Bayesian Networks

Sanja Lazarova-Molnar

Introduction

Suppose you are trying to determine if a patient has inhalational anthrax. You observe the following symptoms:

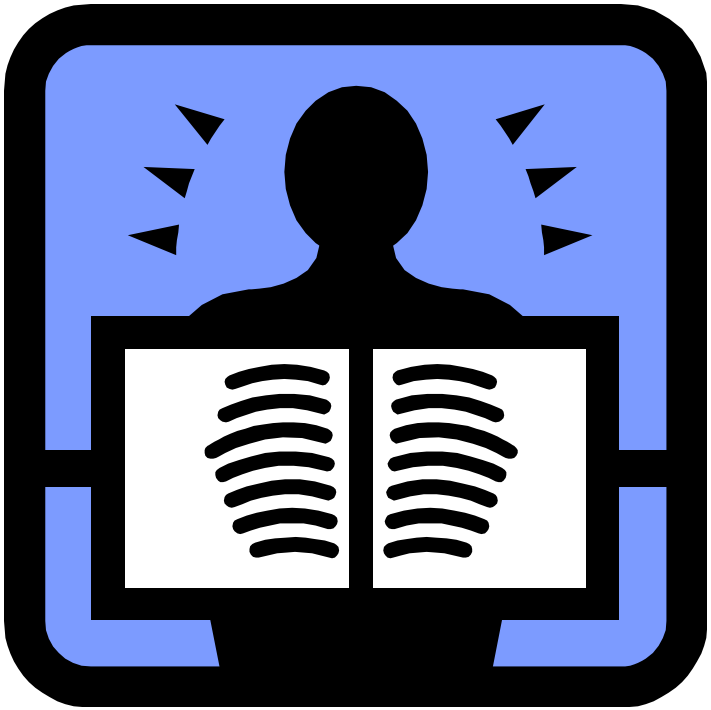
- The patient has a cough
- The patient has a fever
- The patient has difficulty breathing

Introduction

You would like to determine how likely the patient is infected with inhalational anthrax given that the patient has a cough, a fever, and difficulty breathing

We are not 100% certain that the patient has anthrax because of these symptoms. We are dealing with uncertainty!

Introduction



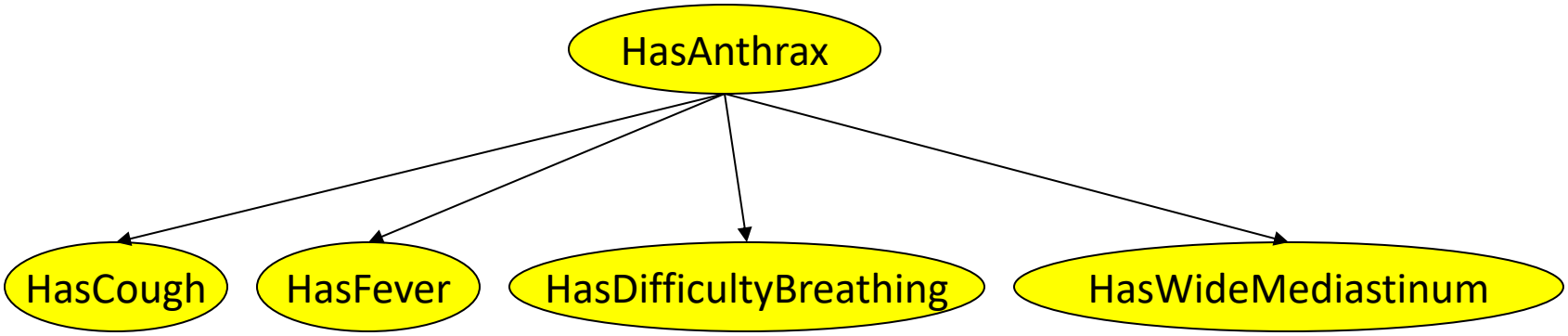
Now suppose you order an x-ray and observe that the patient has a wide mediastinum.

Your belief that that the patient is infected with inhalational anthrax is now much higher.

Introduction

- Observations affected your belief that the patient is infected with anthrax
- Reasoning with uncertainty
- Need methodology for reasoning with uncertainty...

Bayesian Networks



- In the opinion of many AI researchers, Bayesian networks are one of the most significant contribution in AI in the last 20-30 years
- They are used in many applications e.g. spam filtering, speech recognition, robotics, diagnostic systems and even syndromic surveillance

Probability Review: Random Variables

- A random variable is the basic element of probability
- Refers to an event and there is some degree of uncertainty as to the outcome of the event
- For example, the random variable *A* could be the event of getting a heads on a coin flip

Boolean Random Variables

- Simplest type of random variables – Boolean
- Take the values *true* or *false*
- Think of the event as occurring or not occurring
- Examples (Let *A* be a Boolean random variable):

  A = Getting heads on a coin flip

  A = It will rain today

  A = Denmark wins the World Championship in 2022

The Problem with the Joint Distribution

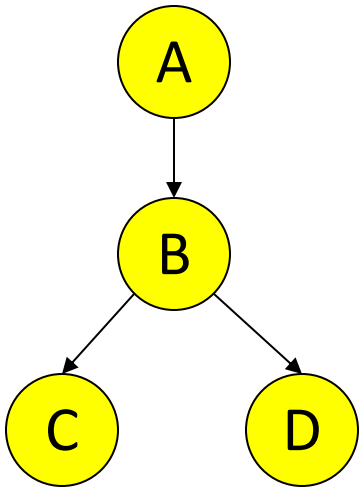
- Lots of entries in the table to fill up!
- For *k* Boolean random variables, you need a table of size 2^k
- How do we use fewer numbers?

A	B	C	P(A,B,C)
false	false	false	0.1
false	false	true	0.2
false	true	false	0.05
false	true	true	0.05
true	false	false	0.3
true	false	true	0.1
true	true	false	0.05
true	true	true	0.15

A Bayesian Network

A Bayesian network is made up of:

1. A Directed Acyclic Graph



2. A set of tables for each node in the graph

A	P(A)
false	0.6
true	0.4

A	B	P(B|A)
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

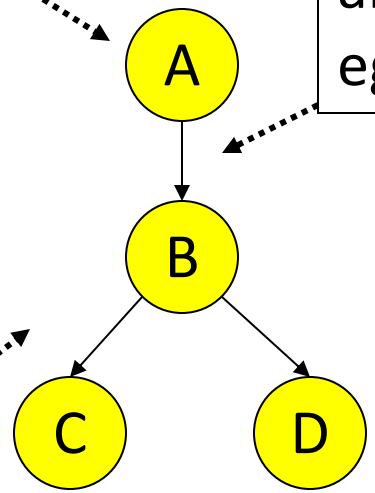
B	D	P(D|B)
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

A Directed Acyclic Graph

Each node in the graph is a random variable

A node *X* is a parent of another node *Y* if there is an arrow from node *X* to node *Y* eg. *A* is a parent of *B*



