CS 112: Computer System Modeling Fundamentals

Prof. Jenn Wortman Vaughan

May 19, 2011

Lecture 15

Reminders & Announcements

Alternative grading scheme…

Reminders & Announcements

Alternative grading scheme…

- If it works out in your favor, we will count the final exam for 45% of your grade and the midterm for 15%

Joint PMFs

- Joint probability mass functions are useful because they let us reason about relationships between random variables

Joint PMFs

- Joint probability mass functions are useful because they let us reason about relationships between random variables

  Let X indicate whether or not someone has a disease

  Let Y indicate whether or not his test results are positive

Joint PMFs

- Joint probability mass functions are useful because they let us reason about relationships between random variables

Let X indicate whether or not someone has a disease

Let Y indicate whether or not his test results are positive

If we know the joint PMF, we can answer questions like:

- How likely is it that the patient has the disease?
- Given that the patient's test result is positive, how likely is it that he has the disease?

The Problem with Joint PMFs

The real world is much more complicated than this

The Problem with Joint PMFs

The real world is much more complicated than this

- Which of two diseases is most likely given that the patient has stomach pains, skin irritation, loss of appetite, negative results on test 1, positive results on tests 2 and 3, and has not yet been given test 4?

The Problem with Joint PMFs

The real world is much more complicated than this

- Which of two diseases is most likely given that the patient has stomach pains, skin irritation, loss of appetite, negative results on test 1, positive results on tests 2 and 3, and has not yet been given test 4?

  n binary variables $\rightarrow 2^n$ probabilities we need to learn

The Problem with Joint PMFs

The real world is much more complicated than this

- Which of two diseases is most likely given that the patient has stomach pains, skin irritation, loss of appetite, negative results on test 1, positive results on tests 2 and 3, and has not yet been given test 4?

n binary variables $\rightarrow$ 2^n probabilities we need to learn

As computer scientists, we are interested in finding *efficient* ways to reason about complex scenarios

Efficient Representations

- We've already seen one example of an assumption that led to a more efficient representation…

Efficient Representations

- We've already seen one example of an assumption that led to a more efficient representation… the Naive Bayes independence assumption

Efficient Representations

- We've already seen one example of an assumption that led to a more efficient representation… the Naive Bayes independence assumption

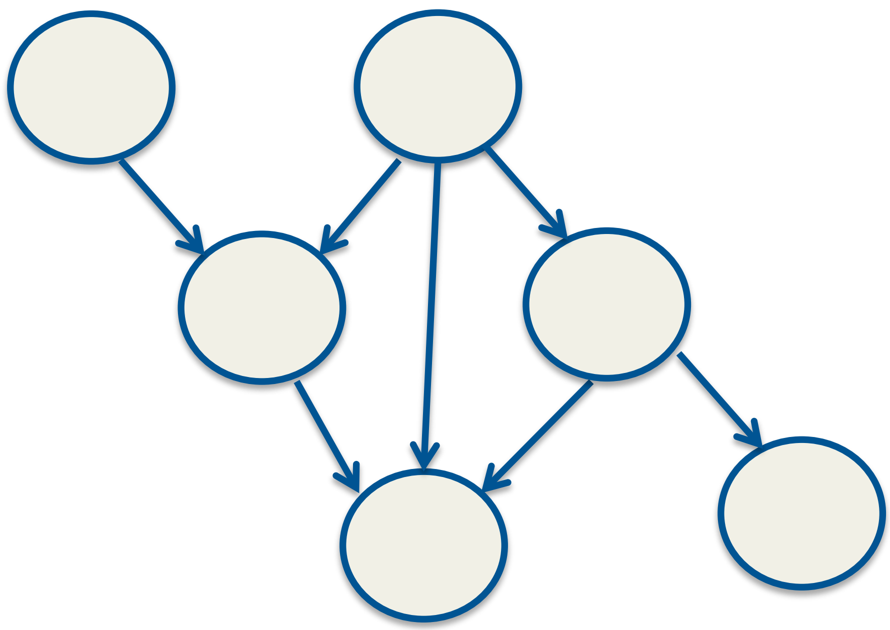
- Graphical models such as Bayesian networks allow us to take advantage of more complex independence relationships between variables

Bayesian Network

- The first component of a Bayesian network is a directed acyclic graph (or DAG)

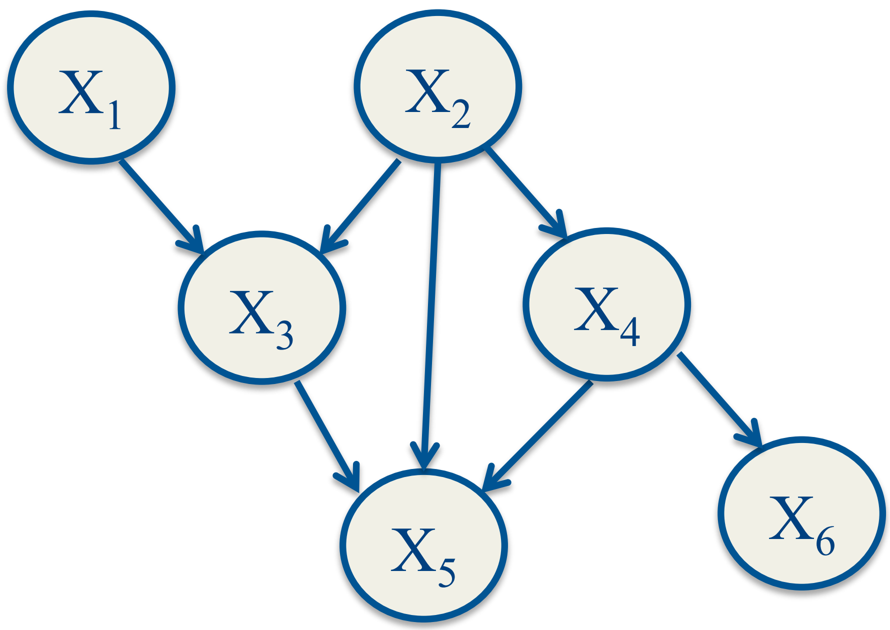
Bayesian Network

- The first component of a Bayesian network is a directed acyclic graph (or DAG)



