Artificial Intelligence

CS4365 --- Fall 2022

Probabilistic Reasoning and Methods

Instructor: Yunhui Guo

Conditional independence

Suppose we have these three events

M : Lecture taught by Mr. B

L : Lecturer arrives late

R : Lecture concerns robots

Suppose Mr. A has **a higher chance of being late** than Mr. B.

Suppose Mr. A has **a higher chance of giving robotics lectures** than Mr. B.

What kind of independence can we find?

How about:

$P(L \mid M) = P(L)$?

$P(R \mid M) = P(R)$?

$P(L \mid R) = P(L)$?

Conditional independence

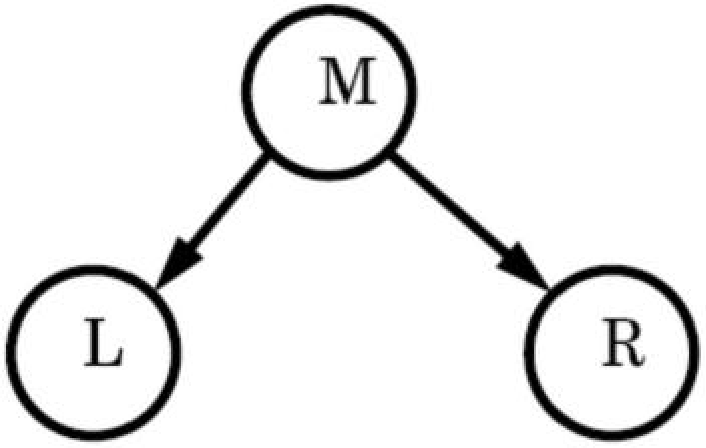
- Once you know who the lecturer is, then whether they arrive late doesn't affect whether the lecture concerns robots.

$$P(R \mid M , L) = P(R \mid M) \text{ and } P(R \mid \sim M , L) = P(R \mid \sim M)$$

We express this in the following way:

"R and L are conditionally independent given M"

Which is also notated by the following diagram.



Given knowledge of M, knowing anything else in the diagram won't help us with L, etc.

Conditional independence formalized

- R and L are conditionally independent given M if

  For all x,y,z in {T,F}:

  $$P(R=x \mid M=y, L=z) = P(R = x \mid M = y)$$

  Also written as $(R \perp\!\!\!\perp L \mid M)$

  R and L are independent events given M

Conditional independence formalized

- More generally:

  Let S1 and S2 and S3 be sets of variables.

  Set-of-variables S1 and set-of-variables S2 are conditionally independent given S3 if for all assignments of values to the variables in the sets,

  P(S1's assignments | S2's assignments and S3's assignments) =
  P(S1's assignments | S3's assignments)

Factorizing into Marginals

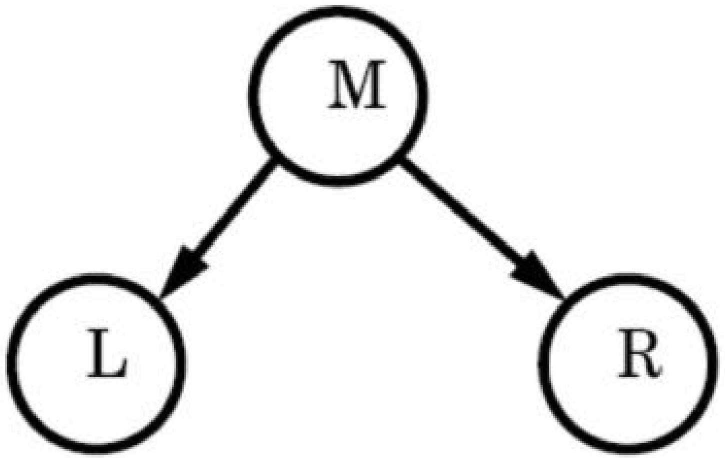
- Can be expressed in slightly different way

$$P(R, L \mid M) = P(R|L, M)P(L|M)$$
$$= P(R|M)\,P(L|M)$$

- That is joint distribution factorizes into product of marginals

Conditional Independence

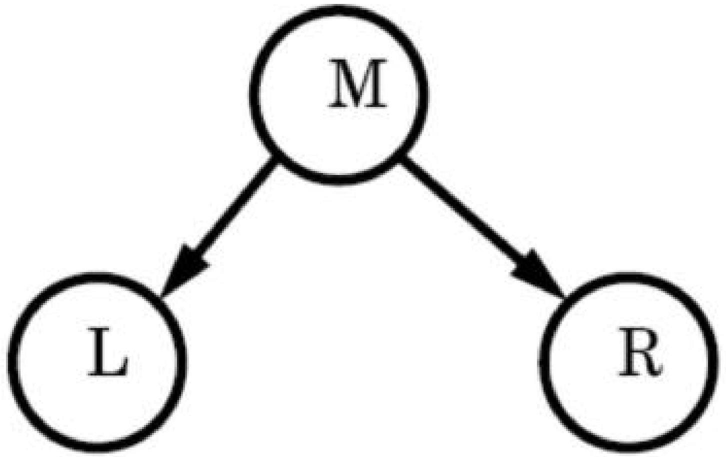
- We can write down P(M). And then, since we know L is only directly influenced by M, we can write down the values of P(L|M) and P(L|~M) and know we've fully specified L's behavior. Same for R.
- $P(M) = 0.6$
- $P(L \mid M) = 0.085$
- $P(L \mid \sim M) = 0.17$
- $P(R \mid M) = 0.3$
- $P(R \mid \sim M) = 0.6$



