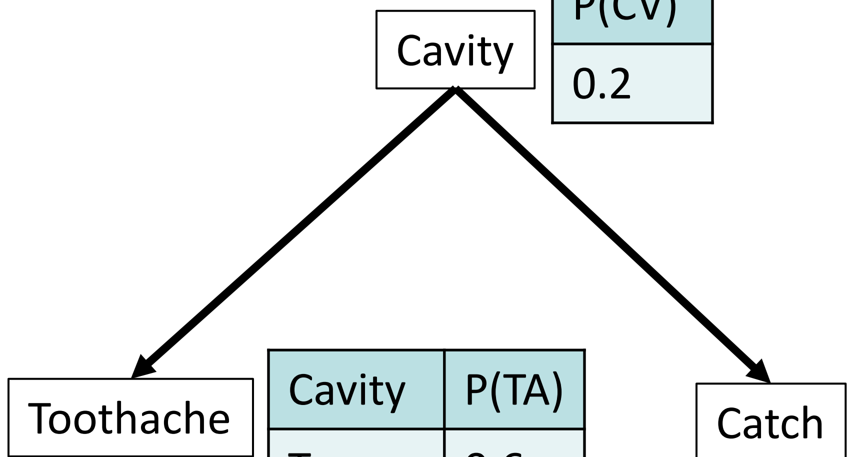
$$P(X_1, \dots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- This equation is part of the definition of Bayesian networks.
- If you do not understand how to use it, you will not be able to solve most problems related to Bayesian networks.

Inference in Bayesian Networks

Weather

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01



P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

Cavity	P(CT)
T	0.9
F	0.2

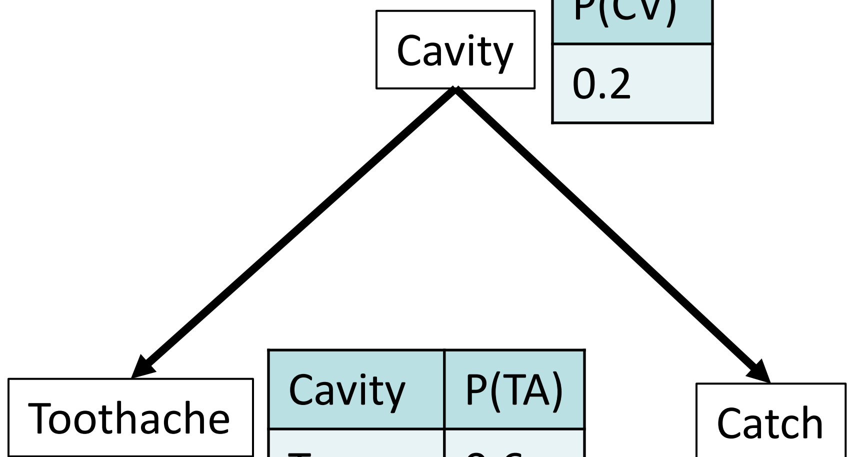
$$P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- In general, probabilistic inference is the problem of computing $P(A_1, \ldots, A_k \mid B_1, \ldots, B_m)$
- In other words, it is the problem of computing the probability of values for some variables given values for some other variables.

Inference in Bayesian Networks

Weather

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01



P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

Cavity	P(CT)
T	0.9
F	0.2

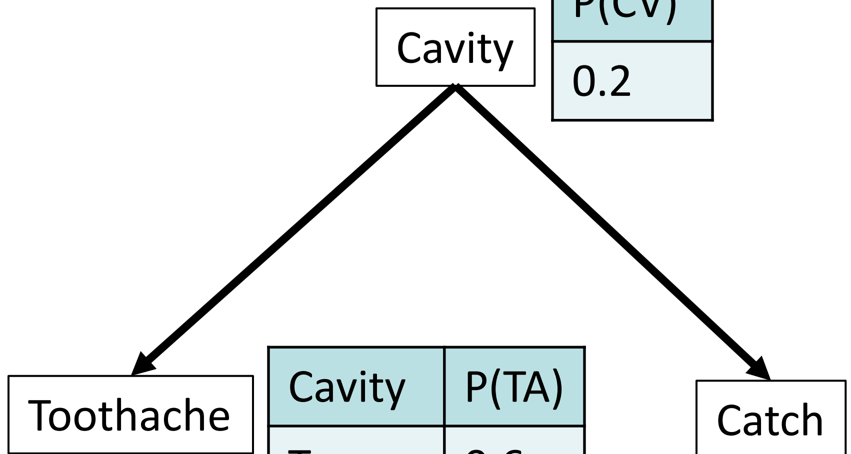
$$P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- In Bayesian networks, all inference problems can be solved by one or more applications of the equation below.
- In many interesting cases there exist better (i.e., faster) methods, but we will not study such methods in this course.

Inference in Bayesian Networks

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01

Weather



P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

Cavity	P(CT)
T	0.9
F	0.2

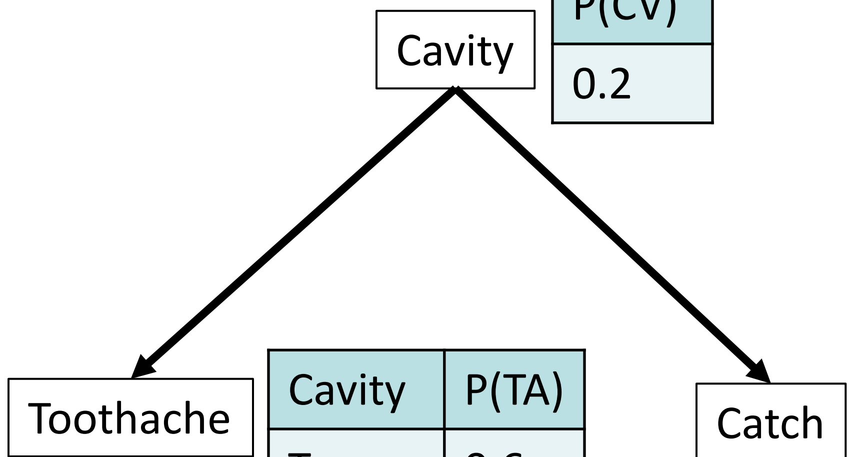
$$P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- For example, compute:
  P(Sunny, not(Cavity), not(Toothache), Catch).
- Based on the equation, how do we compute this?

Inference in Bayesian Networks

Weather

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01



P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

Cavity	P(CT)
T	0.9
F	0.2

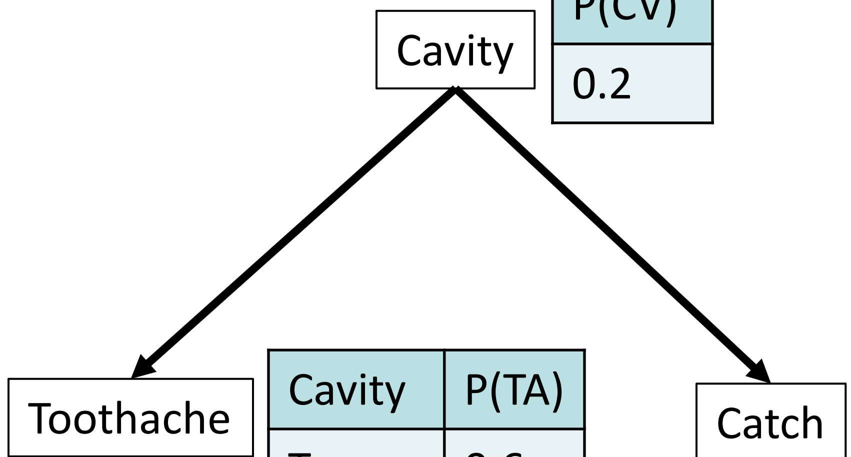
$$P(X_1, \dots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

P(Sunny, not(Cavity), not(Toothache), Catch) =
P(Sunny | Parents(Weather)) *
P(not(Cavity) | Parents(Cavity)) *
P(not(Toothache) | Parents(Toothache)) *
P(Catch | Parents(Catch))

Inference in Bayesian Networks

Weather

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01



P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

Cavity	P(CT)
T	0.9
F	0.2

$$P(X_1, \dots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

P(Sunny, not(Cavity), not(Toothache), Catch) =

P(Sunny) *

P(not(Cavity)) *

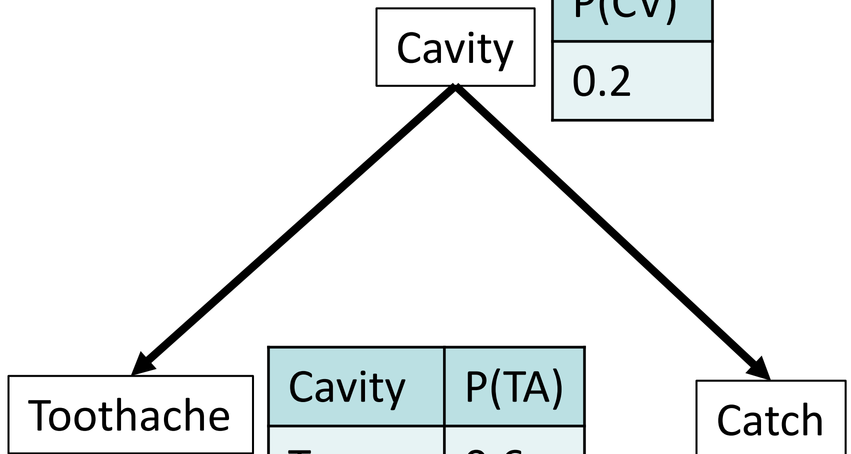
P(not(Toothache) | not(Cavity)) *

P(Catch | not(Cavity))

Inference in Bayesian Networks

Weather

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01



P(CV)
0.2

Cavity	P(TA)
T	0.6
F	0.1

Cavity	P(CT)
T	0.9
F	0.2

$$P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

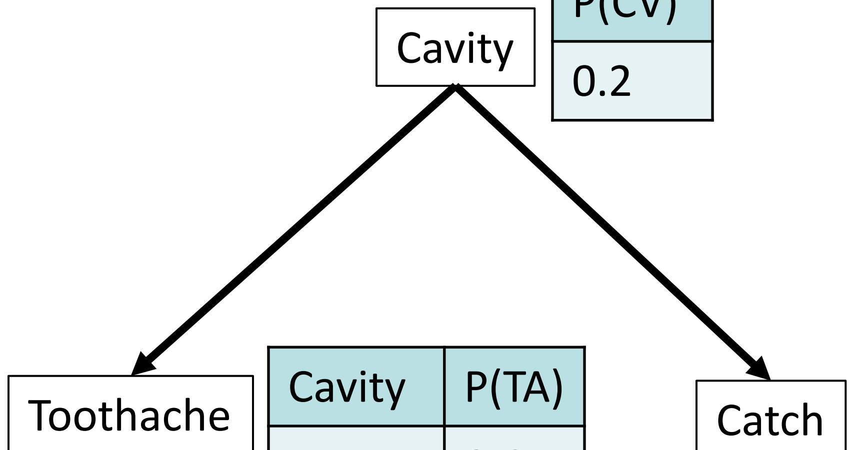
P(Sunny, not(Cavity), not(Toothache), Catch) =