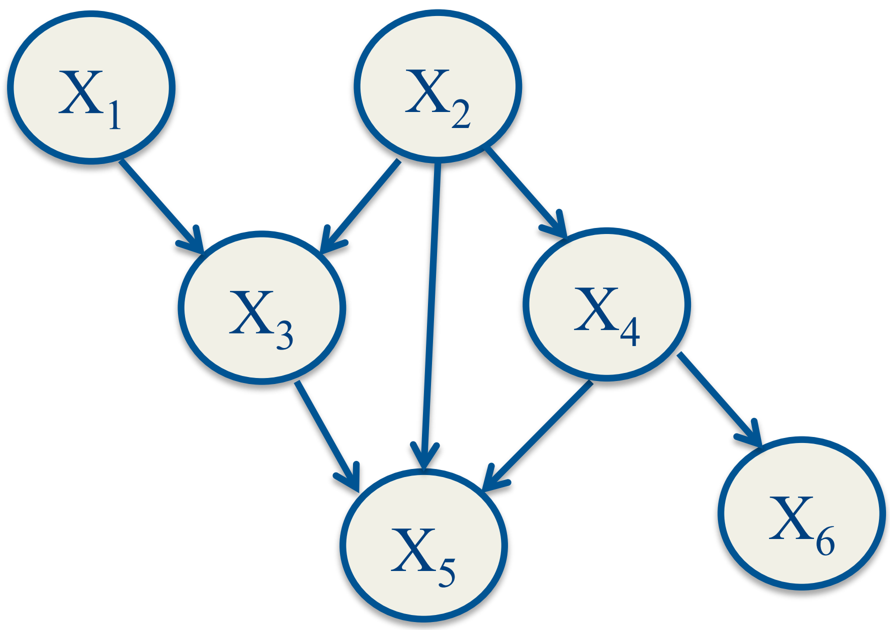
Bayesian Network

- The first component of a Bayesian network is a directed acyclic graph (or DAG)
  - Nodes represents a random variables



Bayesian Network

- The first component of a Bayesian network is a directed acyclic graph (or DAG)
  - Nodes represents a random variables
  - Edges represent dependencies



Bayesian Network

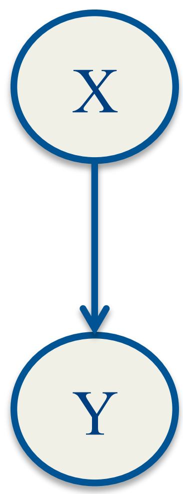
For now, you can think of edges as representing causal effects…

Bayesian Network

For now, you can think of edges as representing causal effects…

Let X indicate whether or not someone has a disease

Let Y indicate whether or not his test results are positive



Bayesian Network

Key Advantage: If dependencies are represented accurately, the joint distribution factors in a nice way

Bayesian Network

Key Advantage: If dependencies are represented accurately, the joint distribution factors in a nice way

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P\left(X_i = x_i \mid Pa(X_i) = pa(X_i)\right)$$

Bayesian Network

Key Advantage: If dependencies are represented accurately, the joint distribution factors in a nice way

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \boxed{P\left(X_i = x_i \mid Pa(X_i) = pa(X_i)\right)}$$

Don't need to learn or store the whole joint PMF!

Can instead learn a small conditional probability table for each random variable.

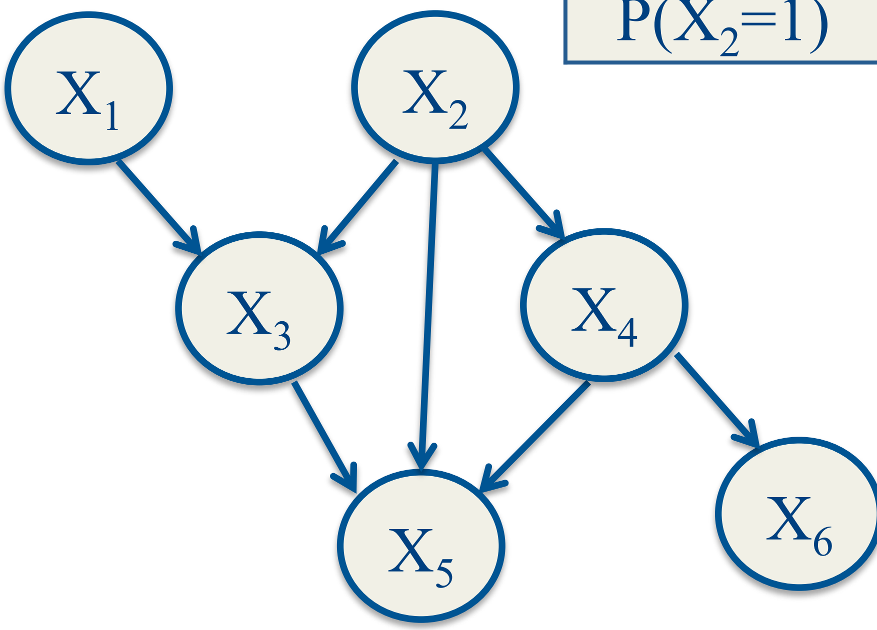
Bayesian Network

- The second component of a Bayesian network is a conditional probability table for each node

Bayesian Network

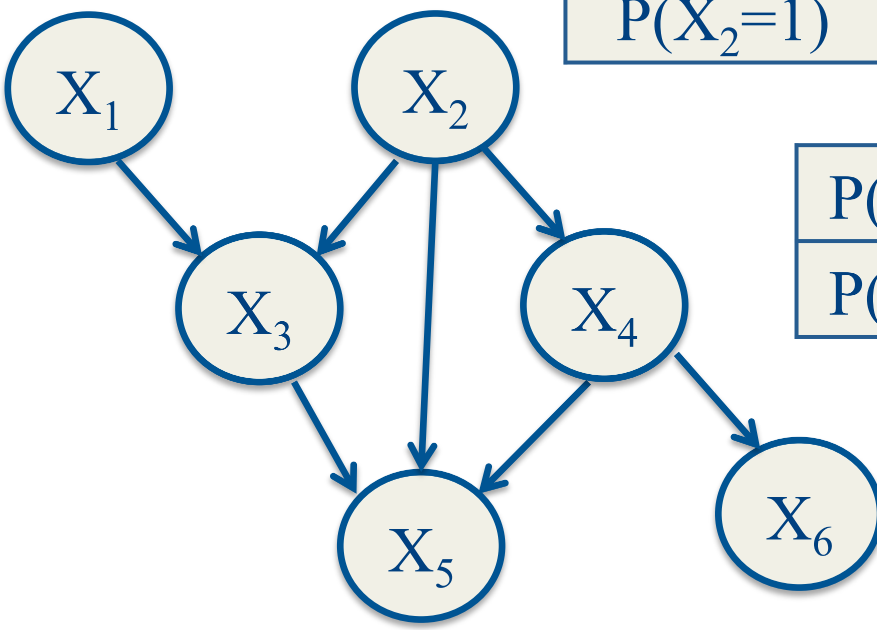
- The second component of a Bayesian network is a conditional probability table for each node