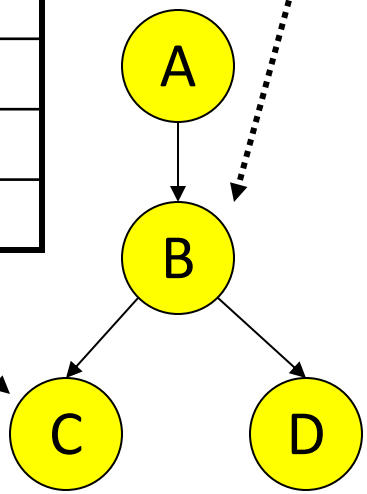
Informally, an arrow from node *X* to node *Y* means *X* has a direct influence on *Y*

A Set of Tables for Each Node

A	P(A)
false	0.6
true	0.4

A	B	P(B|A)
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1



B	D	P(D|B)
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

Each node X_i has a conditional probability distribution $P(X_i \mid \text{Parents}(X_i))$ that quantifies the effect of the parents on the node

The parameters are the probabilities in these conditional probability tables (CPTs)

A Set of Tables for Each Node

Conditional Probability
Distribution for C given B

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

For a given combination of values of the parents (B in this example), the entries for P(C=true | B) and P(C=false | B) must add up to 1
eg. P(C=true | B=false) + P(C=false | B=false)=1

If you have a Boolean variable with k Boolean parents, this table has 2^{k+1} probabilities

Bayesian Network

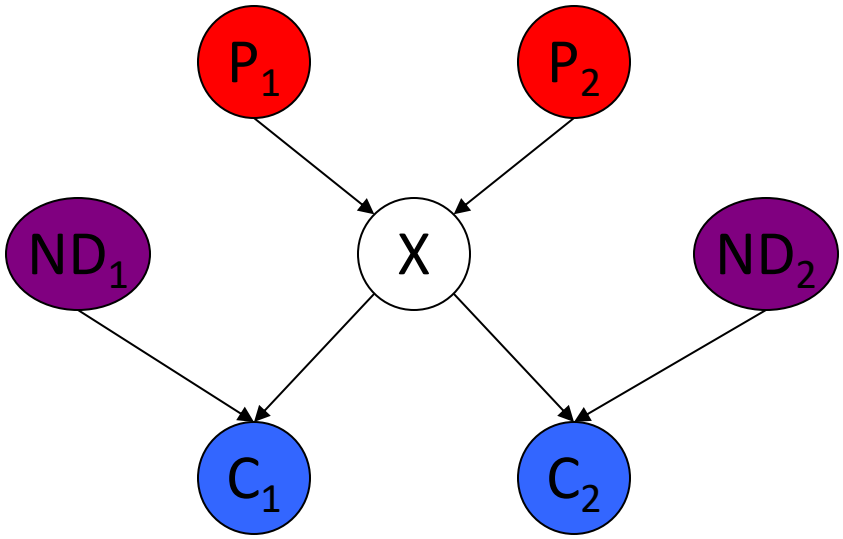
Two important properties:

1. Encodes the conditional independence relationships between the variables in the graph structure
2. Is a compact representation of the joint probability distribution over the variables

Conditional Independence

The Markov condition: given its parents (P_1, P_2),

a node (X) is conditionally independent of its non-descendants (ND_1, ND_2)



Conditional Independence: Example

A is **the height of a child** and **B** is **the number of words that the child knows**.

- When **A** is high, **B** is high too.

A **single piece of information will make A and B completely independent.**

- The child's age.

The height and the # of words known by the kid are **NOT independent**, but they are **conditionally independent** the kid's age is provided.

The Joint Probability Distribution

Due to the Markov condition, we can compute the joint probability distribution over all the variables X_1, …, X_n in the Bayesian net using the formula:

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

Where Parents(X_i) means the values of the Parents of the node X_i with respect to the graph

Using a Bayesian Network Example

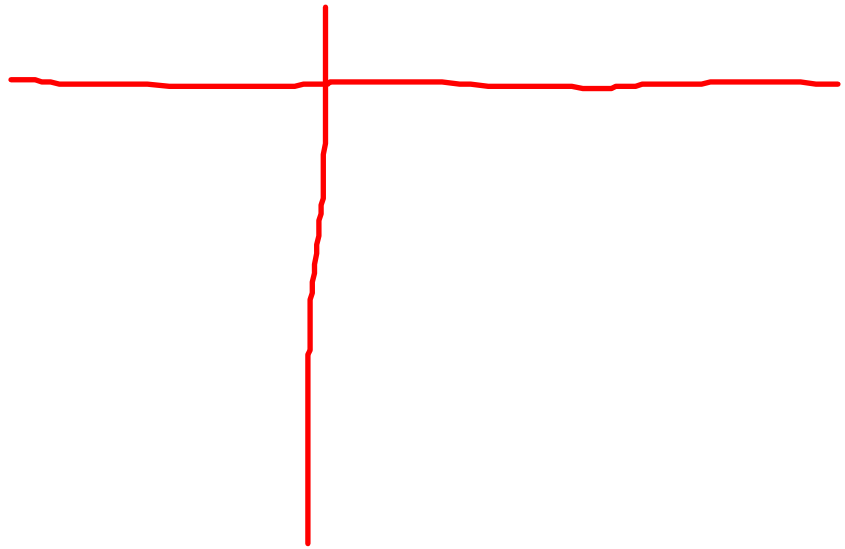
Using the network in the example, suppose you want to calculate:

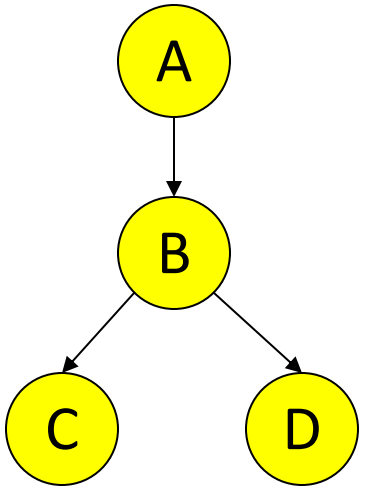
P(A = true, B = true, C = true, D = true)

= P(A = true) * P(B = true | A = true) *

P(C = true | B = true) * P(D = true | B = true)

= (0.4)*(0.3)*(0.1)*(0.95)





Using a Bayesian Network Example

Using the network in the example, suppose you want to calculate:

P(A = true, B = true, C = true, D = true)

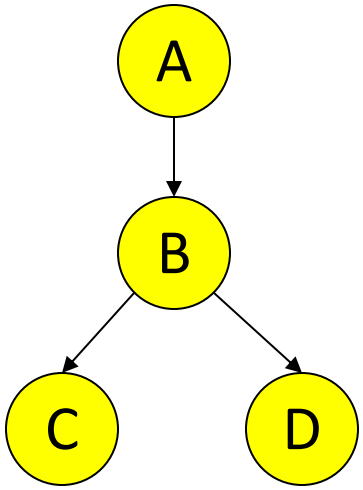
= P(A = true) * P(B = true | A = true) *

P(C = true | B = true) P(D = true | B = true)

= (0.4)*(0.3)*(0.1)*(0.95)