0.6 *
0.8 *
0.9 *
0.2

Inference in Bayesian Networks

Weather

Sunny	0.6
Rain	0.1
Cloudy	0.29
Snow	0.01



P(CV)
0.2

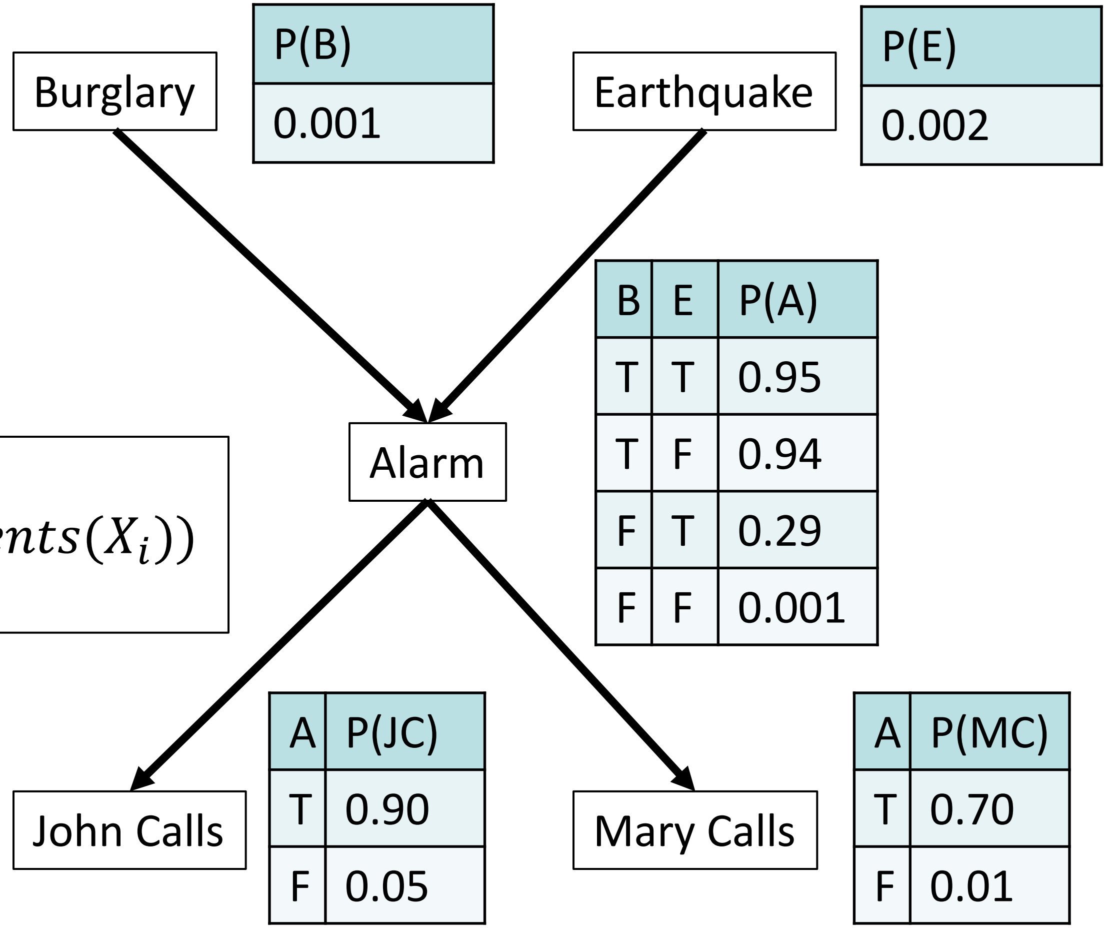
Cavity	P(TA)
T	0.6
F	0.1

Cavity	P(CT)
T	0.9
F	0.2

$$P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

P(Sunny, not(Cavity), not(Toothache), Catch) = 0.6 * 0.8 * 0.9 * 0.2 = 0.0864

Another Example



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

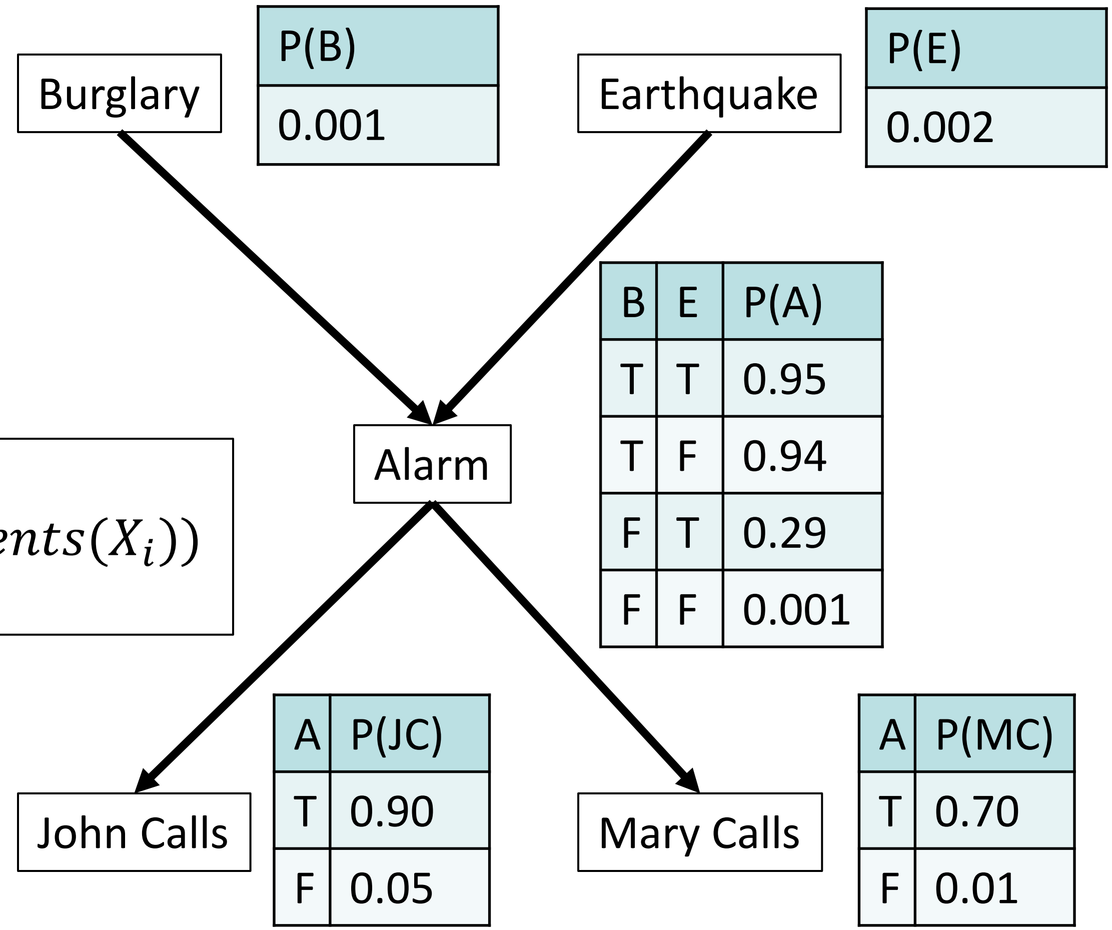
A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

$$P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- Compute P(B, not(E), A, JC, MC):

Another Example



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(JC)
T	0.90
F	0.05

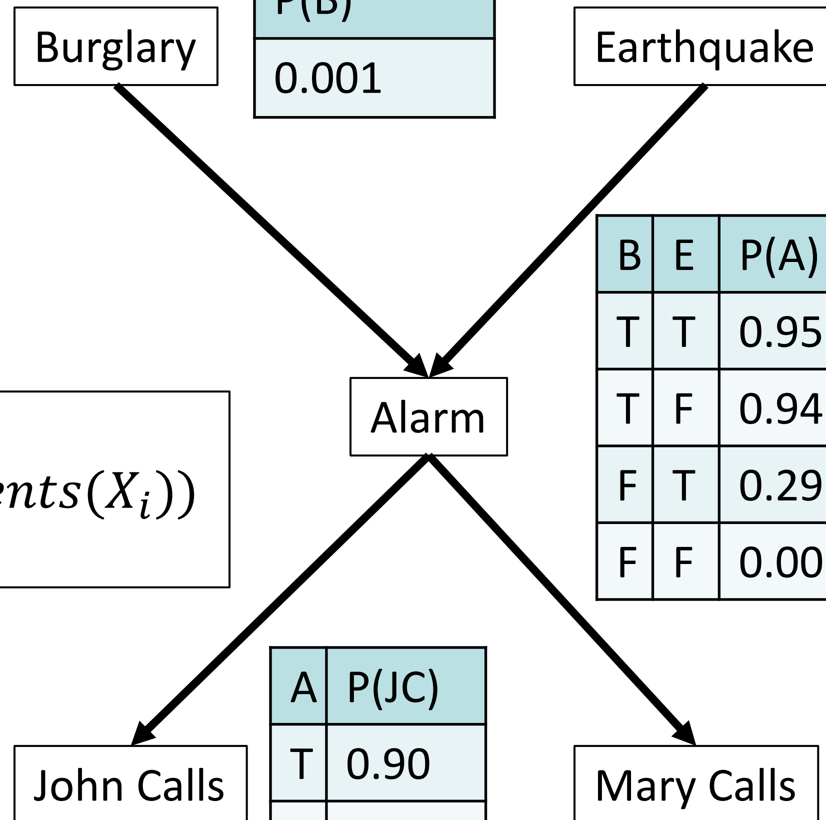
A	P(MC)
T	0.70
F	0.01

$$P(X_1, \dots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- P(B, not(E), A, JC, MC) =

  P(B) * P(not(E)) * P(A | B, not(E)) * P(JC | A) * P(MC | A) =

Another Example



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(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

$$P(X_1, \dots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- P(B, not(E), A, JC, MC) =

  P(B) * P(not(E)) * P(A | B, not(E)) * P(JC | A) * P(MC | A) =

  0.001 * 0.998 * 0.94 * 0.9 * 0.7 = 0.0005910156

A More Complicated Case

- In the previous examples, we computed the probability of cases where all variables were assigned values.
  - We did that by directly applying the equation:

$$\boxed{P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))}$$

- What do we do when some values are unspecified?
- For example, how do we compute P(¬B, JC, MC)?

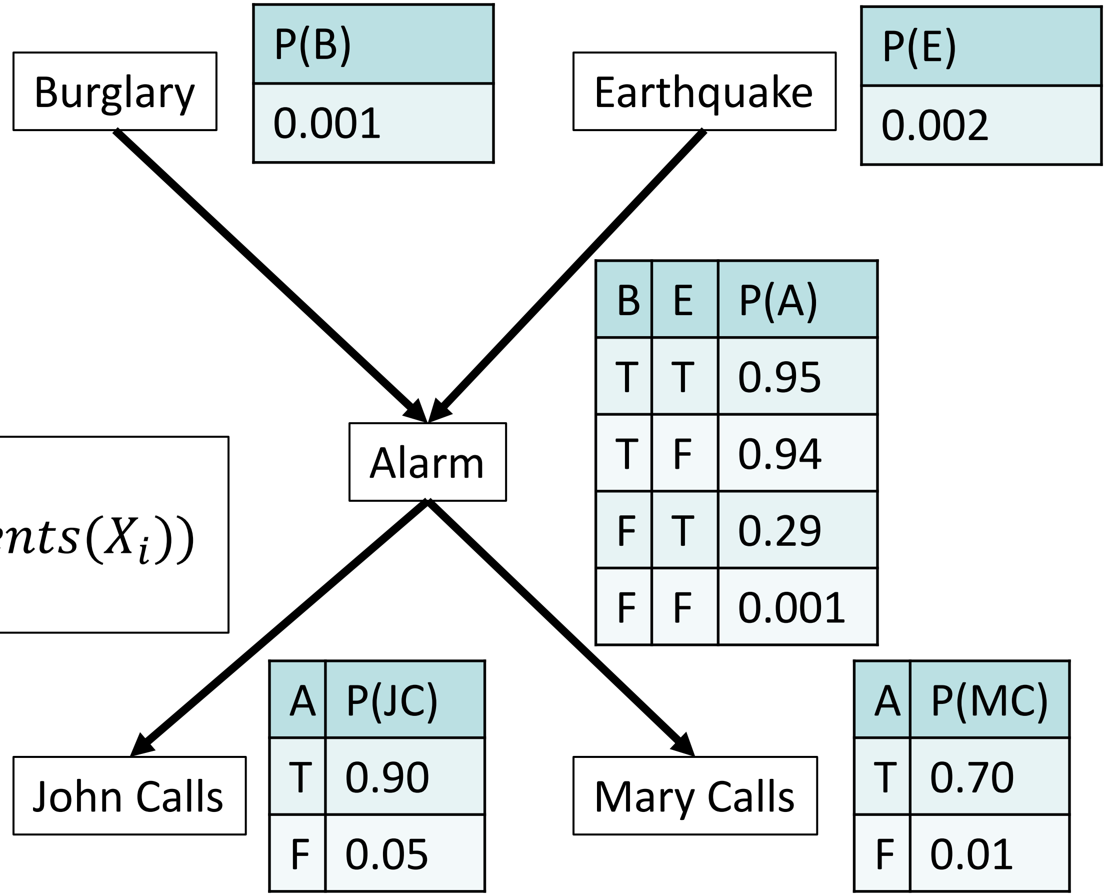
A More Complicated Case

- In the previous examples, we computed the probability of cases where all variables were assigned values.
  - We did that by directly applying the equation:

$$\boxed{P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))}$$

- What do we do when some values are unspecified?
- For example, how do we compute P(¬B, JC, MC)?
  - Answer: we need to apply the above equation repeatedly, and sum over all possible values that are left unspecified.

Another Example



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

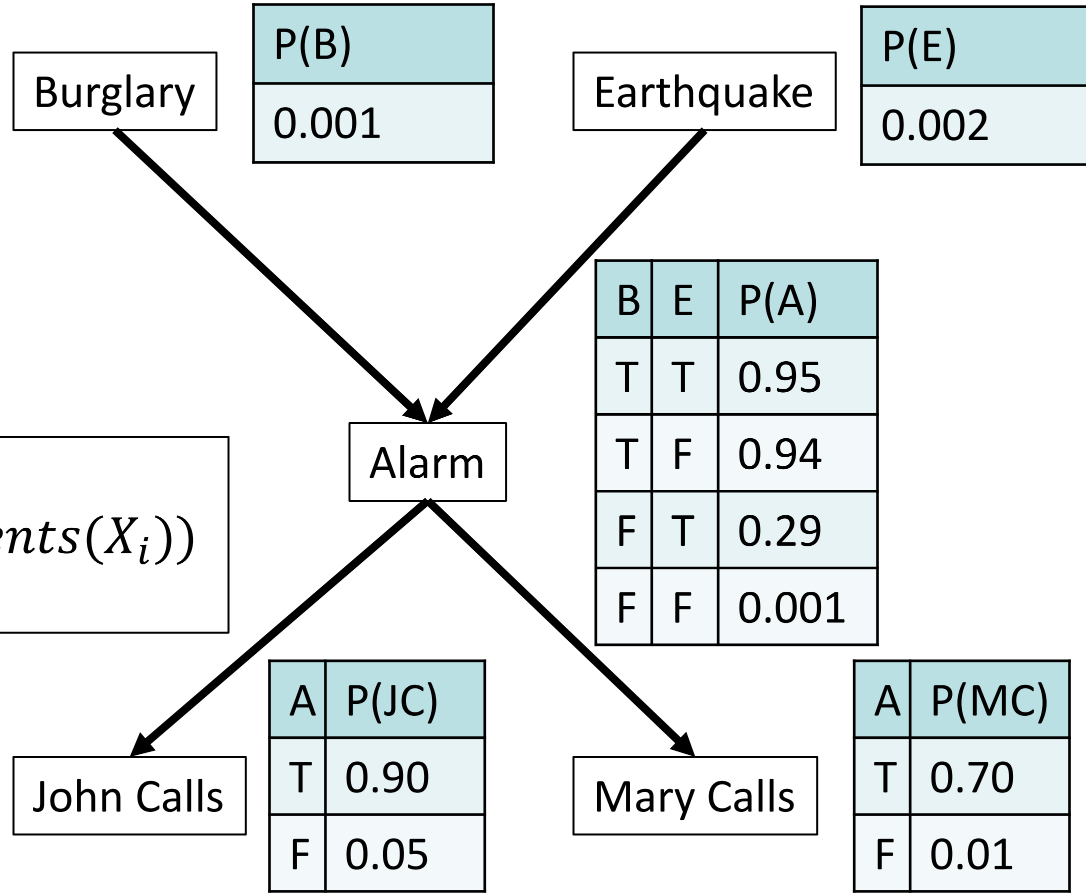
A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

$$P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- Variables E, A are unspecified.
  - Each variable is binary, so we must sum over four possible cases.

Another Example



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

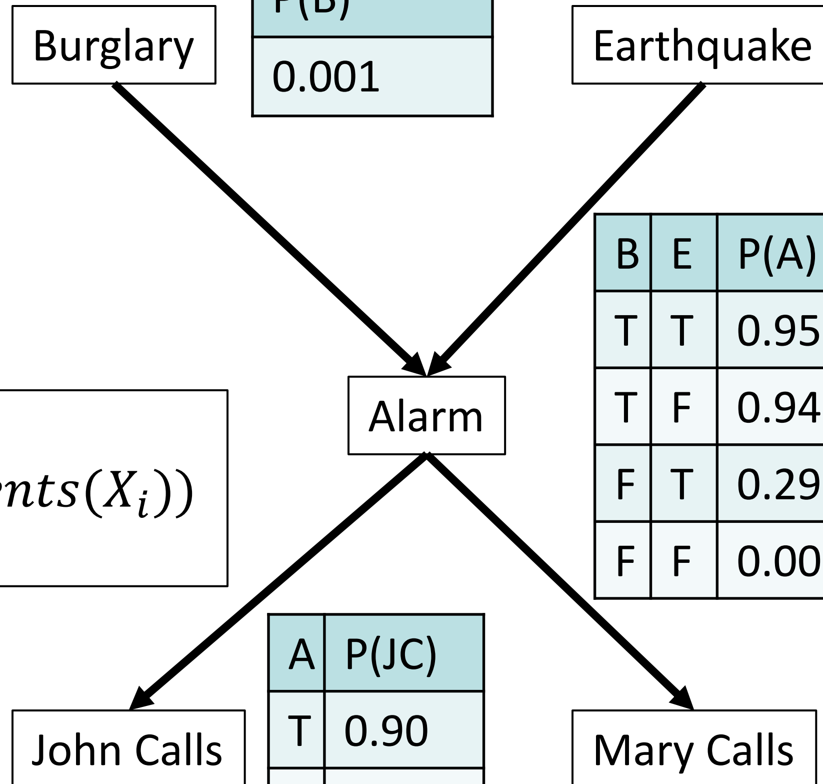
A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

$$P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- Variables E, A are unspecified.
  - Each variable is binary, so we must sum over four possible cases.
- P(¬B, JC, MC) = P(¬B, E, A, JC, MC) +
  P(¬B, E, ¬A, JC, MC) +
  P(¬B, ¬E, A, JC, MC) +
  P(¬B, ¬E, ¬A, JC, MC) = ???