$P(X_2=1)$	0.7



Bayesian Network

- The second component of a Bayesian network is a conditional probability table for each node



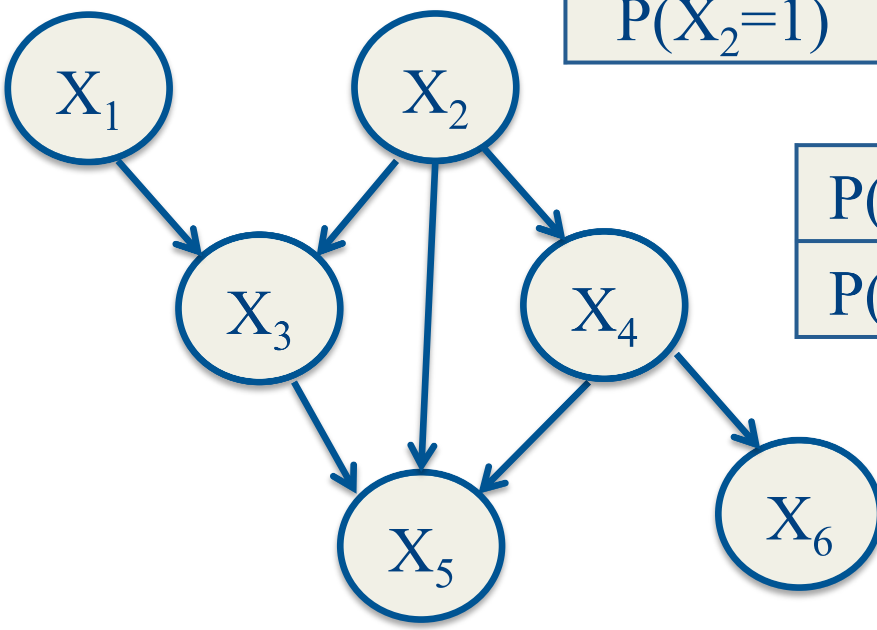
$P(X_2=1)$	0.7

$P(X_4=1 \mid X_2=1)$	0.6
$P(X_4=1 \mid X_2=0)$	0.3

Bayesian Network

- The second component of a Bayesian network is a conditional probability table for each node

$P(X_2=1)$	0.7



$P(X_4=1 \mid X_2=1)$	0.6
$P(X_4=1 \mid X_2=0)$	0.3

- The joint distribution is uniquely specified by these tables

Causal Chains

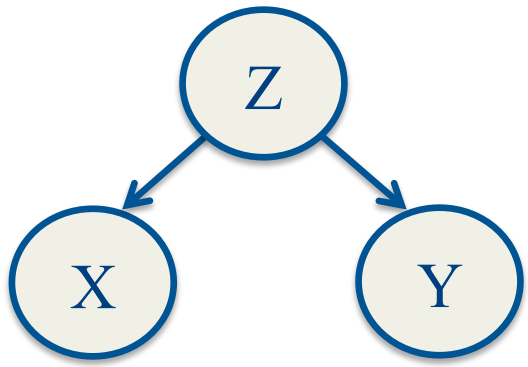
Causal Chains

- X and Y are conditionally independent given Z

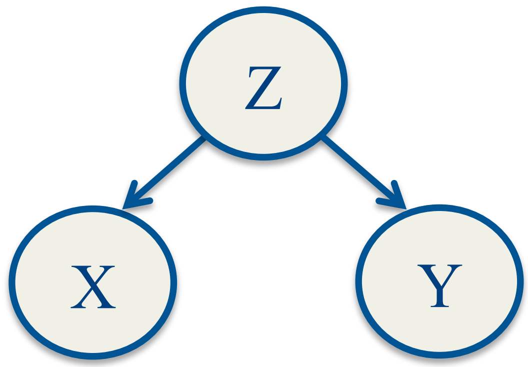
Causal Chains

- X and Y are conditionally independent given Z
- In general, X and Y are *not* independent

Common Cause

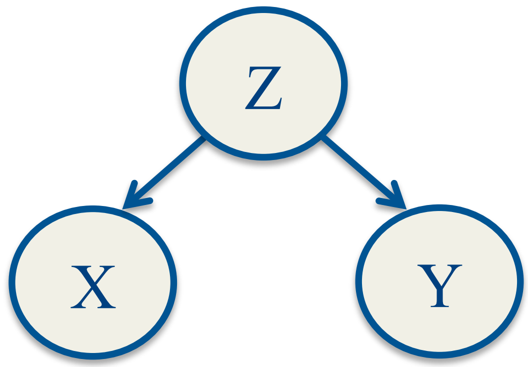


Common Cause



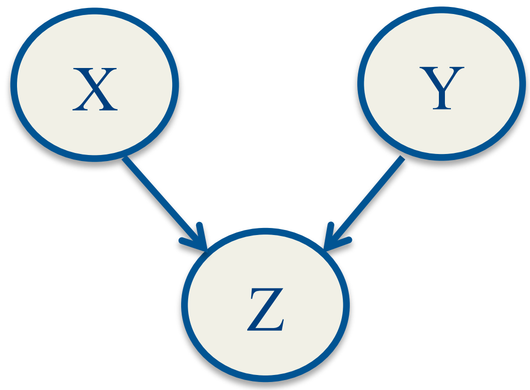
- Again, X and Y are conditionally independent given Z