This is from the graph structure

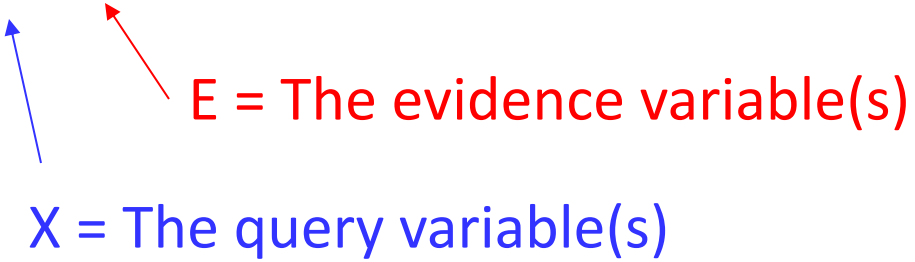
These numbers are from the conditional probability tables



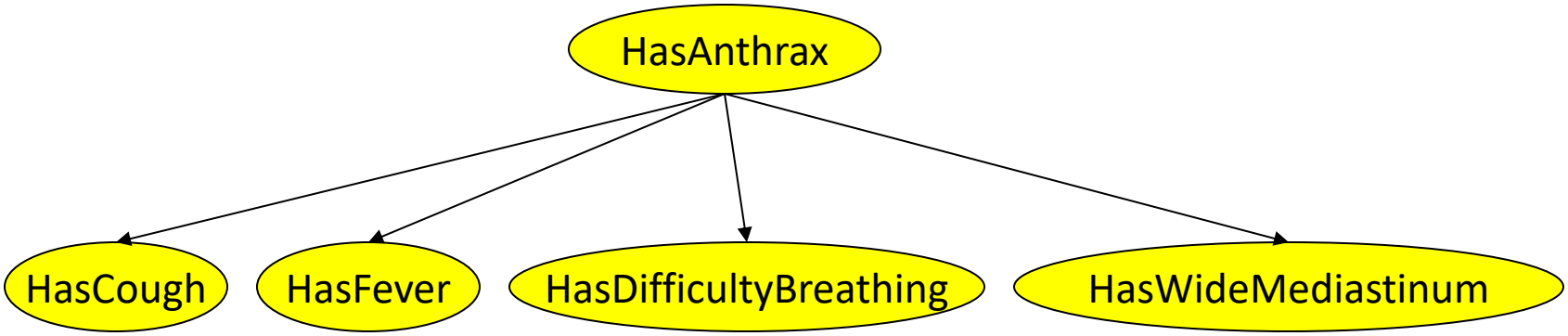
Inference

- Using a Bayesian network to compute probabilities is called inference
- In general, inference involves queries of the form:

  P(X | E)



Inference



- An example of a query would be:

  P(*HasAnthrax = true* | *HasFever = true*, *HasCough = true*)
- Note: Even though *HasDifficultyBreathing* and *HasWideMediastinum* are in the Bayesian network, they are not given values in the query (ie. they do not appear either as query variables or evidence variables)
- They are treated as unobserved variables

Example: Assume five variables

T: The lecture started by 8:45

L: The lecturer arrives late

R: The lecture concerns robots

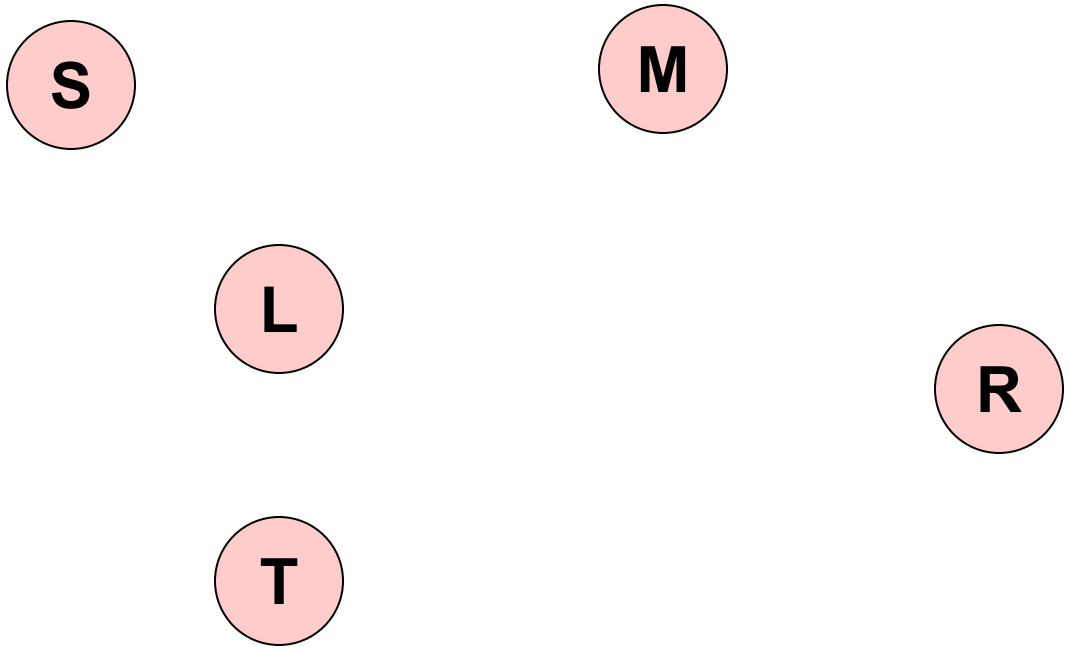
M: The lecturer is Me

S: It is sunny

- T only directly influenced by L (i.e. T is conditionally independent of R,M,S given L)
- L only directly influenced by M and S (i.e. L is conditionally independent of R given M & S)
- R only directly influenced by M (i.e. R is conditionally independent of L,S, given M)
- M and S are independent

Making a Bayes net

T: The lecture started by 8:45
L: The lecturer arrives late
R: The lecture concerns robots
M: The lecturer is Me
S: It is sunny

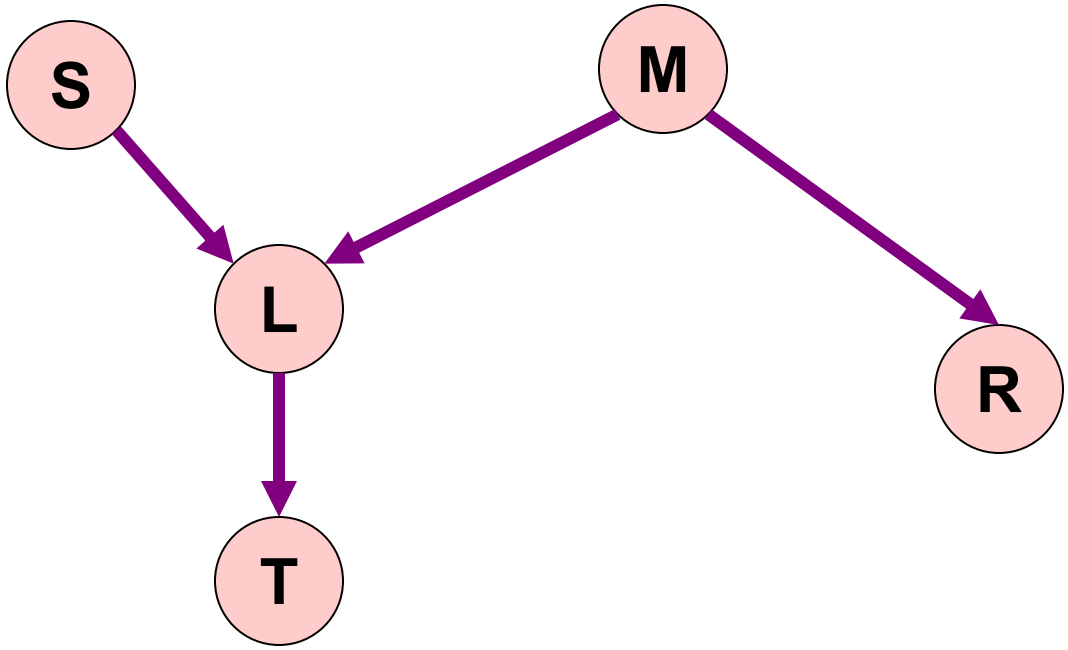


Step One: add variables.