Another Example



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(JC)
T	0.90
F	0.05

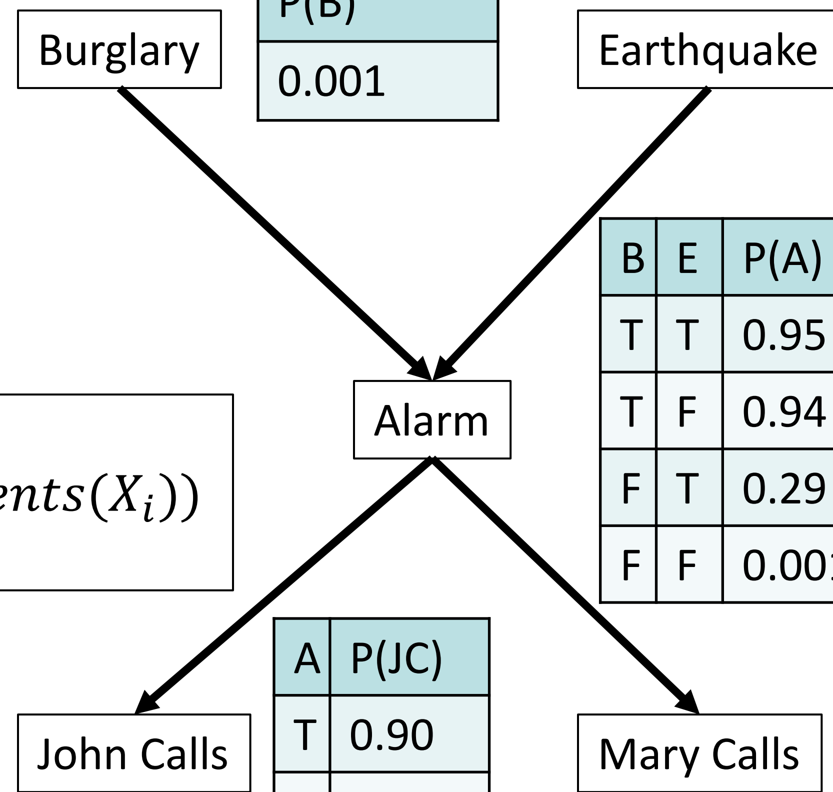
A	P(MC)
T	0.70
F	0.01

$$P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \,|\, parents(X_i))$$

- Variables E, A are unspecified.
  - Each variable is binary, so we must sum over four possible cases.
- P(¬B, JC, MC) = P(¬B, E, A, JC, MC) +
  P(¬B, E, ¬A, JC, MC) +
  P(¬B, ¬E, A, JC, MC) +
  P(¬B, ¬E, ¬A, JC, MC) = ???

Here we apply the equation to each of the four terms separately.

Another Example



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(JC)
T	0.90
F	0.05

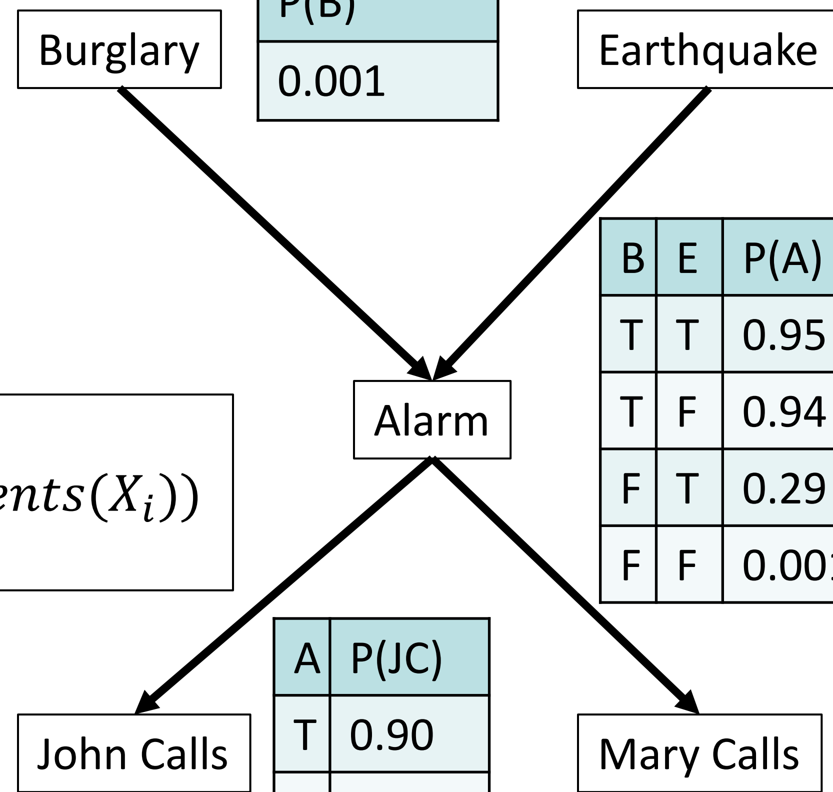
A	P(MC)
T	0.70
F	0.01

$$P(X_1, \dots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- Variables E, A are unspecified.
  - Each variable is binary, so we must sum over four possible cases.
- P(¬B, JC, MC) =

  P(¬B) * P(E) * P(A | ¬B, E) * P(JC | A) * P(MC | A) +
  P(¬B) * P(E) * P(¬A | ¬B, E) * P(JC | ¬A) * P(MC | ¬A) +
  P(¬B) * P(¬E) * P(A | ¬B, ¬E) * P(JC | A) * P(MC | A) +
  P(¬B) * P(¬E) * P(¬A | ¬B, ¬E) * P(JC | ¬A) * P(MC | ¬A)

Another Example



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(JC)
T	0.90
F	0.05

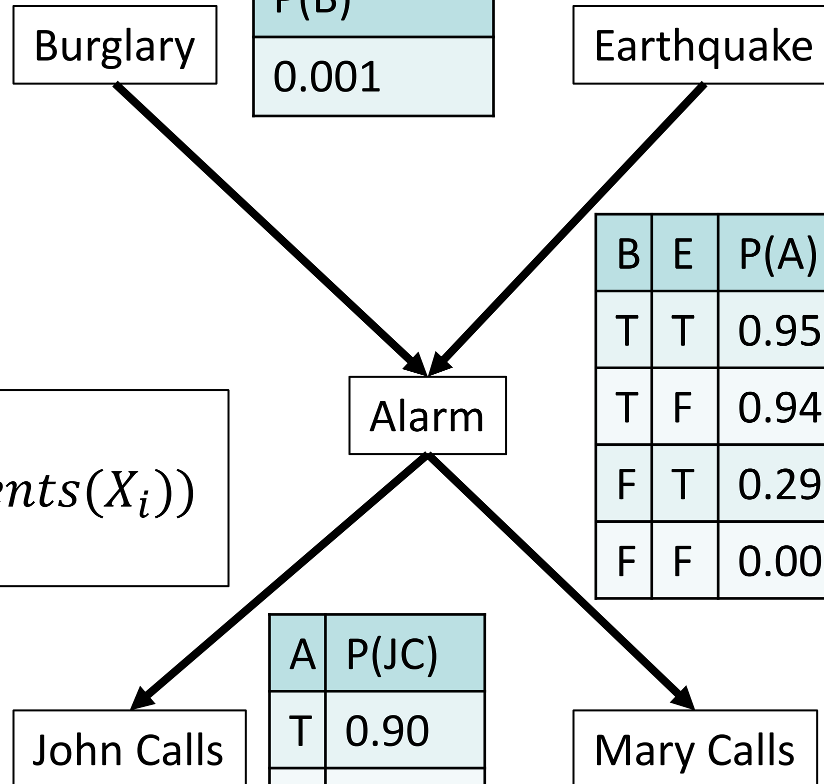
A	P(MC)
T	0.70
F	0.01

$$P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- Variables E, A are unspecified.
  - Each variable is binary, so we must sum over four possible cases.
- P(¬B, JC, MC) =

  0.999 * 0.002 * 0.290 * 0.90 * 0.70 +
  0.999 * 0.002 * 0.710 * 0.05 * 0.01 +
  0.999 * 0.998 * 0.001 * 0.90 * 0.70 +
  0.999 * 0.998 * 0.999 * 0.05 * 0.01

Another Example



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(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

$$P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

- Variables E, A are unspecified.
  - Each variable is binary, so we must sum over four possible cases.
- P(¬B, JC, MC) = 0.0003650 + 0.0000007 + 0.0006281 + 0.0004980
  = 0.0014918

Computing Conditional Probabilities

- So far we have seen how to compute, in Bayesian Networks, these types of probabilities:
  - $P(X_1, \ldots, X_n)$, where we specify values for all n variables of the network.
  - $P(A_1, \ldots, A_k)$, where we specify values for only k of the n variables of the network.
- We now need to cover the case of conditional probabilities:

  $P(A_1, \ldots, A_k \mid B_1, \ldots, B_m)$

- How can we compute this?

Computing Conditional Probabilities

- Using the definition of conditional probabilities, we get:

$$P(A_1, \dots, A_k \mid B_1, \dots, B_m) = \frac{P(A_1, \dots, A_k, B_1, \dots, B_m)}{P(B_1, \dots, B_m)}$$

Now, both the numerator and the denominator are probabilities that we already learned how to compute:

– They are probabilities where values are provided for some, but possibly not all, variables of the network.

Conditional Probability Example

- Here is a more interesting example:
  - John calls, to say the alarm is ringing.
  - Mary also calls, to say the alarm is ringing.
  - What is the probability there is a burglary?
- How do we write our question as a formula? What do we want to compute?

  $P(B \mid JC, MC)$

- How do we compute it? $P(B \mid JC, MC) = \dfrac{P(B, JC, MC)}{P(JC, MC)}$

Conditional Probability Example

- $P(B \mid JC, MC) = \dfrac{P(B, JC, MC)}{P(JC, MC)}$
- First let's compute the denominator, P(JC, MC):

P(JC, MC) =

$P(B, E, A, JC, MC) +$

$P(B, E, \neg A, JC, MC) +$

$P(B, \neg E, A, JC, MC) +$

$P(B, \neg E, \neg A, JC, MC) +$

$P(\neg B, E, A, JC, MC) +$

$P(\neg B, E, \neg A, JC, MC) +$

$P(\neg B, \neg E, A, JC, MC) +$

$P(\neg B, \neg E, \neg A, JC, MC) =$

Conditional Probability Example

- $P(B \mid JC, MC) = \dfrac{P(B, JC, MC)}{P(JC, MC)}$
- First let's compute the denominator, P(JC, MC):

$$
\begin{aligned}
P(JC, MC) = {} \\
& P(B) * P(E) * P(A \mid B, E) * P(JC \mid A) * P(MC \mid A) + \\
& P(B) * P(E) * P(\neg A \mid B, E) * P(JC \mid \neg A) * P(MC \mid \neg A) + \\
& P(B) * P(\neg E) * P(A \mid B, \neg E) * P(JC \mid A) * P(MC \mid A) + \\
& P(B) * P(\neg E) * P(\neg A \mid B, \neg E) * P(JC \mid \neg A) * P(MC \mid \neg A) + \\
& P(\neg B) * P(E) * P(A \mid \neg B, E) * P(JC \mid A) * P(MC \mid A) + \\
& P(\neg B) * P(E) * P(\neg A \mid \neg B, E) * P(JC \mid \neg A) * P(MC \mid \neg A) + \\
& P(\neg B) * P(\neg E) * P(A \mid \neg B, \neg E) * P(JC \mid A) * P(MC \mid A) + \\
& P(\neg B) * P(\neg E) * P(\neg A \mid \neg B, \neg E) * P(JC \mid \neg A) * P(MC \mid \neg A) =
\end{aligned}
$$

Conditional Probability Example

- $P(B \mid JC, MC) = \dfrac{P(B, JC, MC)}{P(JC, MC)}$
- First let's compute the denominator, P(JC, MC):

P(JC, MC) =

$$
\begin{aligned}
&0.001 * 0.002 * 0.950 * 0.90 * 0.70 + \\
&0.001 * 0.002 * 0.050 * 0.05 * 0.01 + \\
&0.001 * 0.998 * 0.940 * 0.90 * 0.70 + \\
&0.001 * 0.998 * 0.060 * 0.05 * 0.01 + \\
&0.999 * 0.002 * 0.290 * 0.90 * 0.70 + \\
&0.999 * 0.002 * 0.710 * 0.05 * 0.01 + \\
&0.999 * 0.998 * 0.001 * 0.90 * 0.70 + \\
&0.999 * 0.998 * 0.999 * 0.05 * 0.01 =
\end{aligned}
$$

Conditional Probability Example

- $P(B \mid JC, MC) = \dfrac{P(B, JC, MC)}{P(JC, MC)}$
- First let's compute the denominator, P(JC, MC):

P(JC, MC) =

0.000001197 +

0.000000000 +

0.000591015 +

0.000000030 +

0.000365034 +

0.000000709 +

0.000628111 +

0.000498002

Conditional Probability Example

- $P(B \mid JC, MC) = \dfrac{P(B, JC, MC)}{P(JC, MC)}$
- First let's compute the denominator, P(JC, MC):

P(JC, MC) = 0.002084098

Conditional Probability Example

- $P(B \mid JC, MC) = \dfrac{P(B, JC, MC)}{P(JC, MC)}$
- First let's compute the denominator, P(JC, MC):

P(JC, MC) = 0.002084098

- Now, let's compute the numerator, P(B, JC, MC):
  - Note: this is a sum over only a subset of the cases that we included in the denominator. So, we have already done most of the work:

Conditional Probability Example

- $P(B \mid JC, MC) = \dfrac{P(B, JC, MC)}{P(JC, MC)}$
- First let's compute the denominator, P(JC, MC):

P(JC, MC) = 0.002084098

- Now, let's compute the numerator, P(B, JC, MC):

P(B, JC, MC) =

$P(B, E, A, JC, MC) +$

$P(B, E, \neg A, JC, MC) +$

$P(B, \neg E, A, JC, MC) +$

$P(B, \neg E, \neg A, JC, MC)$

Conditional Probability Example

- $P(B \mid JC, MC) = \dfrac{P(B, JC, MC)}{P(JC, MC)}$
- First let's compute the denominator, P(JC, MC):

P(JC, MC) = 0.002084098

- Now, let's compute the numerator, P(B, JC, MC):

P(B, JC, MC) =

$$
\begin{aligned}
&P(B) * P(E) * P(A \mid B, E) * P(JC \mid A) * P(MC \mid A) + \\
&P(B) * P(E) * P(\neg A \mid B, E) * P(JC \mid \neg A) * P(MC \mid \neg A) + \\
&P(B) * P(\neg E) * P(A \mid B, \neg E) * P(JC \mid A) * P(MC \mid A) + \\
&P(B) * P(\neg E) * P(\neg A \mid B, \neg E) * P(JC \mid \neg A) * P(MC \mid \neg A)
\end{aligned}
$$

Conditional Probability Example

- $P(B \mid JC, MC) = \dfrac{P(B, JC, MC)}{P(JC, MC)}$
- First let's compute the denominator, P(JC, MC):

P(JC, MC) = 0.002084098

- Now, let's compute the numerator, P(B, JC, MC):

P(B, JC, MC) =

0.001 * 0.002 * 0.950 * 0.90 * 0.70 +

0.001 * 0.002 * 0.050 * 0.05 * 0.01 +

0.001 * 0.998 * 0.940 * 0.90 * 0.70 +

0.001 * 0.998 * 0.060 * 0.05 * 0.01

Conditional Probability Example

- $P(B \mid JC, MC) = \dfrac{P(B, JC, MC)}{P(JC, MC)}$
- First let's compute the denominator, P(JC, MC):

P(JC, MC) = 0.002084098

- Now, let's compute the numerator, P(B, JC, MC):

P(B, JC, MC) =

0.000001197 +

0.000000000 +

0.000591015 +

0.000000030

Conditional Probability Example

- $P(B \mid JC, MC) = \dfrac{P(B, JC, MC)}{P(JC, MC)}$
- First let's compute the denominator, P(JC, MC):

P(JC, MC) = 0.002084098

- Now, let's compute the numerator, P(B, JC, MC):

P(B, JC, MC) = 0.000592242

- Therefore, $P(B \mid JC, MC) = \dfrac{0.000592242}{0.002084098} = 0.284.$
- There is a 28.4% probability that there was a burglary.