Common Cause

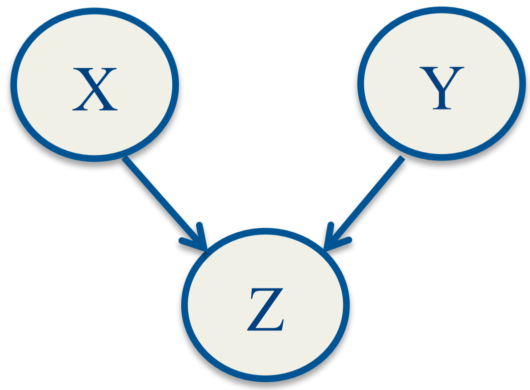


- Again, X and Y are conditionally independent given Z
- In general, X and Y are *not* independent

Common Effect

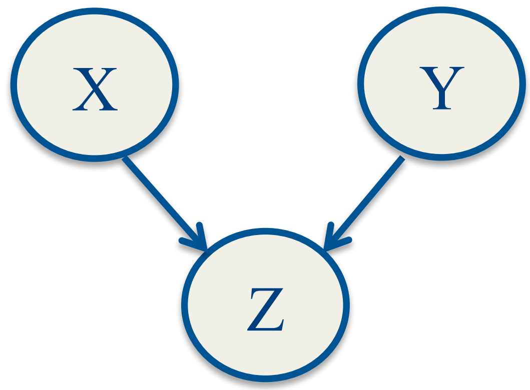


Common Effect



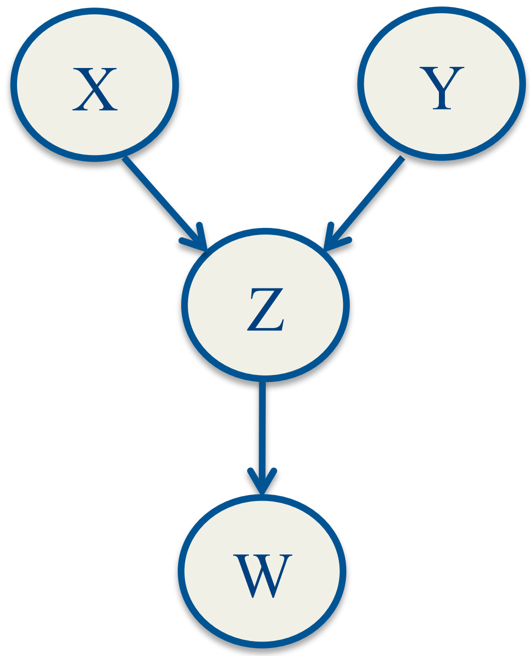
- X and Y are independent

Common Effect



- X and Y are independent
- X and Y are *not* conditionally independent given Z

Common Effect



- X and Y are independent
- X and Y are *not* conditionally independent given Z
- X and Y are *not* conditionally independent given W

Independence Relationships

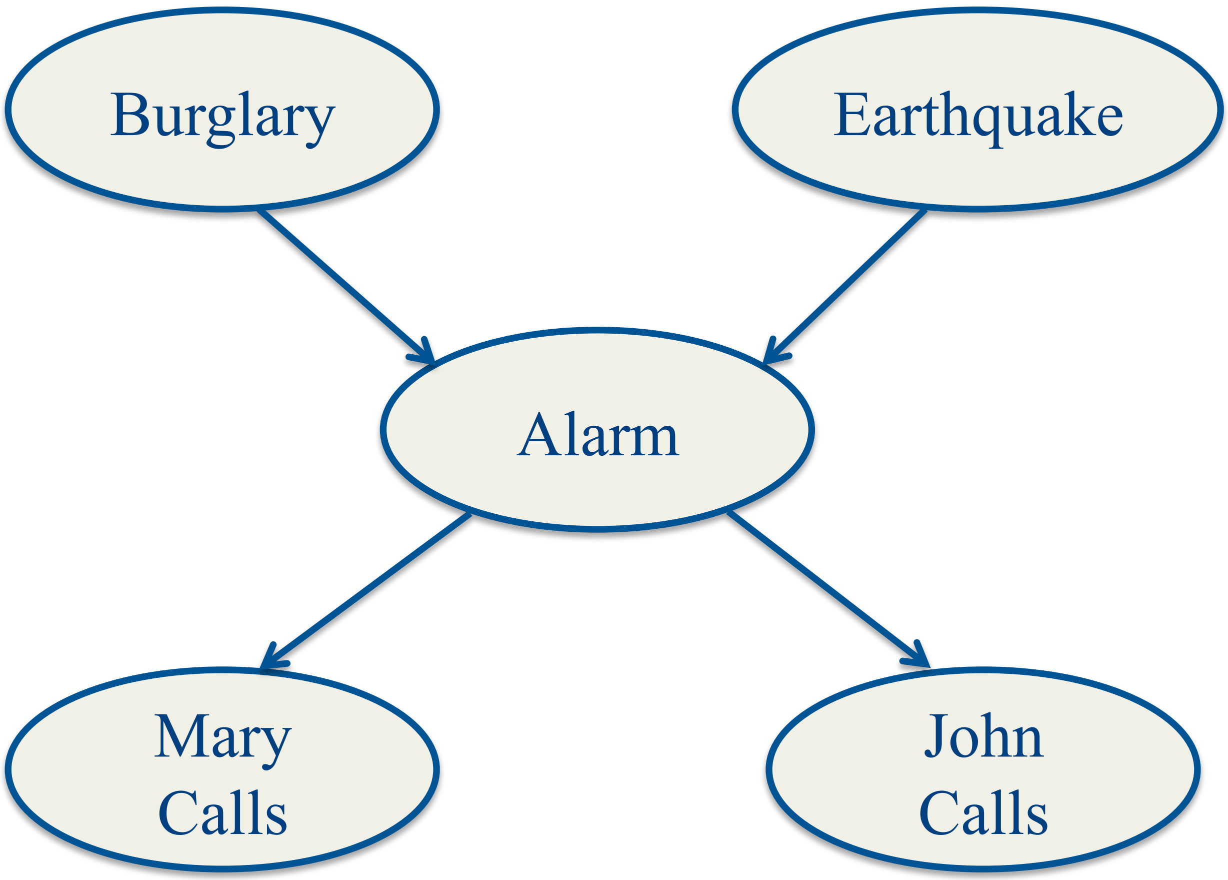
- To determine if two nodes are independent, we just have to check the path (or paths) between them
- If every path is “blocked”, they are independent