- Just choose the variables you'd like to be included in the net.

Making a Bayes net

T: The lecture started by 8:45
L: The lecturer arrives late
R: The lecture concerns robots
M: The lecturer is Me
S: It is sunny

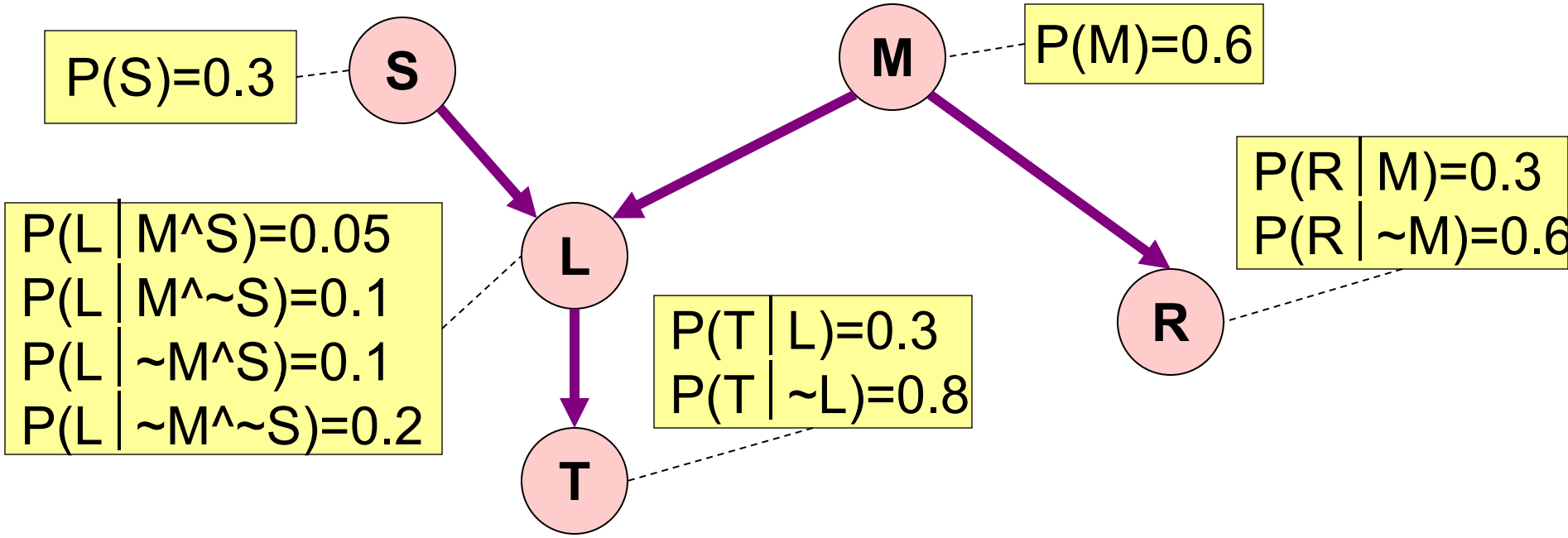


Step Two: add links.

- The link structure must be acyclic.
- If node X is given parents $Q_1,Q_2,..Q_n$ you are promising that any variable that's a non-descendent of X is conditionally independent of X given $\{Q_1,Q_2,..Q_n\}$

Making a Bayes net

T: The lecture started by 8:45
L: The lecturer arrives late
R: The lecture concerns robots
M: The lecturer is Me
S: It is sunny

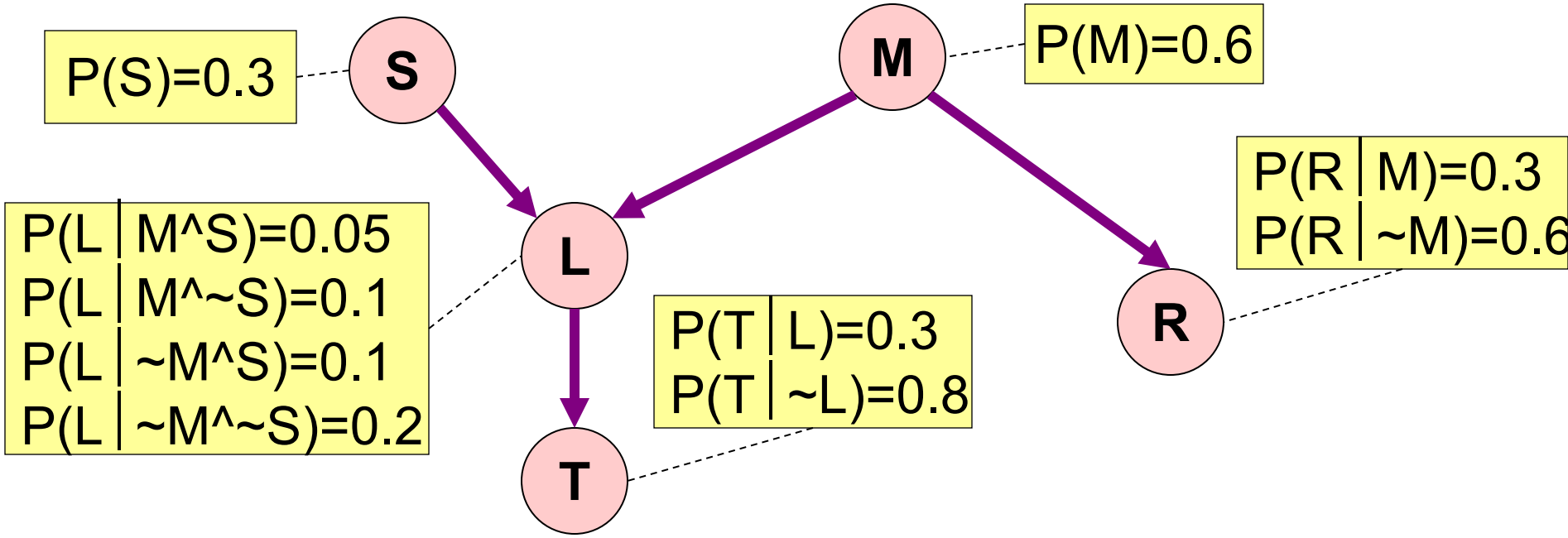


Step Three: add a probability table for each node.

- The table for node X must list P(X|Parent Values) for each possible combination of parent values

Making a Bayes net

T: The lecture started by 8:45
L: The lecturer arrives late
R: The lecture concerns robots
M: The lecturer is Me
S: It is sunny



- Two unconnected variables may still be correlated
- Each node is conditionally independent of all non-descendants in the tree, given its parents.
- You can deduce many other conditional independence relations from a Bayes net.