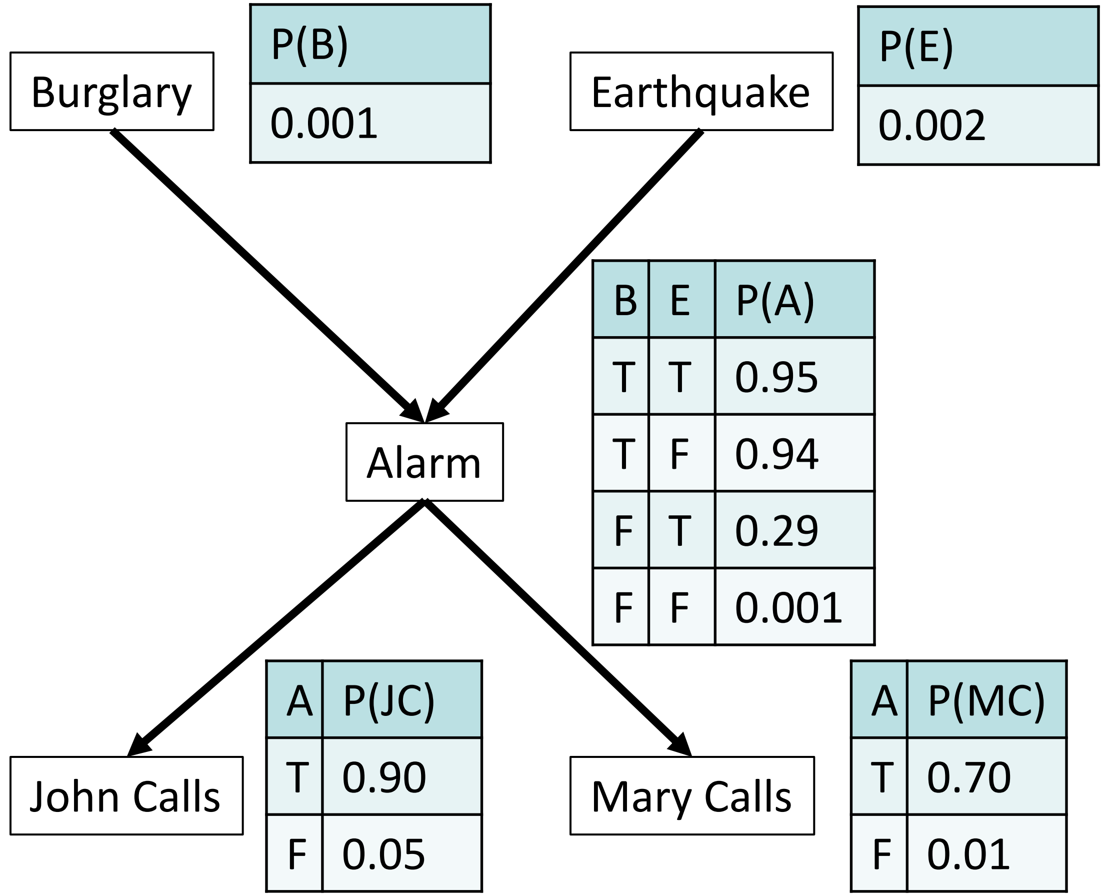
Complexity of Inference

- What is the complexity of the inference algorithm we have been using in the previous examples?
- We sum over probabilities of various combinations of values.
- In the worst case, how many combinations of values do we need to consider?
  - All possible combinations of values of all variables in the Bayesian network.
- This is NOT any faster than inference by enumeration using a joint distribution table.
  - We are still doing inference by enumeration, but using a Bayesian network.
- As mentioned before, in some cases (but not always) there are polynomial time inference algorithms for Bayesian networks (e.g., the **variable elimination algorithm**, textbook chapter 14.4.2).
- However, we will not go over such algorithms in this course.

Complexity of Inference

- So, our inference method using Bayesian networks is not any faster than using joint distribution tables.
- The big advantage over using joint distribution tables is space.
- To define a joint distribution table, we need space exponential to n (the number of variables).
- To define a Bayesian network, the space we need is linear to n, and exponential to r, where:
  - n is the number of variables.
  - r is the maximum number of parents that any node in the network has.
- In the typical case, r << n, and thus Bayesian networks require much fewer numbers to be specified, compared to joint distribution tables.

Simplified Calculations



P(B)
0.001

P(E)
0.002

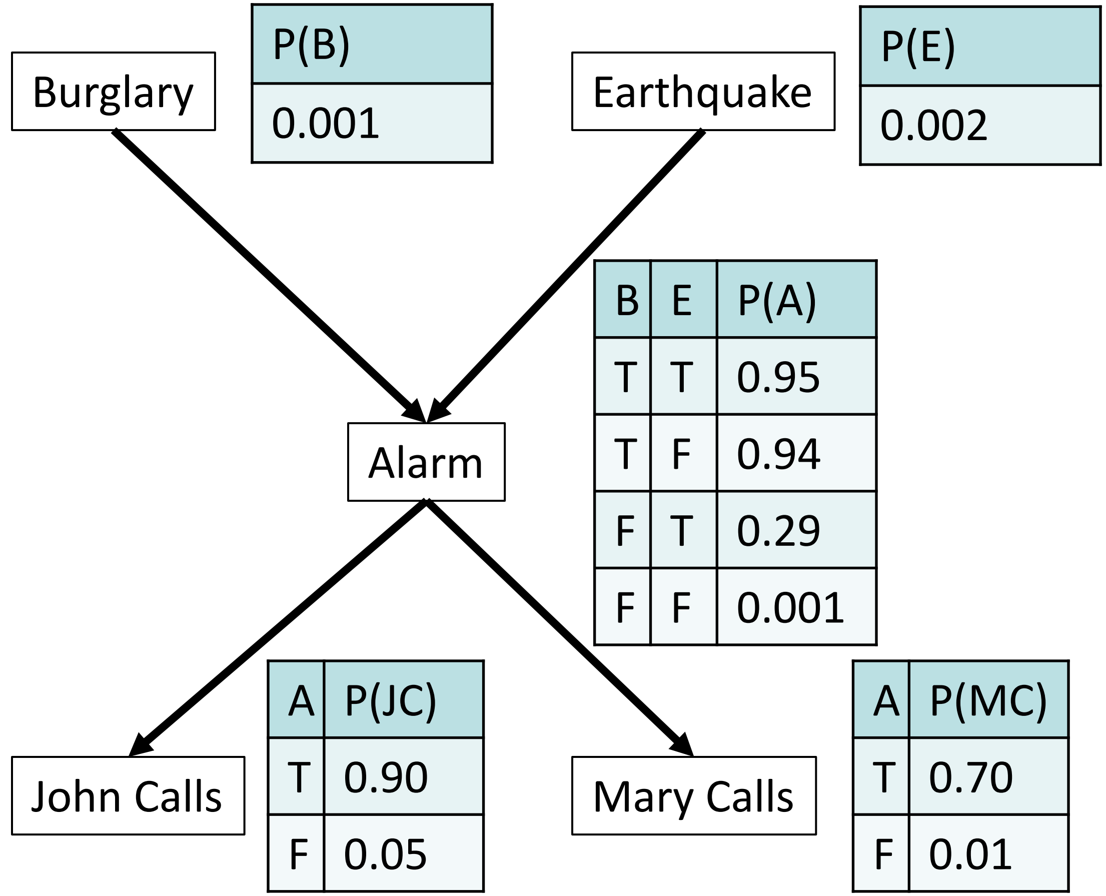
B	E	P(A)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

- Some times, we can compute some probabilities in a more simple manner than using enumeration.
- For example: compute P(B, E).
  - We could sum over the eight possible combinations of A, JC, MC.
  - Or, we could just remember that B and E are independent, so:
    P(B, E) = P(B) * P(E) = 0.001 * 0.002.

Simplified Calculations

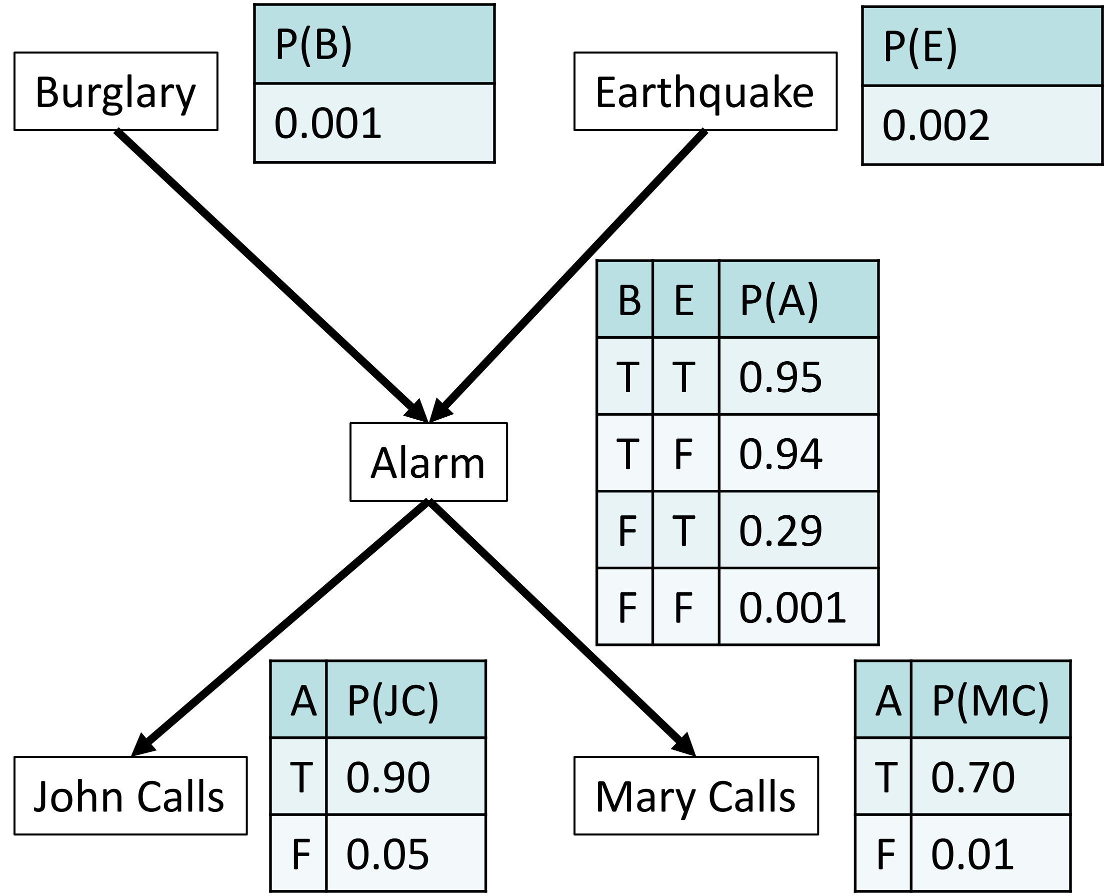


- Another example: compute P(JC, ¬MC | A).
- Again, we can do inference by enumeration, or we can simply recognize that JC and MC are conditionally independent given A.
- Therefore, P(JC, ¬MC | A) = P(JC | A) * P(¬MC | A) = 0.9 * 0.3.

Markov Blanket

- A node A is conditionally independent of any other node in the network, as long as we know the values of:
  - The parents of A.
  - The children of A.
  - The parents of the children of A.
- This set of nodes (parents, children, children's parents) is called the **Markov Blanket** of A.

Markov Blanket



P(B)
0.001

P(E)
0.002

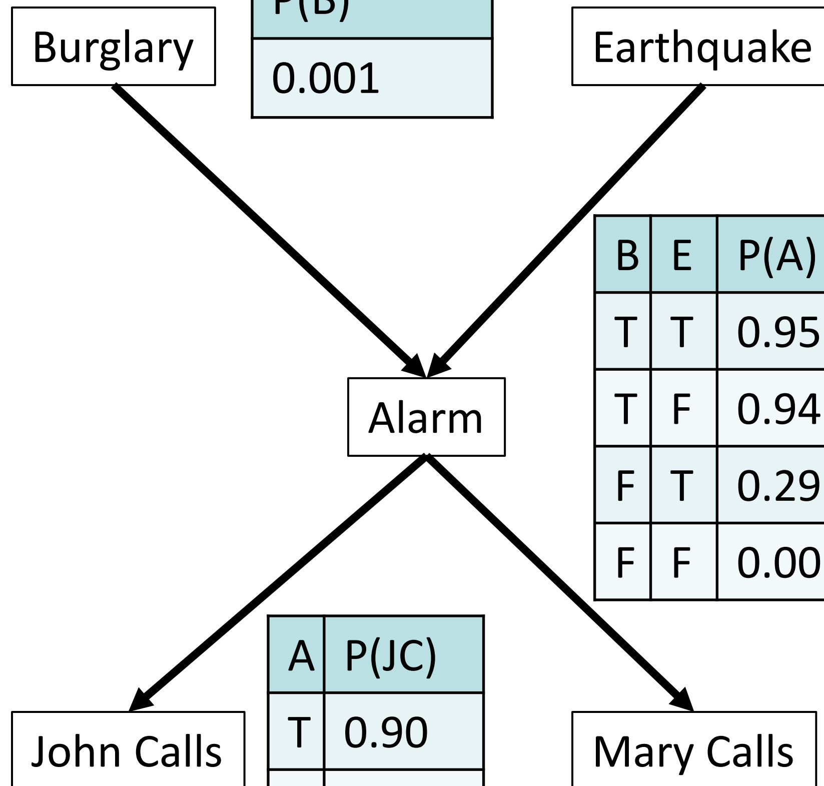
B	E	P(A)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

- Why do we also need values for the children’s parents?
- Here is an example: are B and E conditionally independent given A?
- How do we approach that question? What quantities do we need to compute?