The Alarm Network

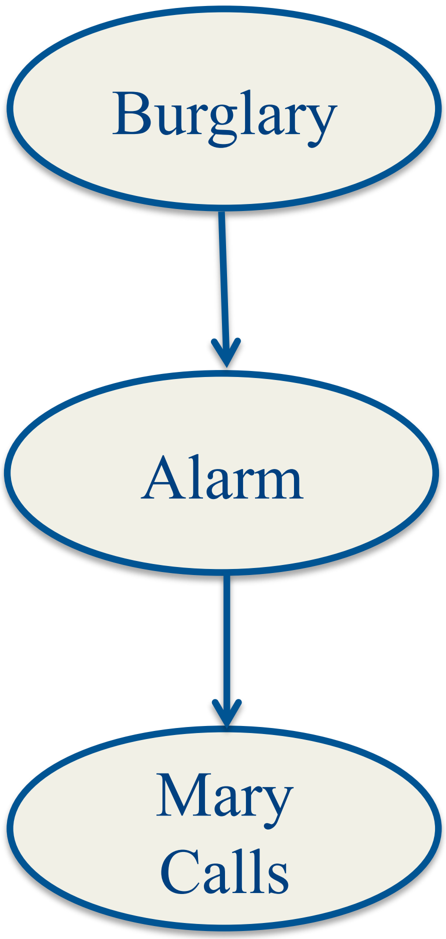
Suppose I have a new burglar alarm installed. It's pretty reliable at detecting burglaries, but sometimes responds to minor earthquakes. I have 2 neighbors (John, Mary) who promise to always call when they hear the alarm. John always calls when he hears the alarm but sometimes confuses his telephone with the alarm and calls then too. Mary likes loud music and sometimes misses the alarm.

Given some knowledge about some of these events, we'd like to reason about some others.

The Alarm Network

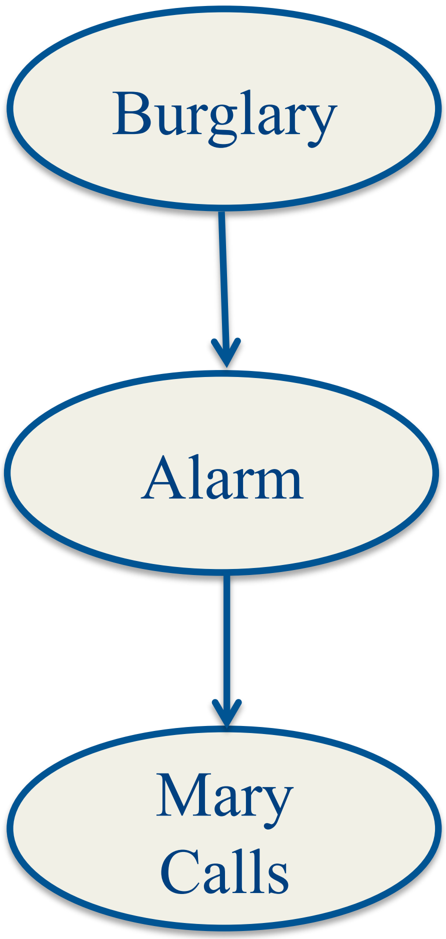


The Alarm Network



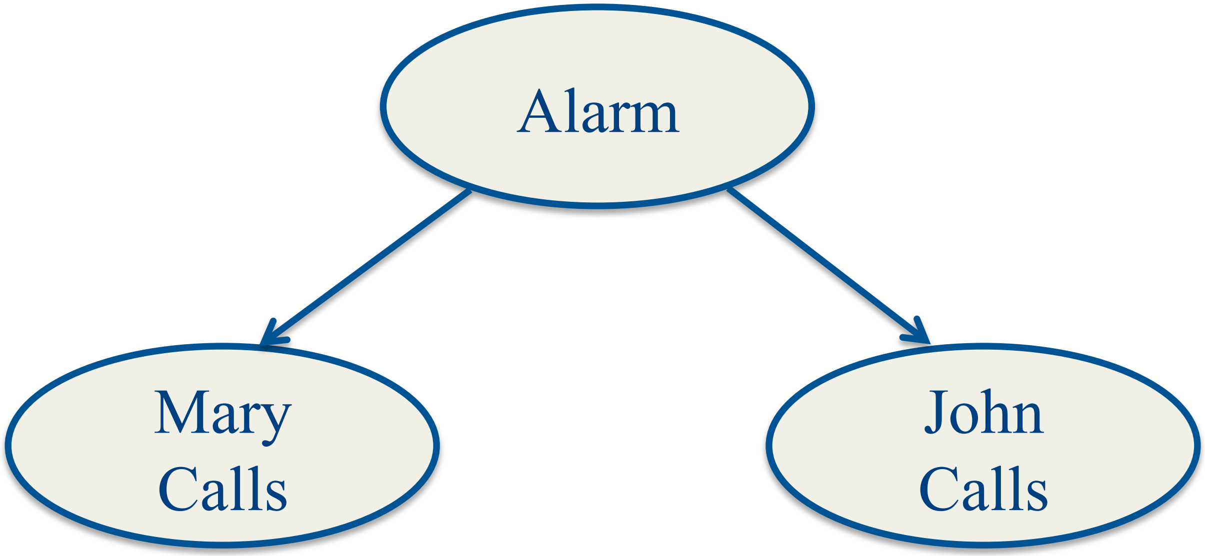
- Is a burglary independent of Mary calling?