Bayes Nets Formalized

A Bayes net (also called a belief network) is an augmented directed acyclic graph, represented by the pair V , E where:

- V is a set of vertices.
- E is a set of directed edges joining vertices. No loops of any length are allowed.

Each vertex in V contains the following information:

- The name of a random variable
- A probability distribution table indicating how the probability of this variable's values depends on all possible combinations of parental values.

Building a Bayes Net

1. Choose a set of relevant variables and an ordering for them.
2. Assume they're called $X_1, \ldots, X_m$ (where X_1 is the first in the ordering, X_2 is the second, etc.)
3. For *i = 1 to m*:
    1. Add the X_i node to the network
    2. Set *Parents*(X_i) to be a minimal subset of $\{X_1 \ldots X_{i-1}\}$ such that we have conditional independence of X_i and all other members of $\{X_1 \ldots X_{i-1}\}$ given *Parents*(X_i)
    3. Define the probability table of
    $P(X_i = k \mid \text{Assignments of } Parents(X_i))$.

Example Bayes Net Building

Suppose we're building a nuclear power station.

There are the following random variables:

GRL: Gauge reads low

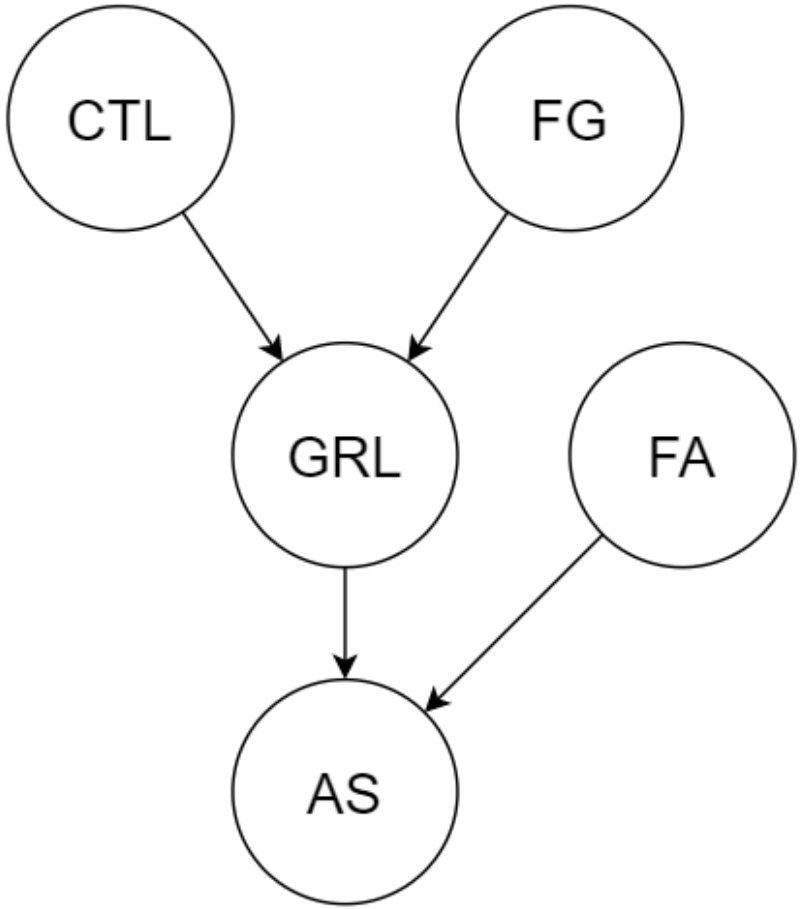
CTL: Core temperature is low

FG: Gauge is faulty

FA: Alarm is faulty

AS: Alarm sound

- If alarm working properly, the alarm is meant to sound if the gauge stops reading a low temp.
- If gauge working properly, the gauge is meant to read the temp of the core.

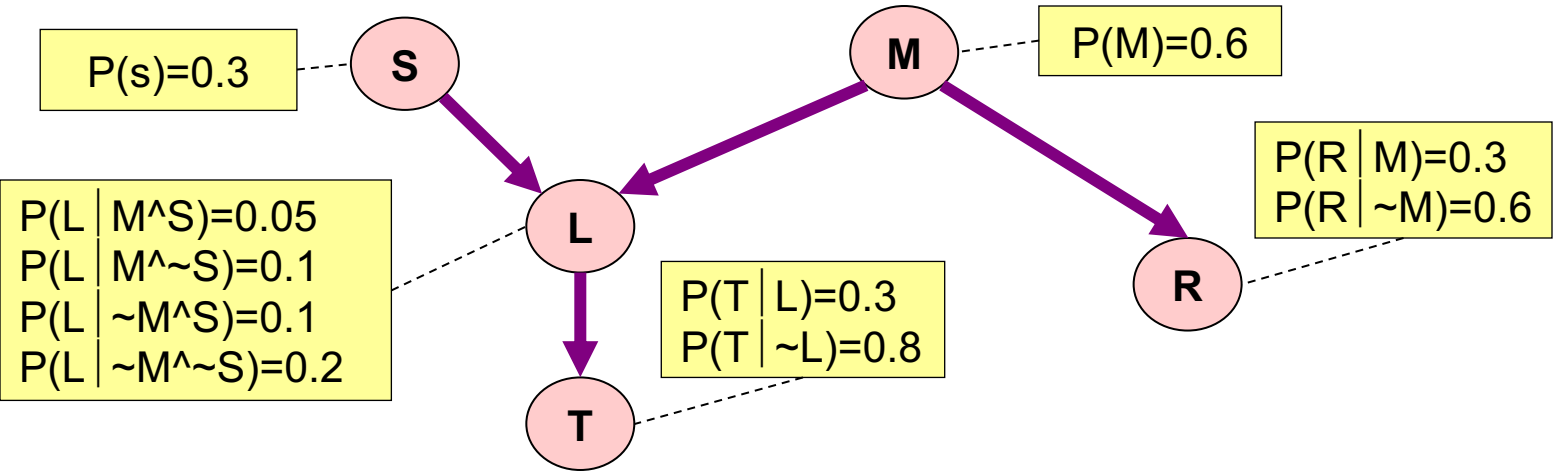


courtesy of Thomas Steenfeldt Laursen

Computing a Joint Entry

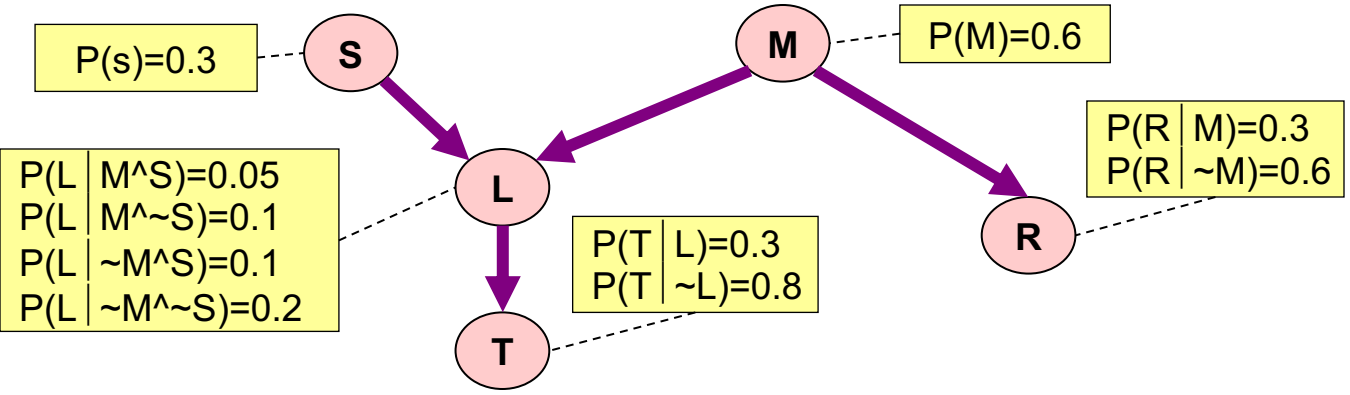
How to compute an entry in a joint distribution?