Markov Blanket



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(JC)
T	0.90
F	0.05

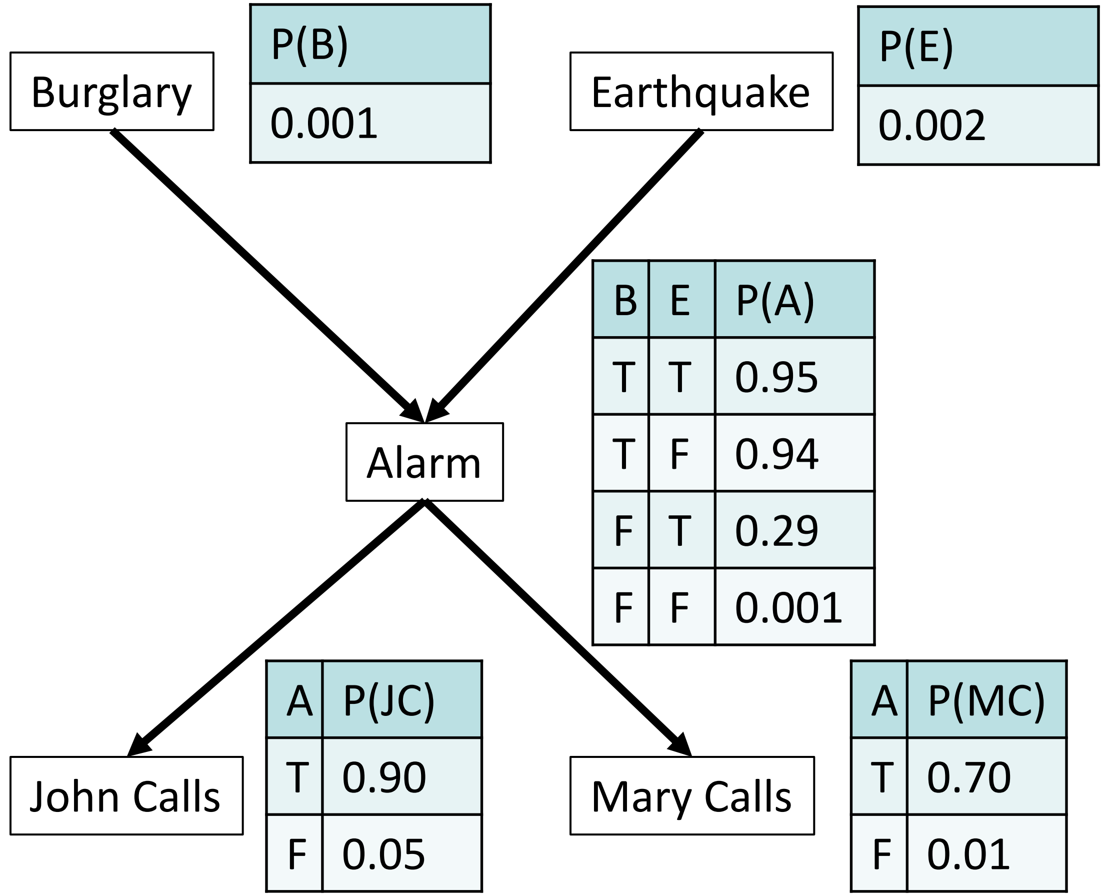
A	P(MC)
T	0.70
F	0.01

- Are B and E conditionally independent given A?
- To answer this, we need to compare two quantities:

P(B | A) and P(B | A, E).

- If those two quantities are equal, then B and E are conditionally independent given A.

Markov Blanket



P(B)
0.001

P(E)
0.002

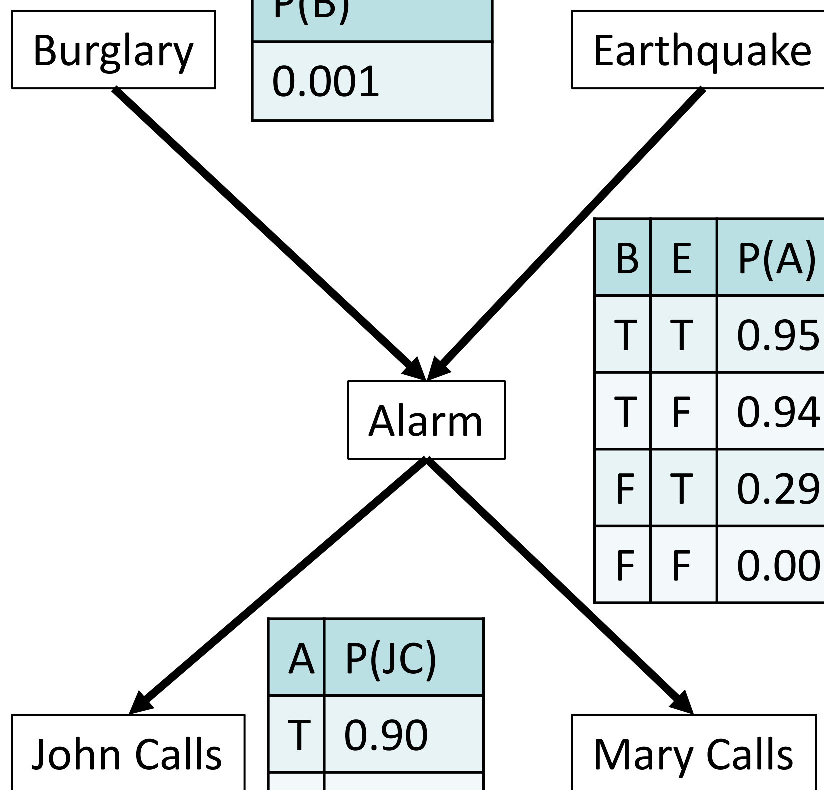
B	E	P(A)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

- P(B | A) = ???

Markov Blanket



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

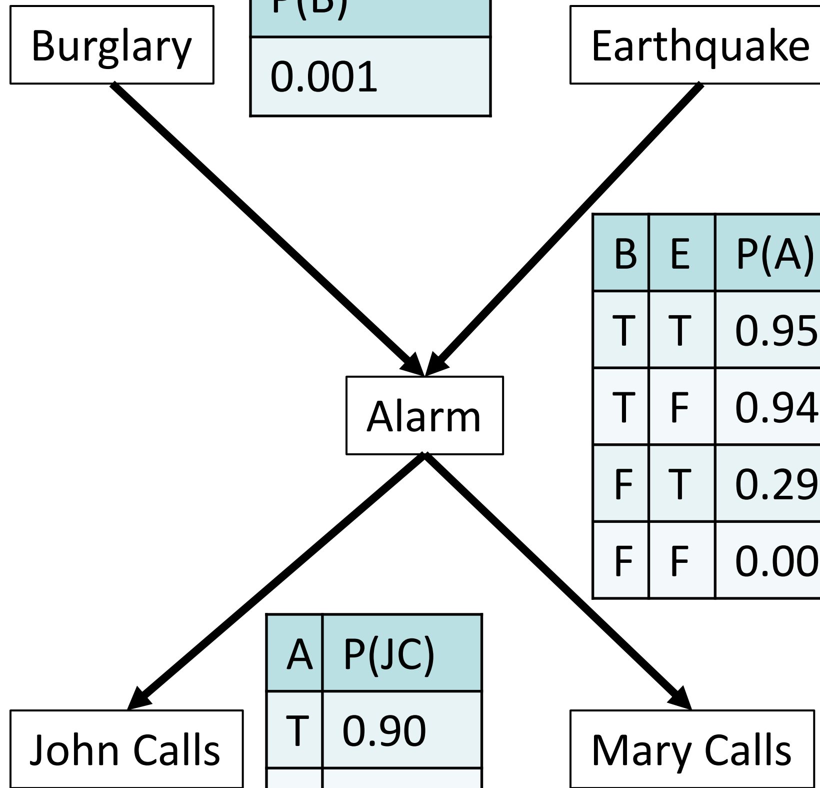
A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

$$\mathrm{P(B \mid A)} = \frac{P(A,B)}{P(A)} = \frac{P(A,B,E)+P(A,B,\neg E)}{P(A,B,E)+P(A,B,\neg E)+P(A,\neg B,E)+P(A,\neg B,\neg E)}$$

$$= \frac{P(B)*P(E)*P(A\,|B,E)+P(B)*P(\neg E)*P(A\,|B,\neg E)}{P(A,B,E)+P(A,B,\neg E)+P(\neg B)*P(E)*P(A\,|\neg B,E)+P(\neg B)*P(\neg E)*P(A\,|\neg B,\neg E)}$$

Markov Blanket



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(JC)
T	0.90
F	0.05

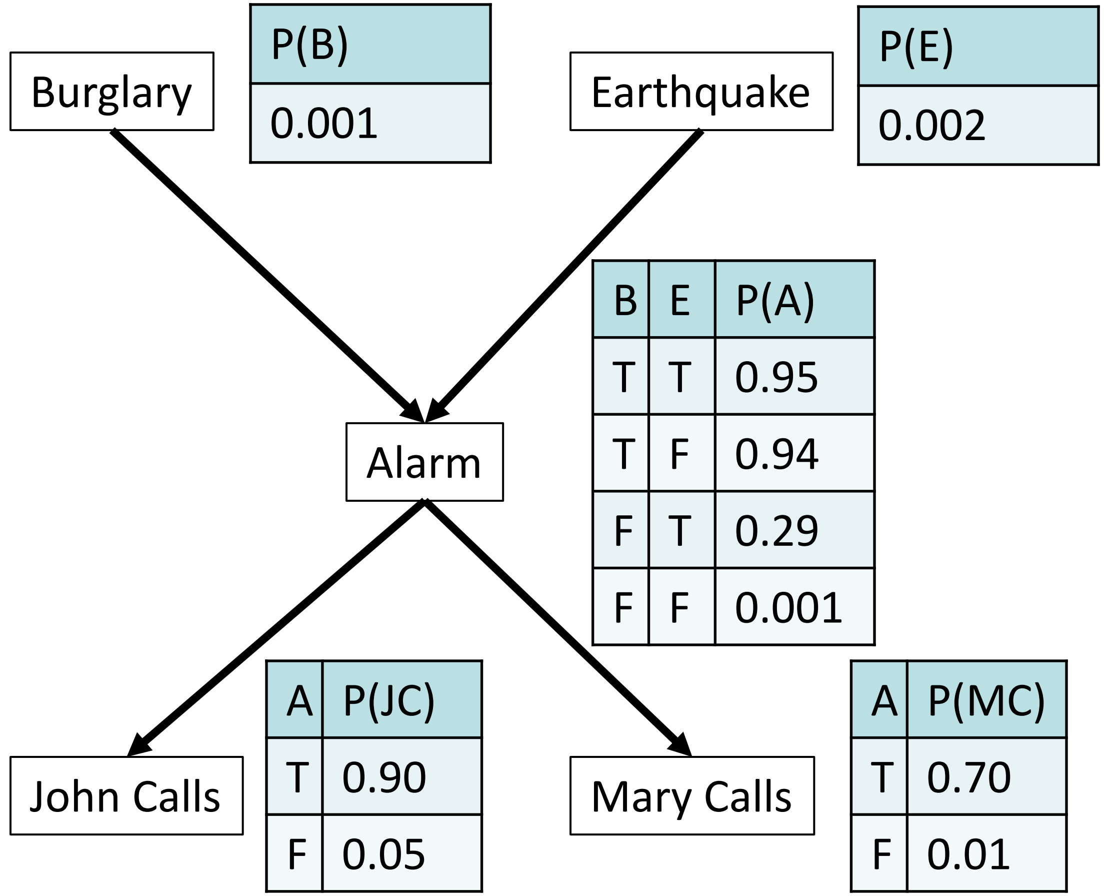
A	P(MC)
T	0.70
F	0.01

$$\frac{P(B)*P(E)*P(A\,|B,E)+P(B)*P(\neg E)*P(A\,|B,\neg E)}{P(A,B,E)+P(A,B,\neg E)+P(\neg B)*P(E)*P(A\,|\neg B,E)+P(\neg B)*P(\neg E)*P(A\,|\neg B,\neg E)}=$$

$$\frac{0.001*0.002*0.95+0.001*0.998*0.94}{0.001*0.002*0.95+0.001*0.998*0.94+0.999*0.002*0.29+0.999*0.998*0.001}=$$

$$\frac{0.00094002}{0.00251644}=0.3735.$$

Markov Blanket



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(JC)
T	0.90
F	0.05

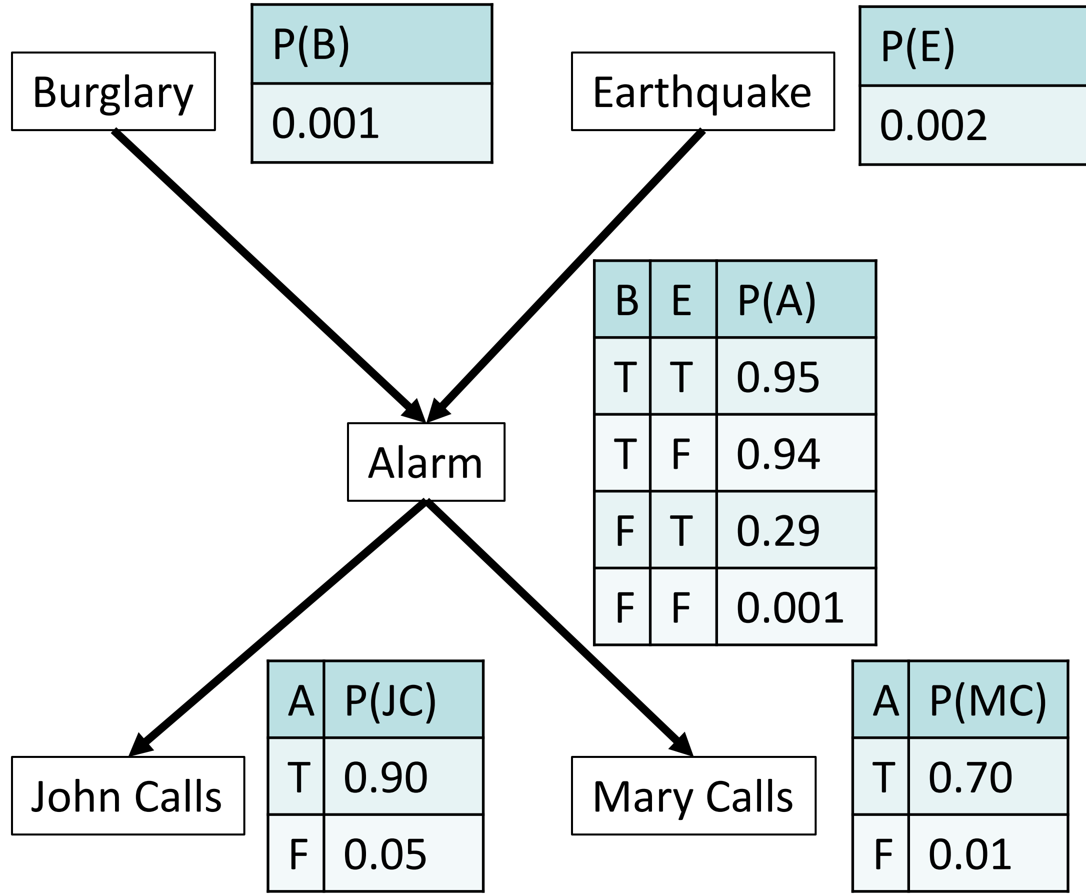
A	P(MC)
T	0.70
F	0.01

$$\mathrm{P(B \mid A, E)} = \frac{P(A,B,E)}{P(A,E)} = \frac{P(A,B,E)}{P(A,B,E)+P(A,\neg B,E)}$$

$$= \frac{P(B) * P(E) * P(A \mid B,E)}{P(B) * P(E) * P(A \mid B,E) + P(\neg B) * P(E) * P(A \mid \neg B,E)} =$$

$$\frac{0.001*0.002*0.95}{0.001*0.002*0.95 + 0.999*0.002*0.29} = \frac{0.0000019}{0.0005813} = 0.0032$$