The Alarm Network



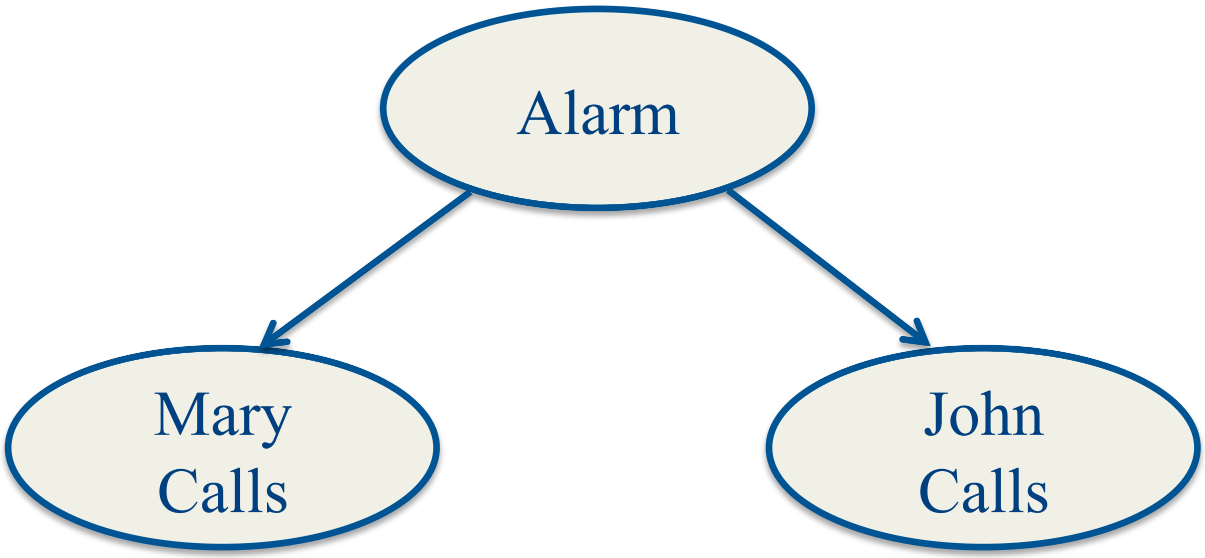
- Is a burglary independent of Mary calling?
- What if we know that the alarm went off?

The Alarm Network



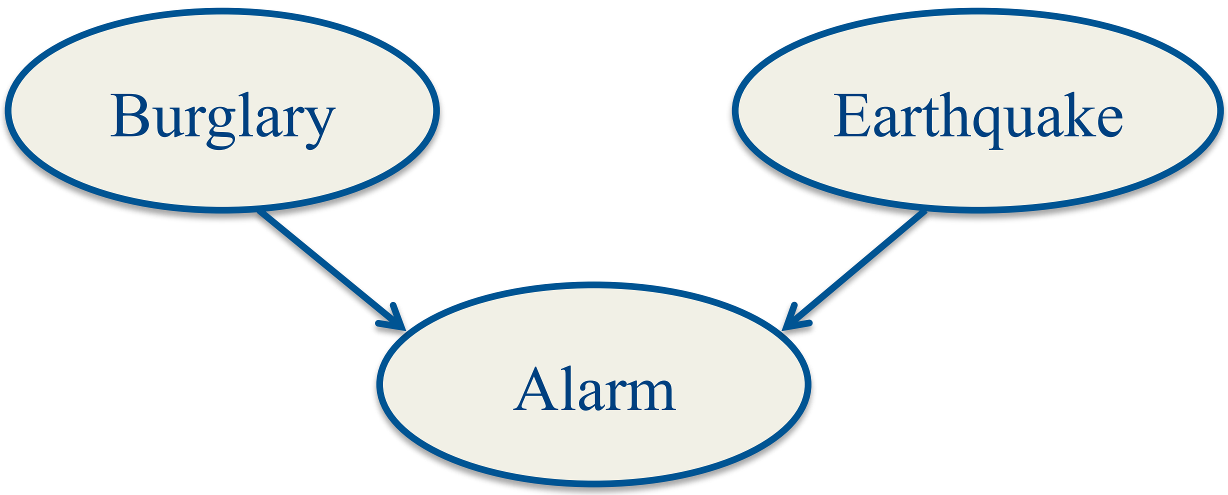
- Is Mary calling independent of John calling?

The Alarm Network



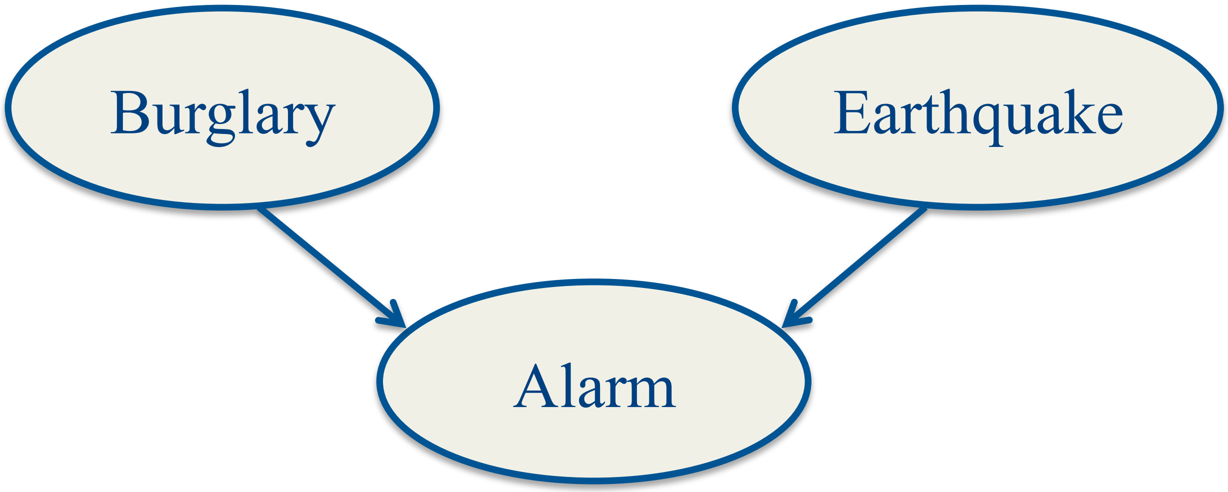
- Is Mary calling independent of John calling?
- What if we know that the alarm went off?