E.G: What is P(S ^ ~M ^ L ^ ~R ^ T)?



T: The lecture started by 8:45
L: The lecturer arrives late
R: The lecture concerns robots
M: The lecturer is Me
S: It is sunny

Computing with Bayes Net



P(T ^ ~R ^ L ^ ~M ^ S) =
= P(T | ~R ^ L ^ ~M ^ S) * P(~R ^ L ^ ~M ^ S) =
= P(T | L) * P(~R ^ L ^ ~M ^ S) =
= P(T | L) * P(~R | L ^ ~M ^ S) * P(L ^ ~M ^ S) =
= P(T | L) * P(~R | ~M) * P(L ^ ~M ^ S) =
= P(T | L) * P(~R | ~M) * P(L | ~M ^ S) * P(~M ^ S) =
= P(T | L) * P(~R | ~M) * P(L | ~M ^ S) * P(~M | S) * P(S) =
= P(T | L) * P(~R | ~M) * P(L | ~M ^ S) * P(~M) * P(S).

The general case

$P(X_1=x_1 \wedge X_2=x_2 \wedge \ldots X_{n-1}=x_{n-1} \wedge X_n=x_n) =$

$P(X_n=x_n \wedge X_{n-1}=x_{n-1} \wedge \ldots X_2=x_2 \wedge X_1=x_1) =$

$P(X_n=x_n \mid X_{n-1}=x_{n-1} \wedge \ldots X_2=x_2 \wedge X_1=x_1) * P(X_{n-1}=x_{n-1} \wedge \ldots X_2=x_2 \wedge X_1=x_1) =$

$P(X_n=x_n \mid X_{n-1}=x_{n-1} \wedge \ldots X_2=x_2 \wedge X_1=x_1) * P(X_{n-1}=x_{n-1} \mid \ldots X_2=x_2 \wedge X_1=x_1) *$

$P(X_{n-2}=x_{n-2} \wedge \ldots X_2=x_2 \wedge X_1=x_1) =$

:

:

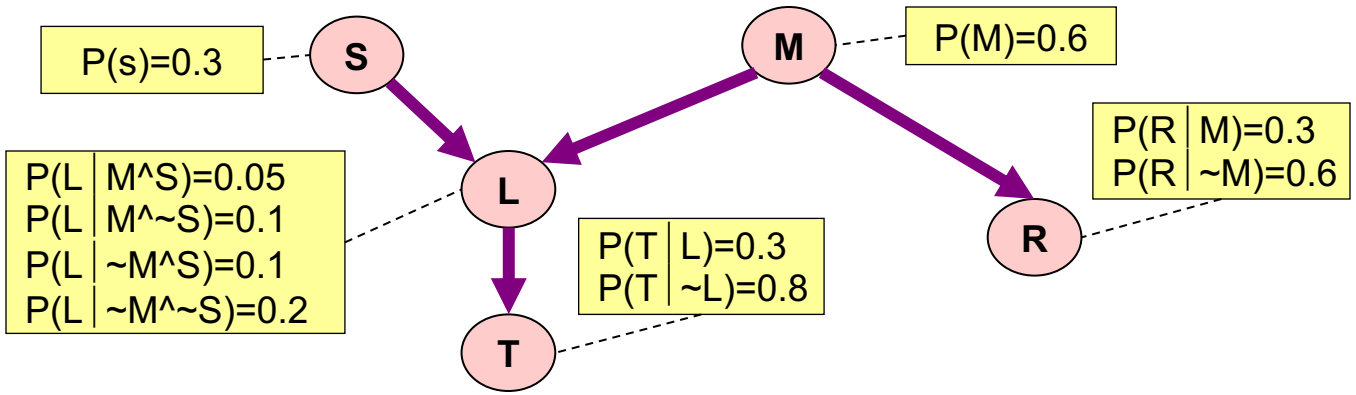
$$= \prod_{i=1}^{n} P\left(\left(X_i = x_i\right) \middle| \left(\left(X_{i-1} = x_{i-1}\right) \wedge \ldots \left(X_1 = x_1\right)\right)\right)$$

$$= \prod_{i=1}^{n} P\left(\left(X_i = x_i\right) \middle| \text{Assignments of Parents}\left(X_i\right)\right)$$

So any entry in joint pdf table can be computed. And so **any conditional probability** can be computed.

Where are we now?

- We have a methodology for building Bayes nets.
- No exponential storage to hold our probability table. Only exponential in the maximum number of parents of any node.
- We can compute probabilities in time linear with the number of nodes.
- So we can also compute answers to any question.



E.G. What could we do to compute P(R | T,~S)?

Where are we now?

Step 1: Compute P(R ^ T ^ ~S)

Step 2: Compute P(~R ^ T ^ ~S)

Step 3: Return

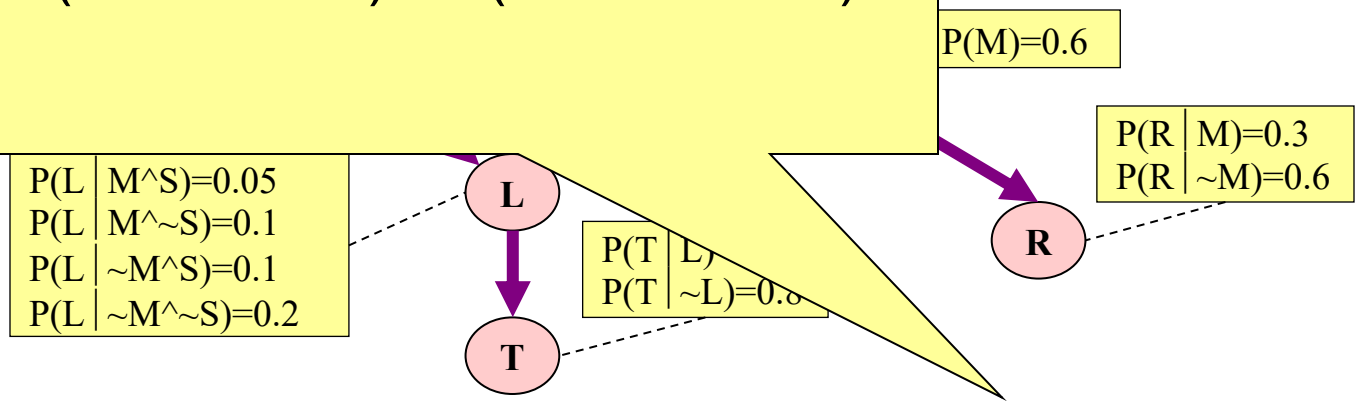
$$\frac{P(R \wedge T \wedge \neg S)}{P(R \wedge T \wedge \neg S) + P(\neg R \wedge T \wedge \neg S)}$$

g Bayes nets.

ge to hold our probability
mum number of parents of

y given assignment of truth
do it in time linear with the

any questions.