Markov Blanket



P(B)
0.001

P(E)
0.002

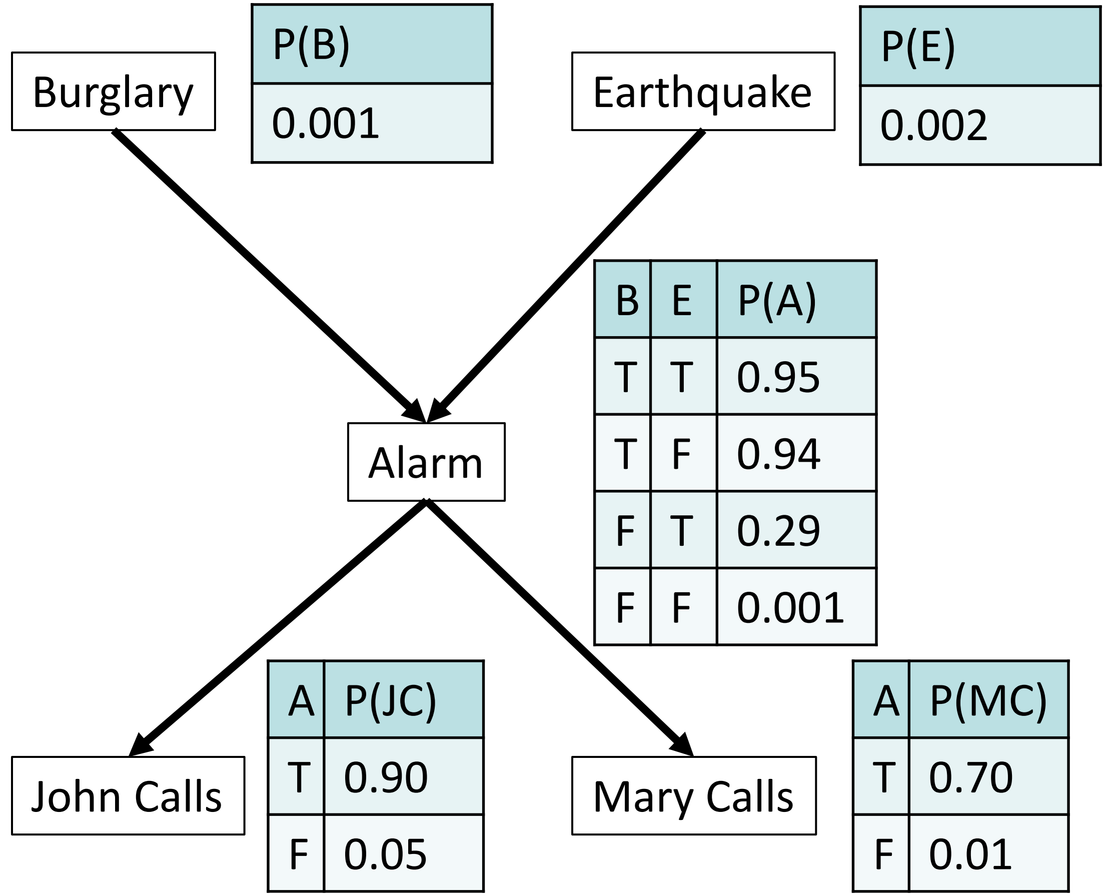
B	E	P(A)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

- Are B and E conditionally independent given A?
- To answer this, we need to compare: two quantities:
  - P(B | A) = 37.3%.
  - P(B | A, E) = 0.33%.
- Conclusion: B and E are NOT conditionally independent given A.

Markov Blanket



P(B)
0.001

P(E)
0.002

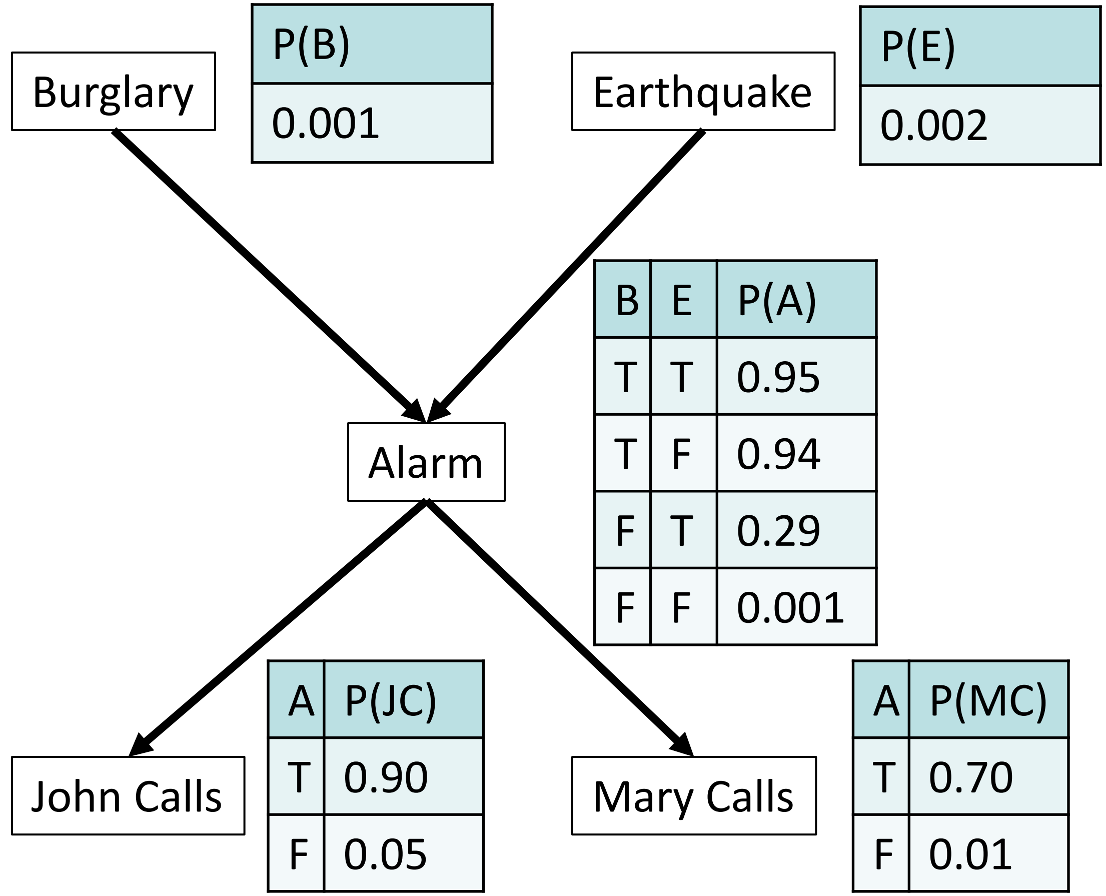
B	E	P(A)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

- Intuitively, how can we explain these results?
  - P(B | A) = 37.3%.
  - P(B | A, E) = 0.33%.

Markov Blanket



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(JC)
T	0.90
F	0.05

A	P(MC)
T	0.70
F	0.01

- Intuitively, how can we explain these results?
  - P(B | A) = 37.3%.
  - P(B | A, E) = 0.33%.
- If we know there was an alarm, burglary becomes more likely, because burglary causes alarms.
- However, if we know there was an alarm AND an earthquake, then the earthquake “explains” the alarm, and burglary becomes much less likely.

The Influence of Children’s Parents

- Overall, a child’s multiple parents are all possible causes for the child.
- If the child (the effect) is true, that makes all causes more likely.
- However, if the child is true, and some causes are also true, that makes other causes less likely.

The Influence of Children's Parents

- Another example: suppose you turn on the light in your room, and the light does not turn on.
- What is your first thought?

The Influence of Children's Parents

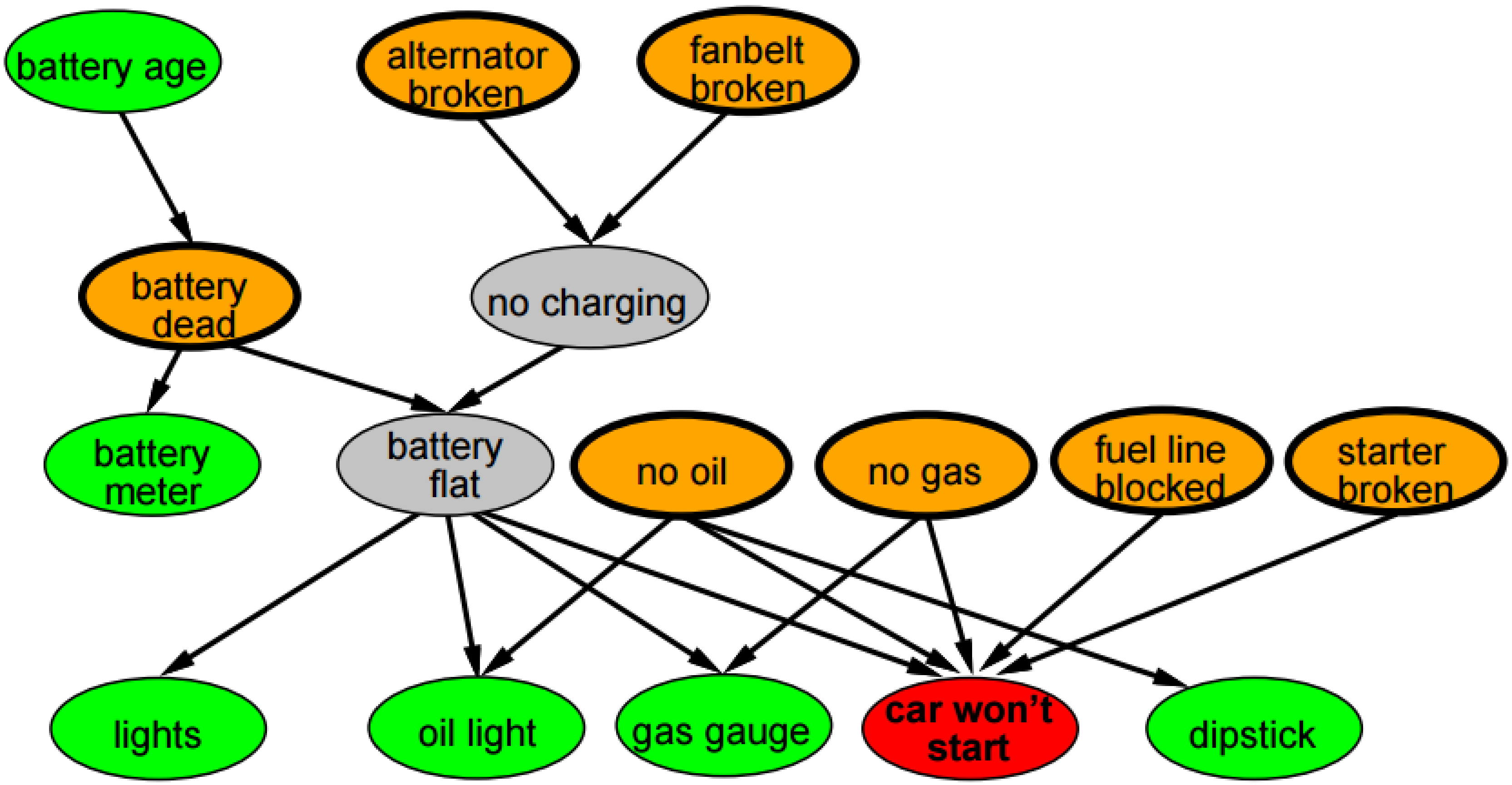
- Another example: suppose you turn on the light in your room, and the light does not turn on.
- What is your first thought?
  - Different people may answer this differently, but my first thought would be that the light is burned out.
- Now, suppose that I find out that there is a blackout.
  - What do you think now about the chance of the light being burned out?

The Influence of Children's Parents

- Another example: suppose you turn on the light in your room, and the light does not turn on.
- What is your first thought?
  - Different people may answer this differently, but my first thought would be that the light is burned out.
- Now, suppose that I find out that there is a blackout.
  - Then, I don't really think anymore that the light is burned out.
- So, if we define:
  - LB to stand for the light burned out.
  - LNTO to stand for the light not turning on.
  - BO to stand for black-out.
- $P(LB \mid LNTO) > P(LB)$, because LB is the most likely cause for LNTO.
- $P(LB \mid LNTO, BO) < P(LB \mid LNTO)$, because BO explains LNTO.

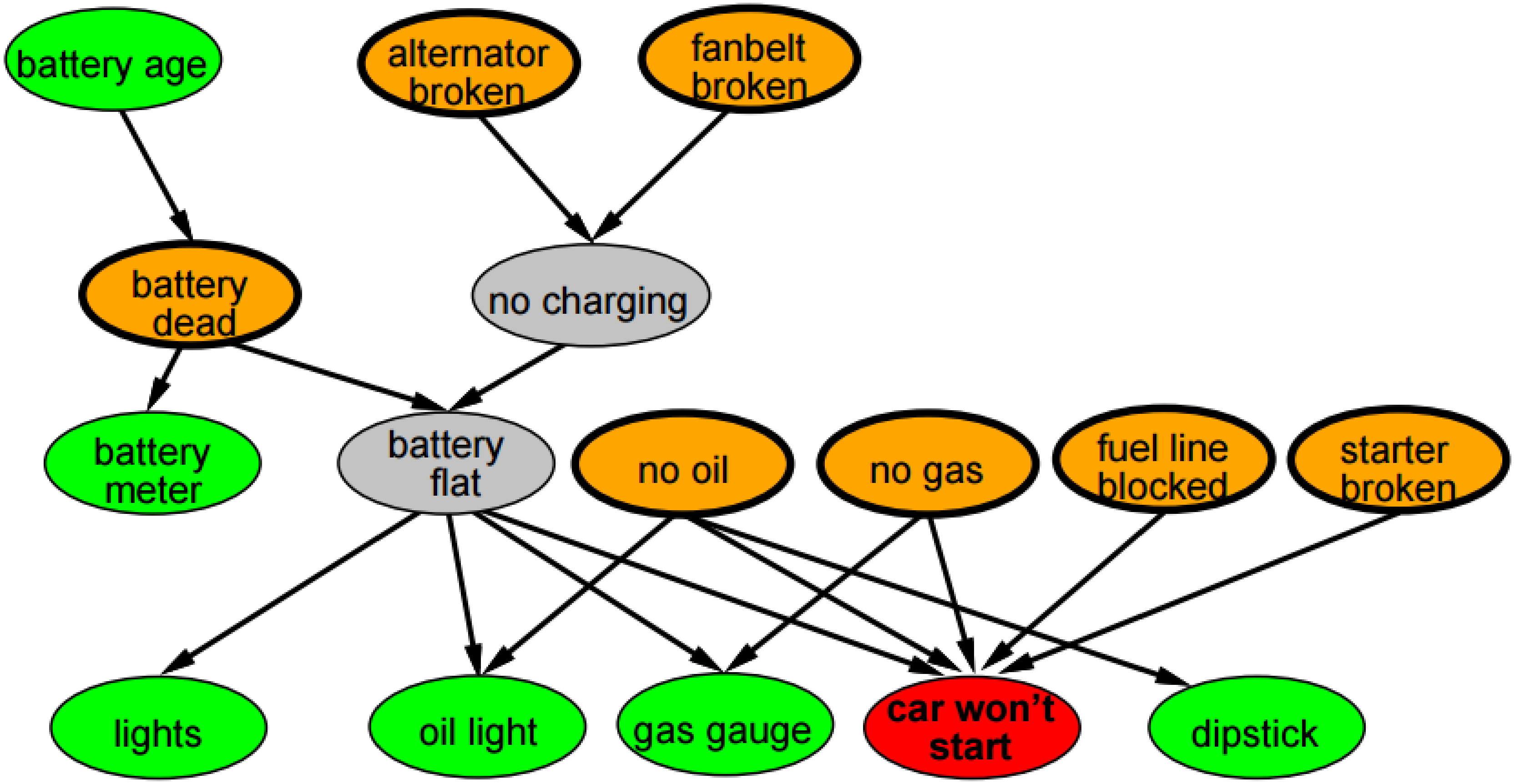
Markov Blanket

- Here is a more complicated network, from the textbook, for car diagnostics.
- How does P(battery dead) compare to P(battery dead | car won't start)?