The Alarm Network



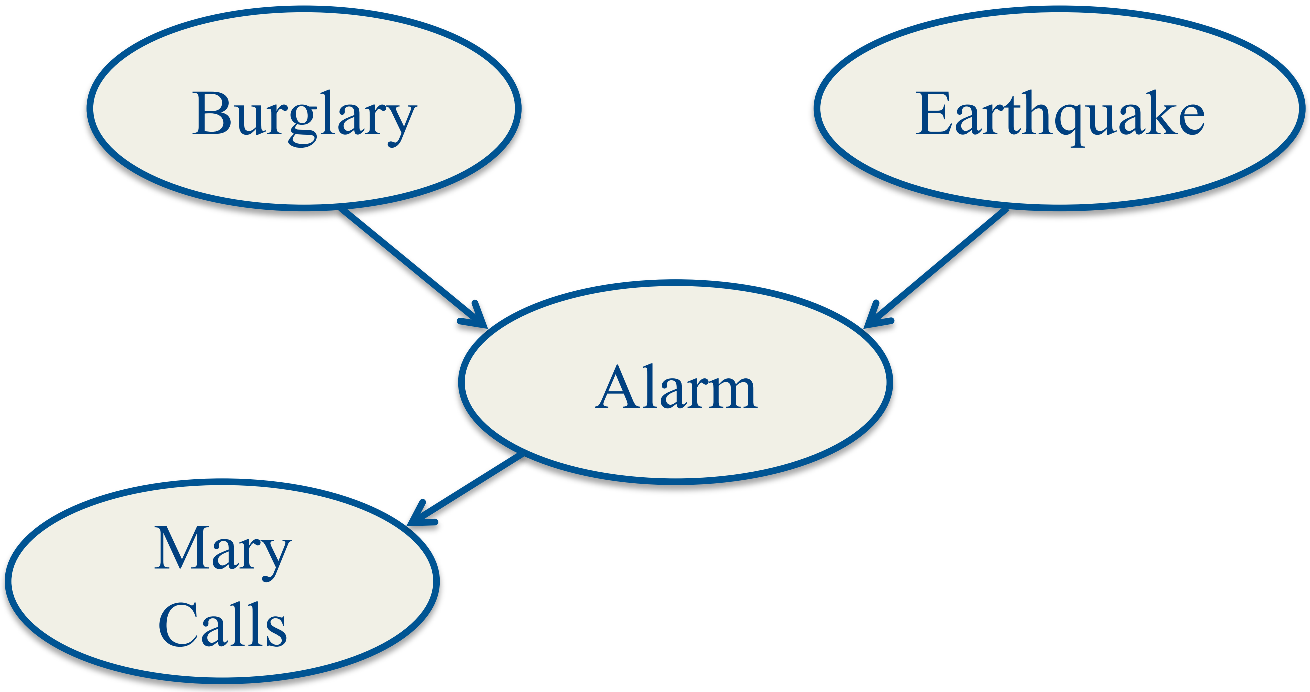
- Is a burglary independent of an earthquake?

The Alarm Network



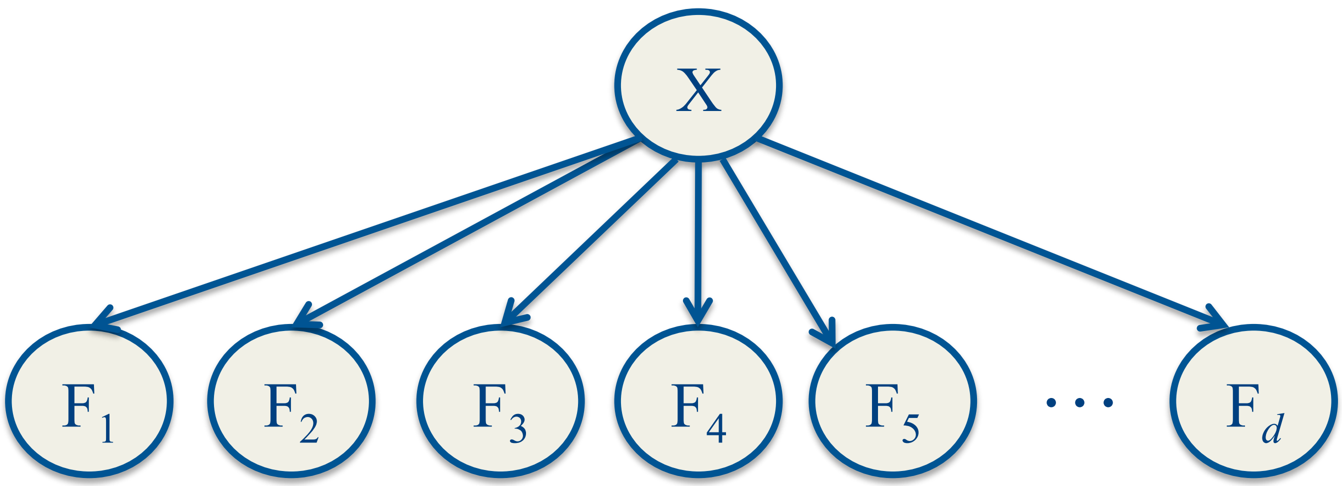
- Is a burglary independent of an earthquake?
- What if we know that the alarm went off?