E.G. What could we do to compute P(R | T,~S)?

Where are we now?

Step 1: Compute P(R ^ T ^ ~S)

Step 2: Compute P(~R ^ T ^ ~S)

Step 3: Return

$$\frac{P(R \wedge T \wedge \sim S)}{P(R \wedge T \wedge \sim S) + P(\sim R \wedge T \wedge \sim S)}$$

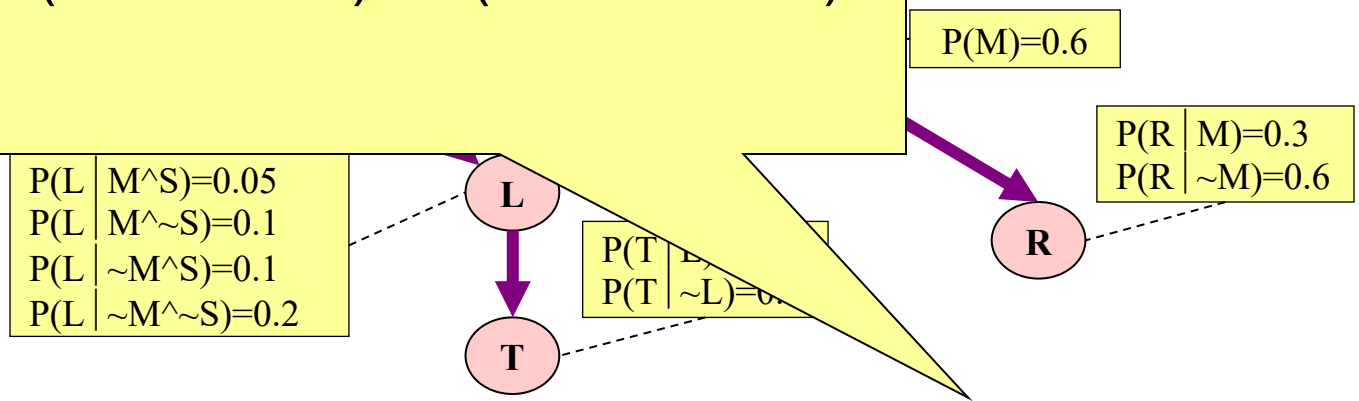
Sum of all the rows in the Joint that match R ^ T ^ ~S

Sum of all the rows in the Joint that match ~R ^ T ^ ~S

ximum number of parents of

n do it in time linear with the

o any questions.



E.G. What could we do to compute P(R | T,~S)?

Where are we now?

Step 1: Compute P(R ^ T ^ ~S)

Step 2: Compute P(~R ^ T ^ ~S)

Step 3: Return

$$\frac{P(R \wedge T \wedge \sim S)}{P(R \wedge T \wedge \sim S) + P(\sim R \wedge T \wedge \sim S)}$$

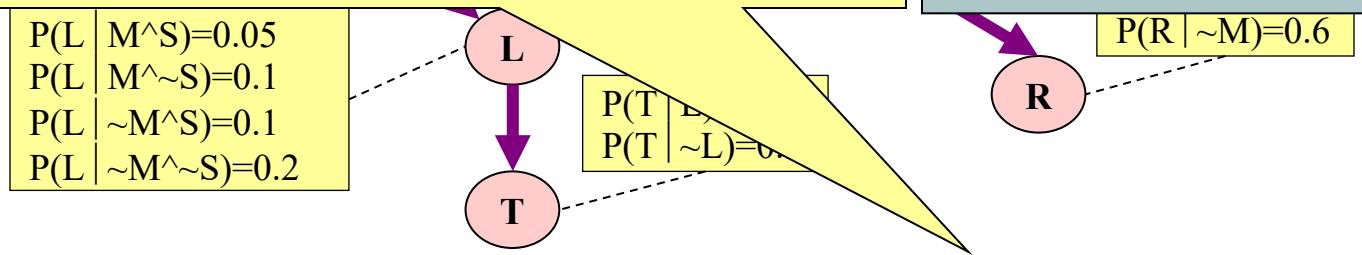
Sum of all the rows in the Joint that match R ^ T ^ ~S

4 joint computes

Sum of all the rows in the Joint that match ~R ^ T ^ ~S

4 joint computes

Each of these obtained by the "computing a joint probability entry" method of the earlier slides



E.G. What could we do to compute P(R | T,~S)?

The good news

We can do inference. We can compute any conditional probability:

P(Some variable | Some other variable values)

$$P(E_1 \mid E_2) = \frac{P(E_1 \wedge E_2)}{P(E_2)} = \frac{\sum_{\text{joint entries matching } E_1 \text{ and } E_2} P(\text{joint entry})}{\sum_{\text{joint entries matching } E_2} P(\text{joint entry})}$$

The good news

We can do inference. We can compute any conditional probability:

P(Some variable | Some other variable values)

$$P(E_1 \mid E_2) = \frac{P(E_1 \wedge E_2)}{P(E_2)} = \frac{\sum_{\text{joint entries matching } E_1 \text{ and } E_2} P(\text{joint entry})}{\sum_{\text{joint entries matching } E_2} P(\text{joint entry})}$$

Suppose you have *m* binary-valued variables in your Bayes Net and expression E_2 mentions *k* variables.

How much work is the above computation?

The sad, bad news

Conditional probabilities by enumerating all matching entries in the joint are expensive:

Exponential in the number of variables.

The sad, bad news

Conditional probabilities by enumerating all matching entries in the joint are expensive:

Exponential in the number of variables.

But perhaps there are faster ways of querying Bayes nets?

- If ever asked to manually do a Bayes Net inference -> many tricks to save you time -> not topic of this class though ☹

The sad, bad news

Conditional probabilities by enumerating all matching entries in the joint are expensive:

Exponential in the number of variables.

But perhaps there are faster ways of querying Bayes nets?

- If ever asked to manually do a Bayes Net inference -> many tricks to save you time -> not topic of this class though ☹

Sadder and worse news:

General querying of Bayes nets is NP-complete.