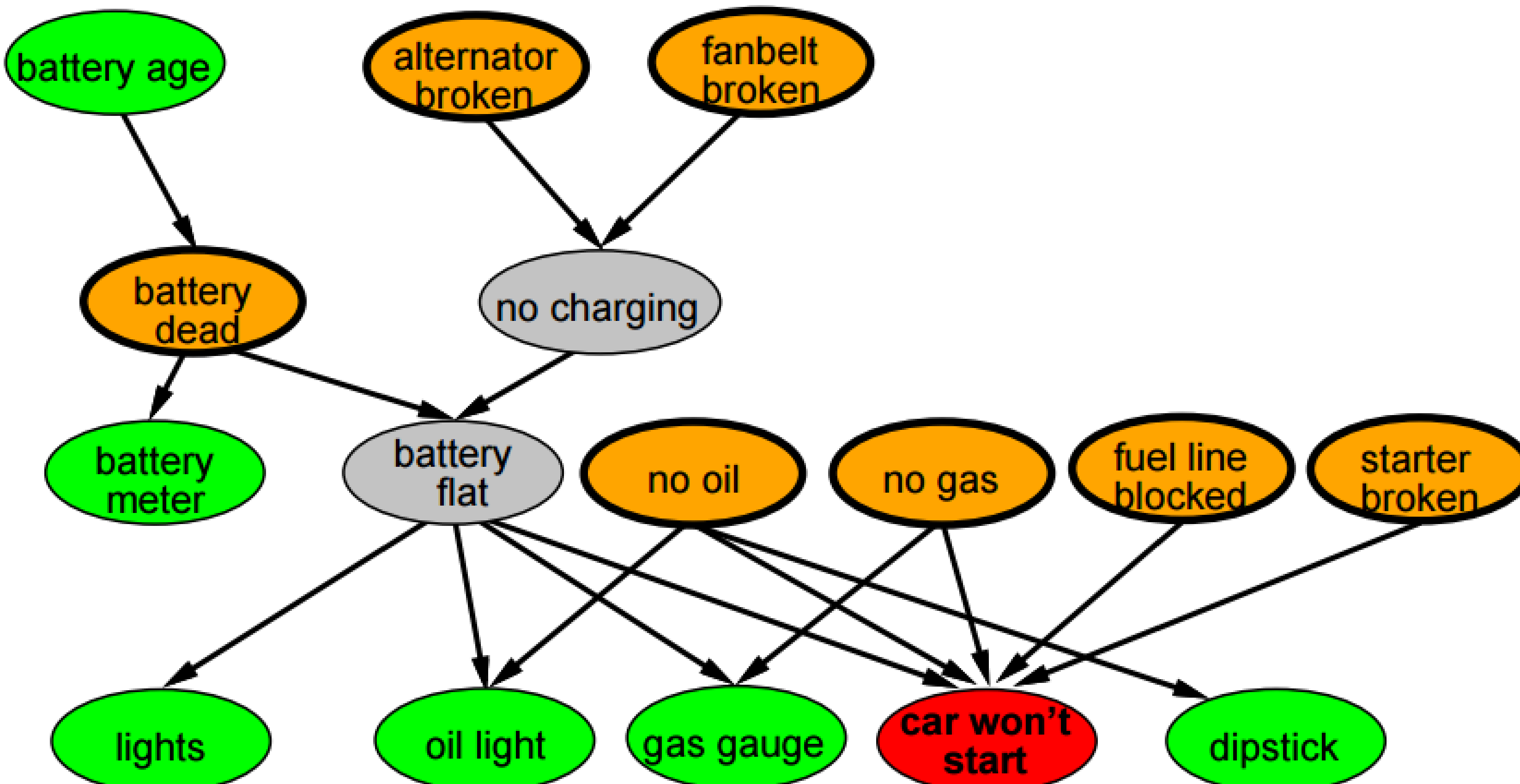
Markov Blanket

- Here is a more complicated network, from the textbook, for car diagnostics.
- How does P(battery dead) compare to P(battery dead | car won't start)?
- P(battery dead) < P(battery dead | car won't start), since "battery dead" is a possible cause (indirectly, through "battery flat") of "car won't start".



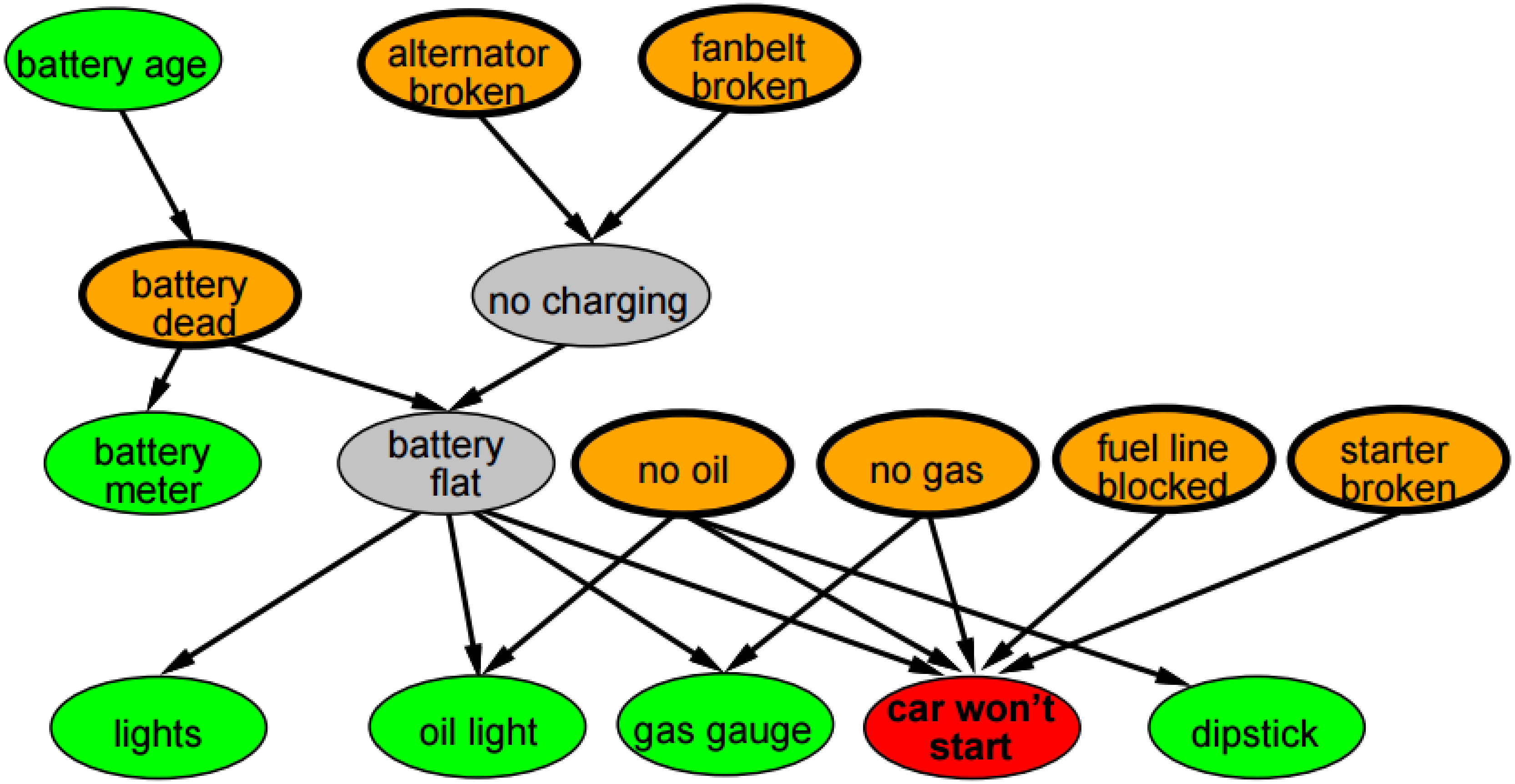
Markov Blanket

- How does P(battery dead | battery flat) compare to P(battery dead | battery flat, car won't start)?



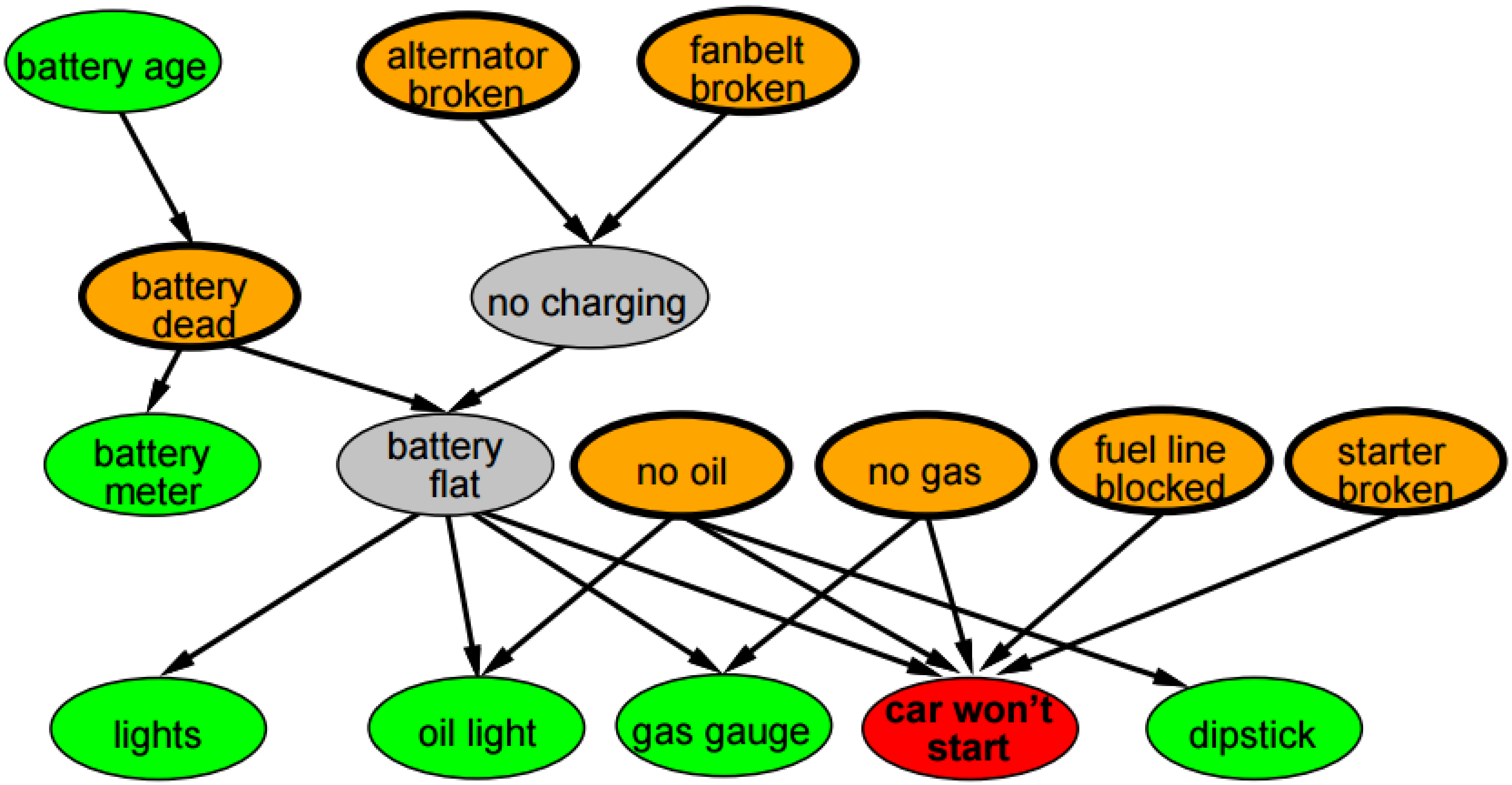
Markov Blanket

- P(battery dead | battery flat) = P(battery dead | battery flat, car won't start). If we know that the battery is flat, the fact that the car won't start does not tell us anything more about the battery being dead.