The Alarm Network

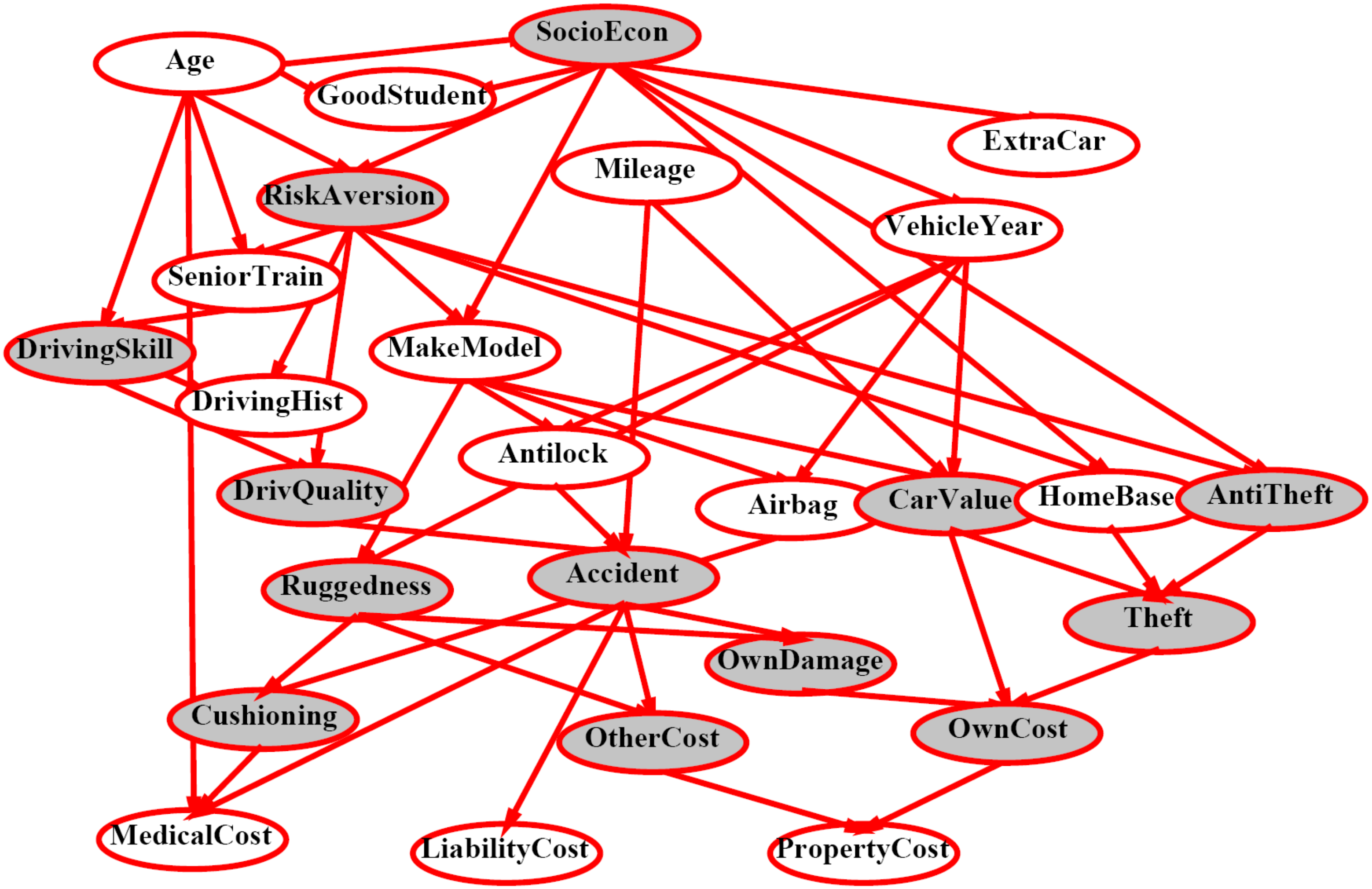


- Is a burglary independent of an earthquake?
- What if we know that the alarm went off?
- What if we know that Mary called?

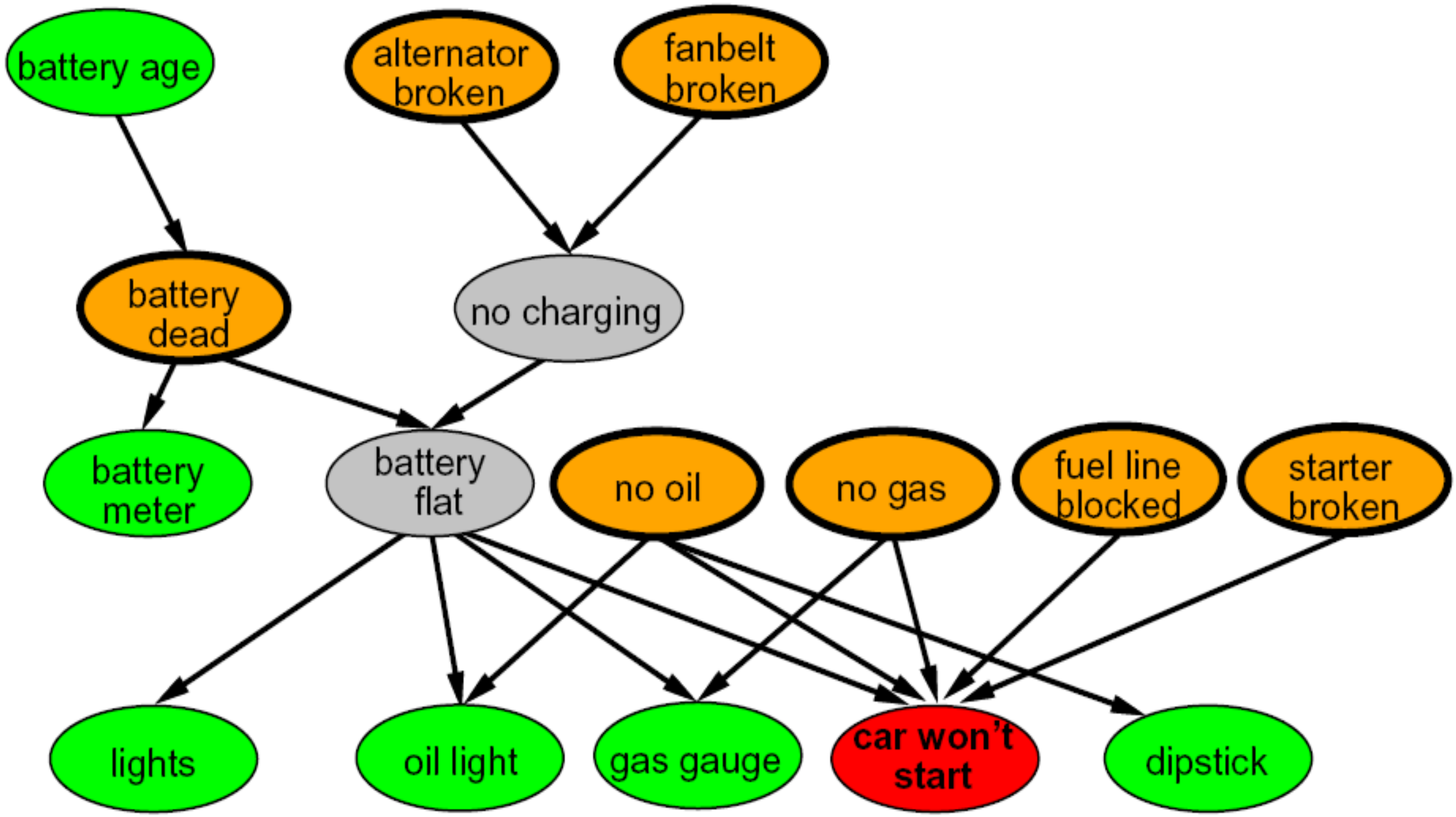
Naive Bayes Network



Insurance Network



Car Network


battery age
alternator broken
fanbelt broken
battery dead
no charging
battery meter
battery flat
no oil
no gas
fuel line blocked
starter broken
lights
oil light
gas gauge
car won't start
dipstick

Other Common Applications

- Medical diagnostics
- Bioinformatics
- Computer vision (e.g., object recognition)
- Text analysis (e.g., discovery of common topics in text)
- Time series analysis (e.g., analyzing stock market data)

- Any application in which there are lots of random variables and lots of structure