Summary of the Bad News

- Exact inference is feasible in small to medium-sized networks
- Exact inference in large networks takes a very long time
- We resort to approximate inference techniques which are much faster and give pretty good results

One last unresolved issue…

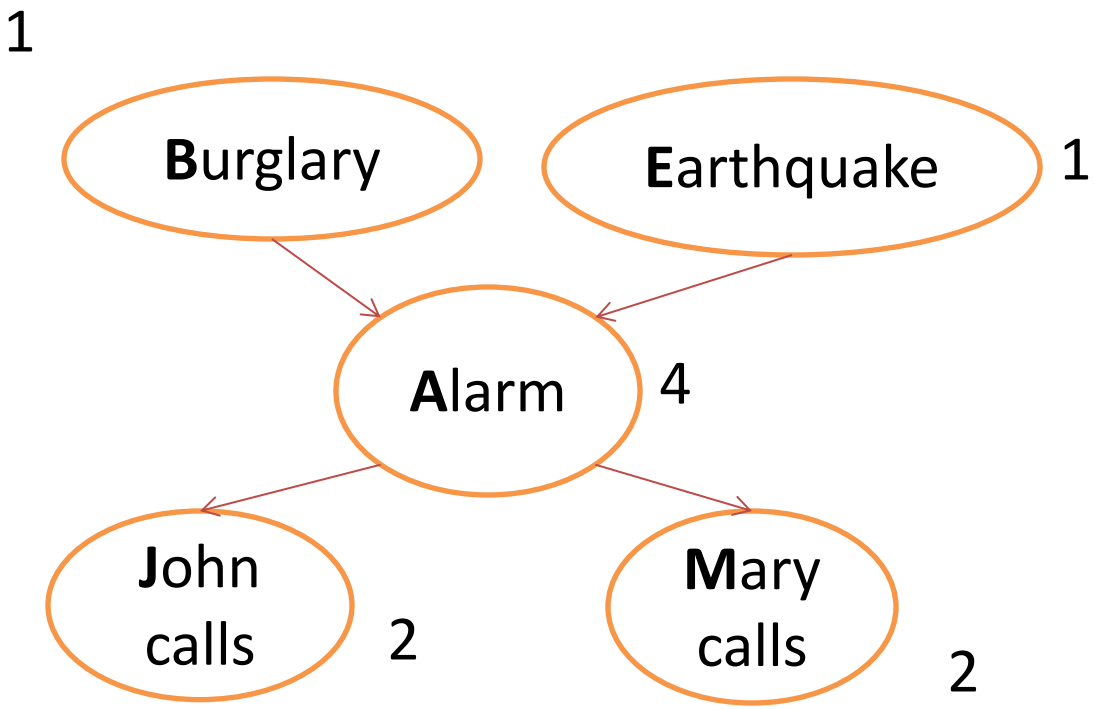
Where do we get the Bayesian network from?

Two options:

- Get an expert to design it
- Learn it from data

More examples and tutorials:
http://www.bayesserver.com/Live.aspx
http://dsaldana.github.io/sallybn/doc/tutorial/tutorial.html

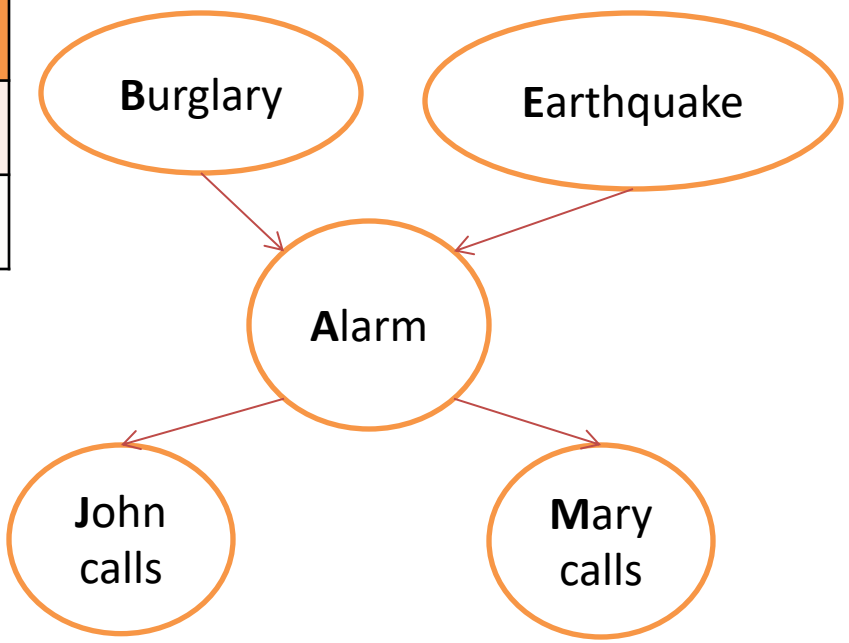
Example: Alarm Network



How many parameters? 10

Example: Alarm Network

B	P(B)
+b	0.001
¬b	0.999



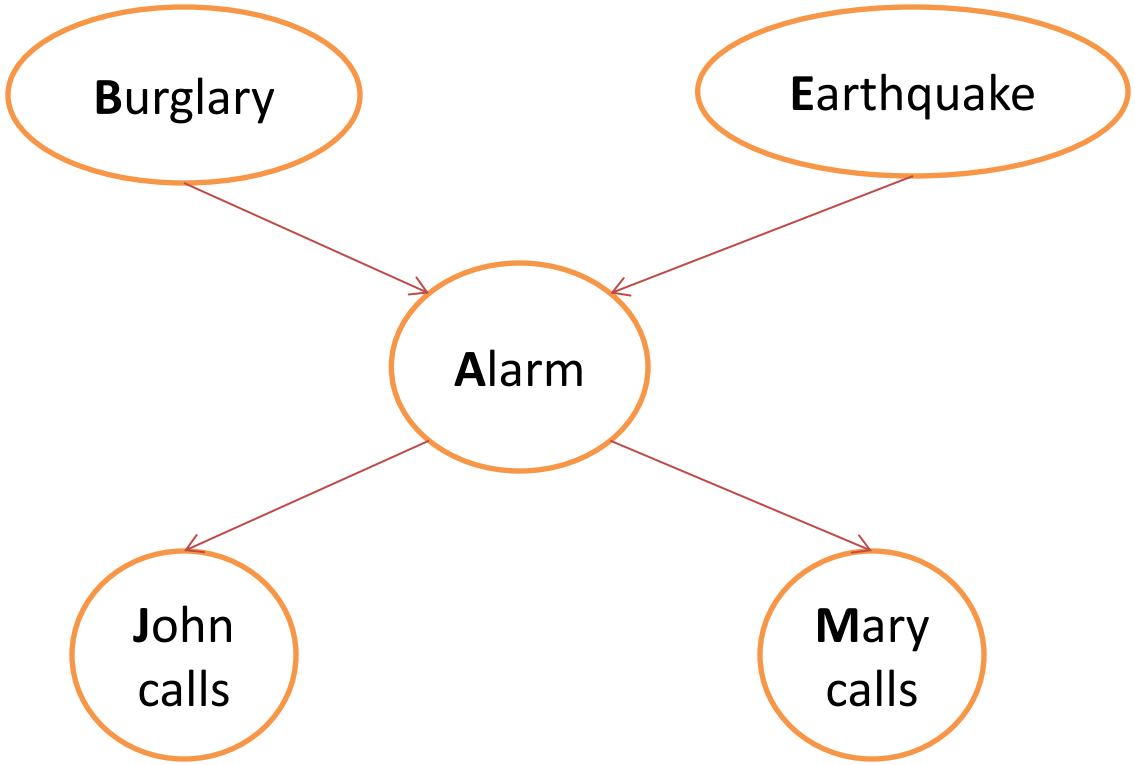
E	P(E)
+e	0.002
¬e	0.998

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

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

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

Example: Alarm Network



$$\prod_i P(X_i|\text{Parents}(X_i)) = P(B) \cdot P(E) \cdot P(A|B,E) \cdot P(J|A) \cdot P(M|A)$$

Bayes Net for Liver Diagnosis