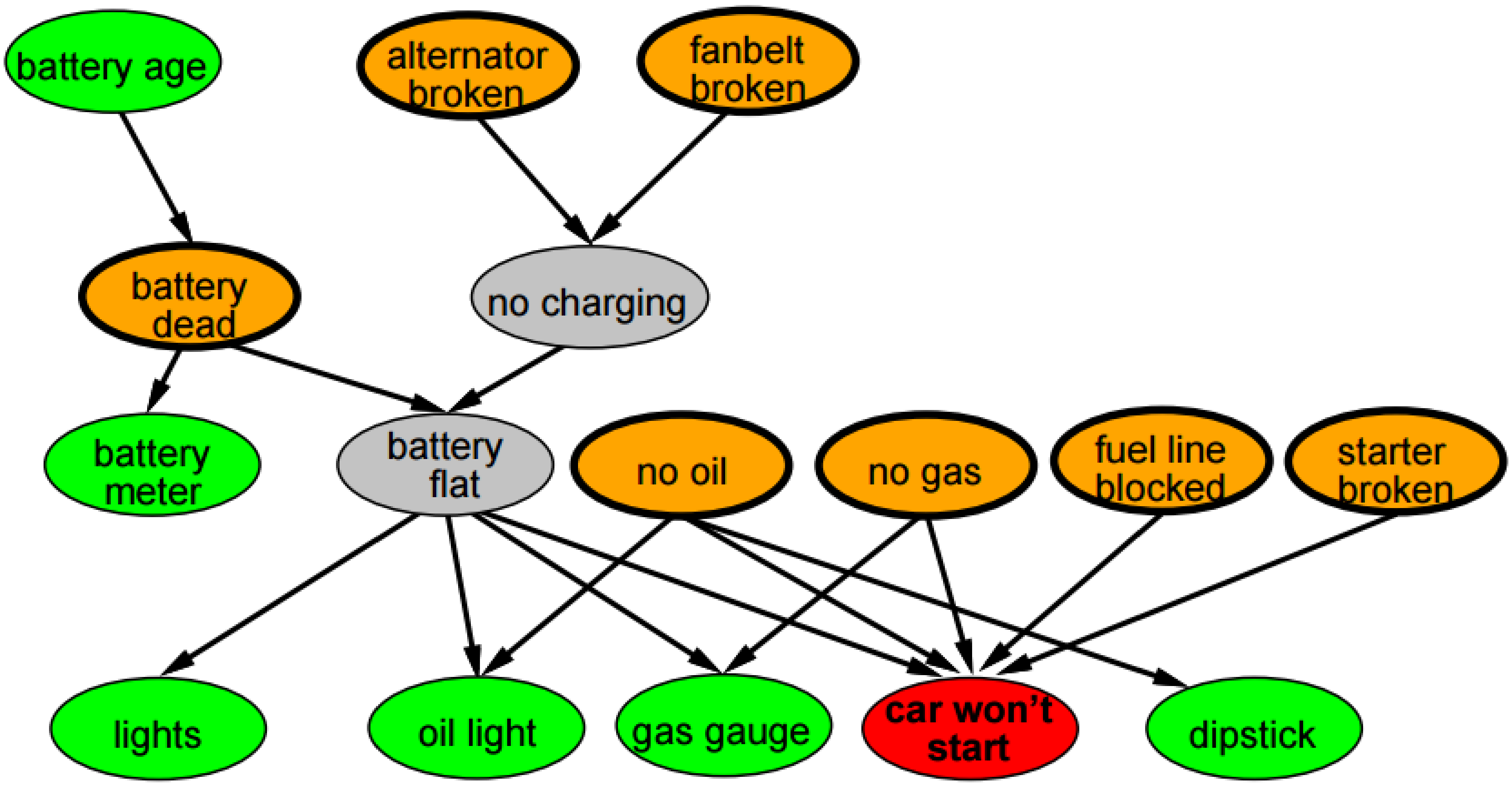
Markov Blanket

- How does P(battery dead | battery flat) compare to P(battery dead | battery flat, no charging)?



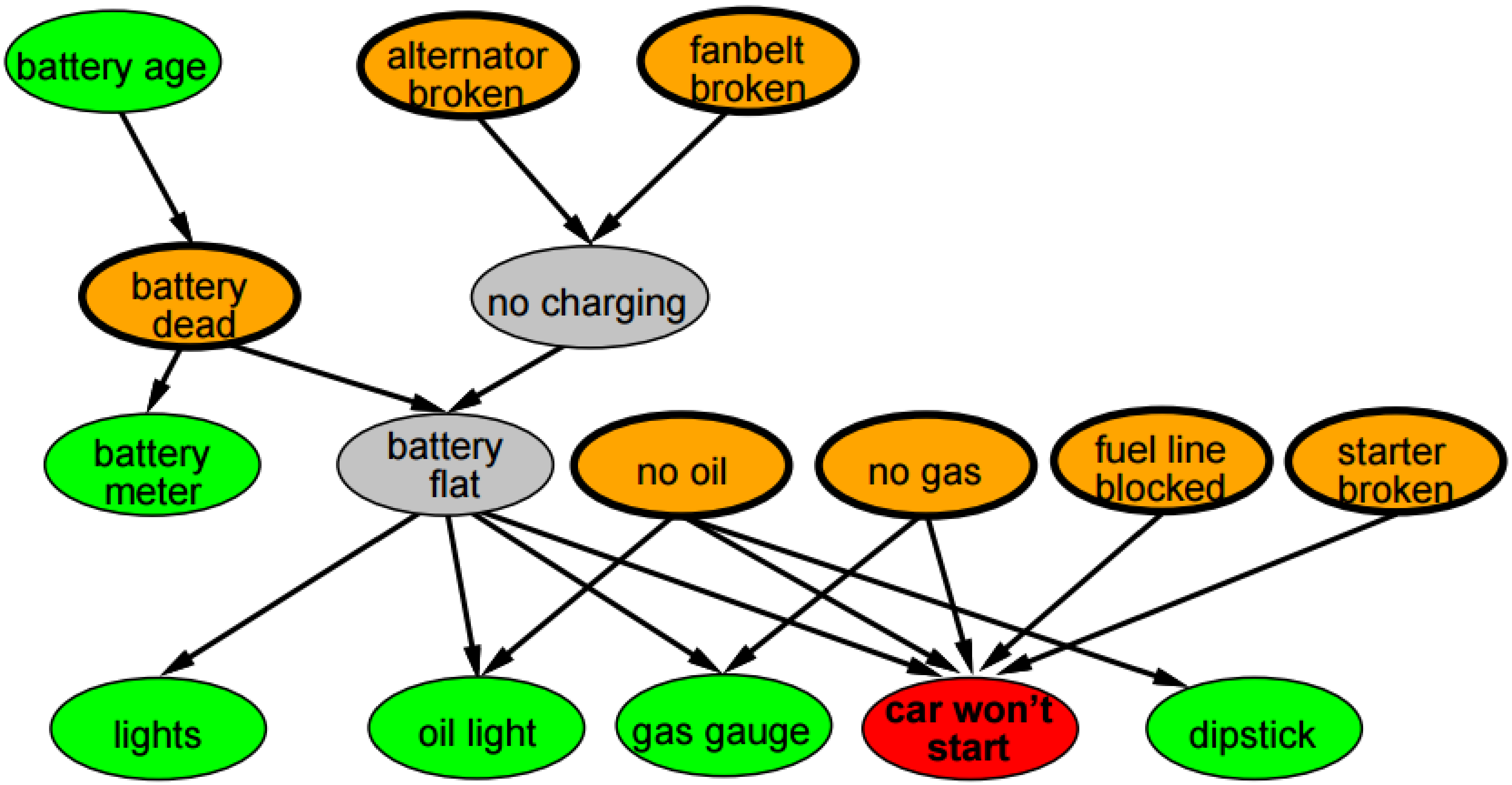
Markov Blanket

- How does P(battery dead | battery flat) compare to P(battery dead | battery flat, no charging)?
- P(battery dead | battery flat) > P(battery dead | battery flat, no charging), since "no charging" is another cause of "battery flat".



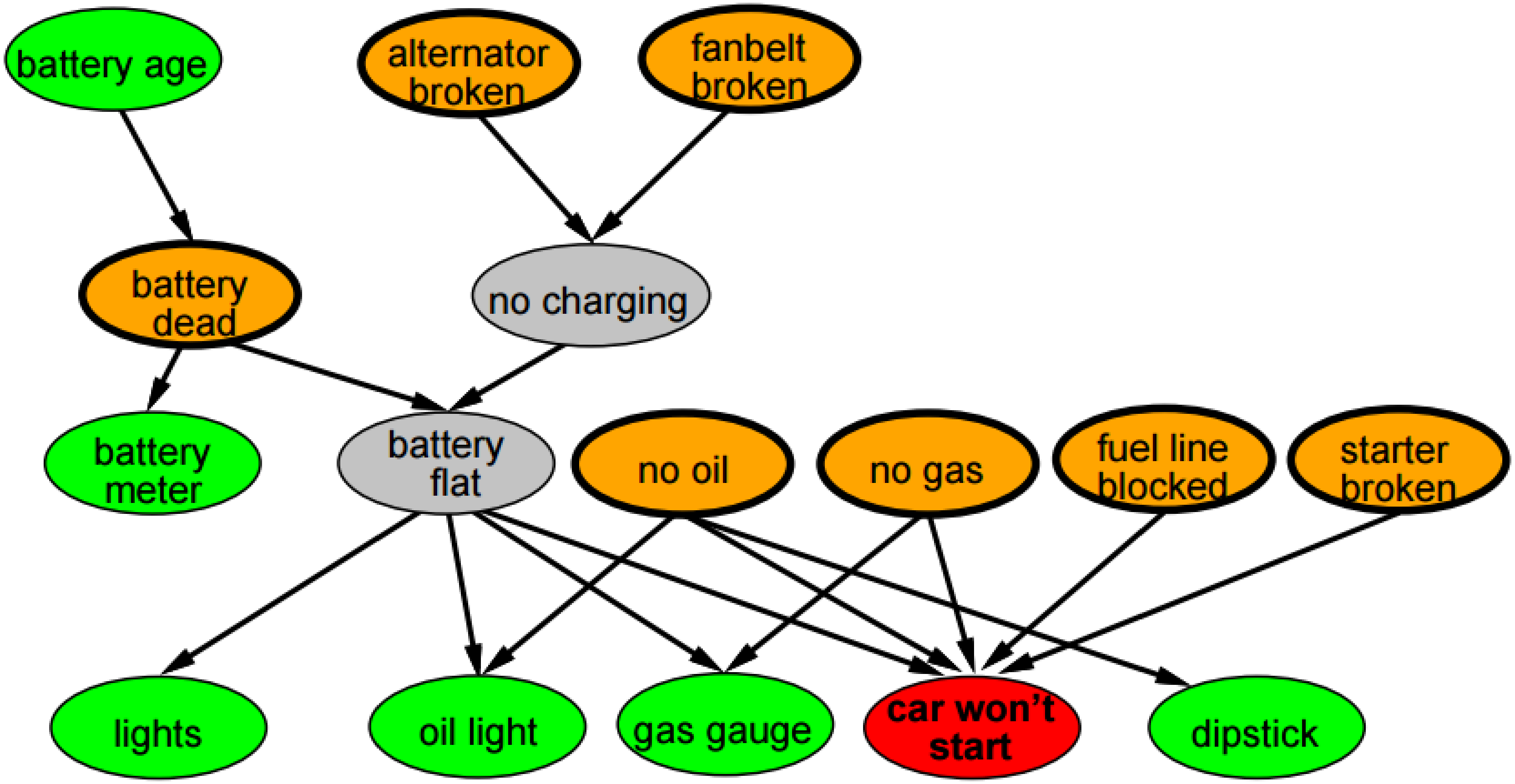
Markov Blanket

- What is the Markov Blanket of “no charging”?



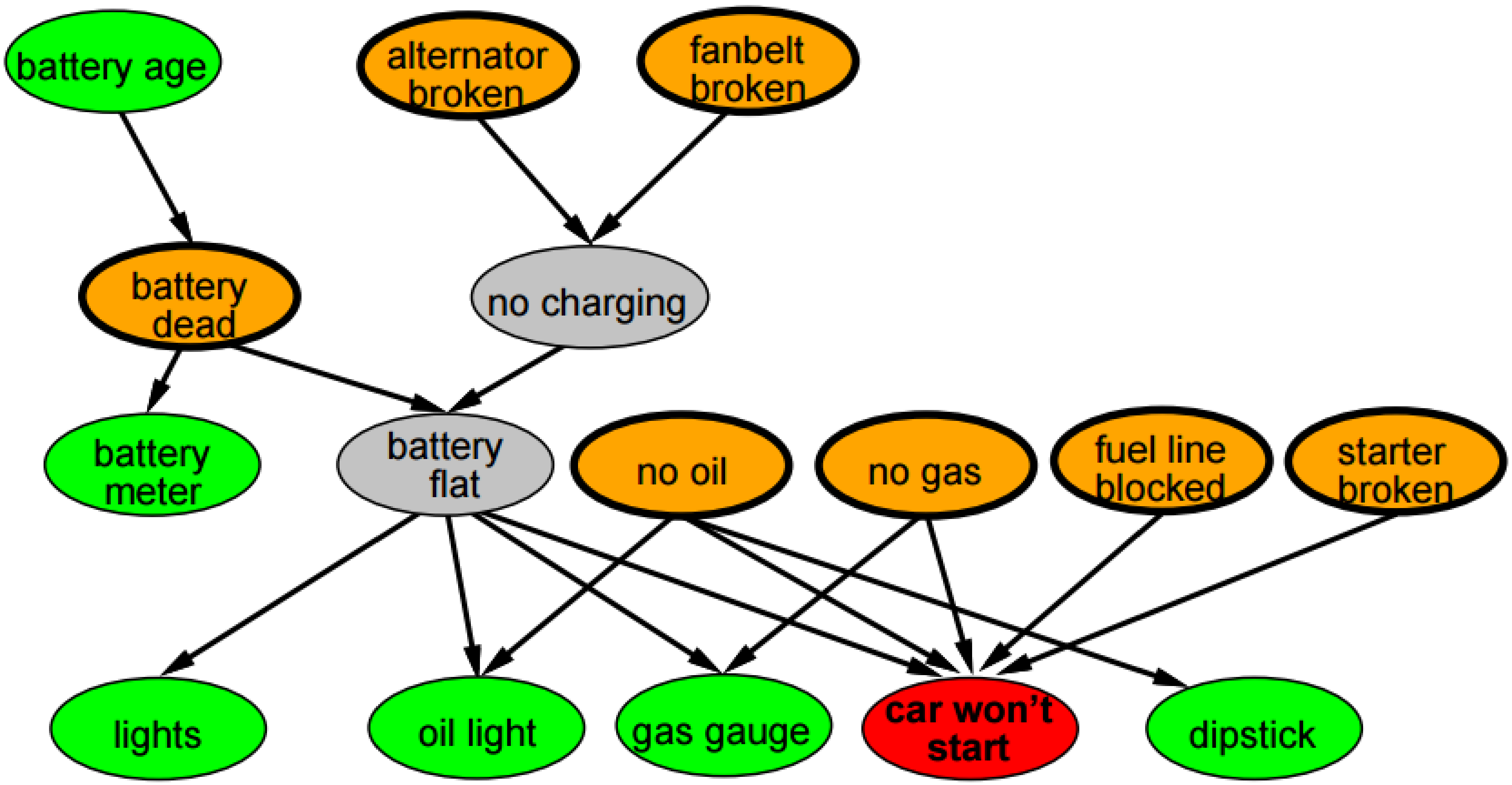
Markov Blanket

- What is the Markov Blanket of “no charging”?
- Its parents: “alternator broken”, “fanbelt broken”.
- Its children: “battery flat”.
- Its children’s parents: “battery dead”.



Markov Blanket

- Therefore, if we know the values of "alternator broken", "fanbelt broken", "battery flat", "battery dead", no other value gives us any more information about the "no charging" variable.



Bayesian Networks - Recap

- Bayesian networks are useful for:
  - Representing joint distributions with much fewer numbers than using a joint distribution table.
  - Doing inference faster than using enumeration (though enumeration is the only inference method that we cover in this course).
- Inference by enumeration (which takes exponential time) is done using repeated applications of the joint distribution equation:

$$\boxed{P(X_1, \ldots, X_i) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))}$$

- The Markov Blanket provides useful intuition about how different nodes affect each other.
  - If A causes B, then A being true makes B more likely.
  - If A causes B, then B being true makes A more likely.
  - If A causes B, and B is true, then competing causes of B make A less likely.