'R and L conditionally independent given M'

Conditional Independence

- P(M) = 0.6
- P(L | M) = 0.085
- P(L | ~M) = 0.17
- P(R | M) = 0.3
- P(R | ~M) = 0.6



Conditional Independence:

- P(R|M , L) = P(R|M) ,
- P(R | ~M , L) = P(R | ~M)

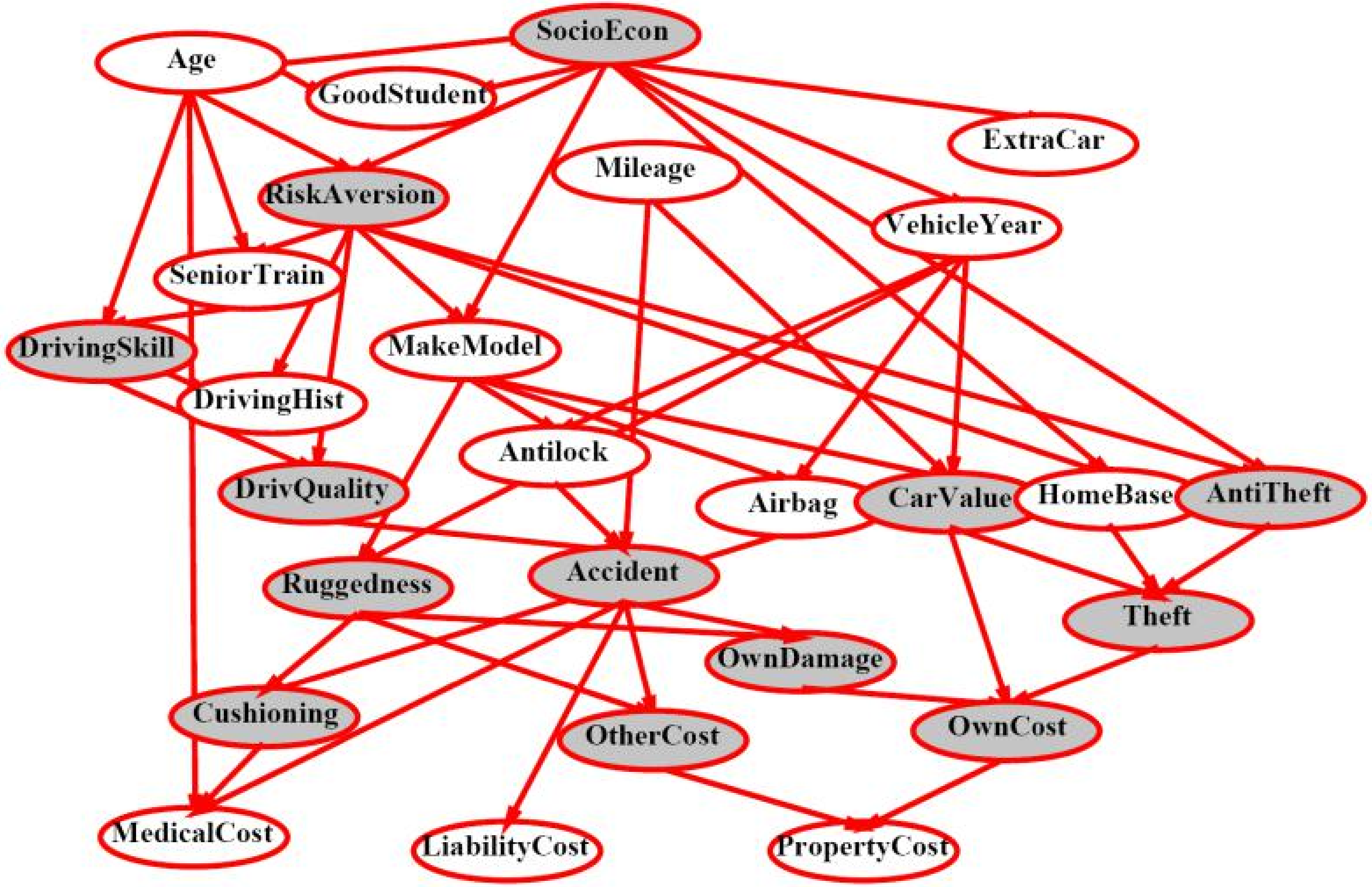
Once again, we can obtain any member of the joint probability distribution that we desire:

- P(L, R, M) =

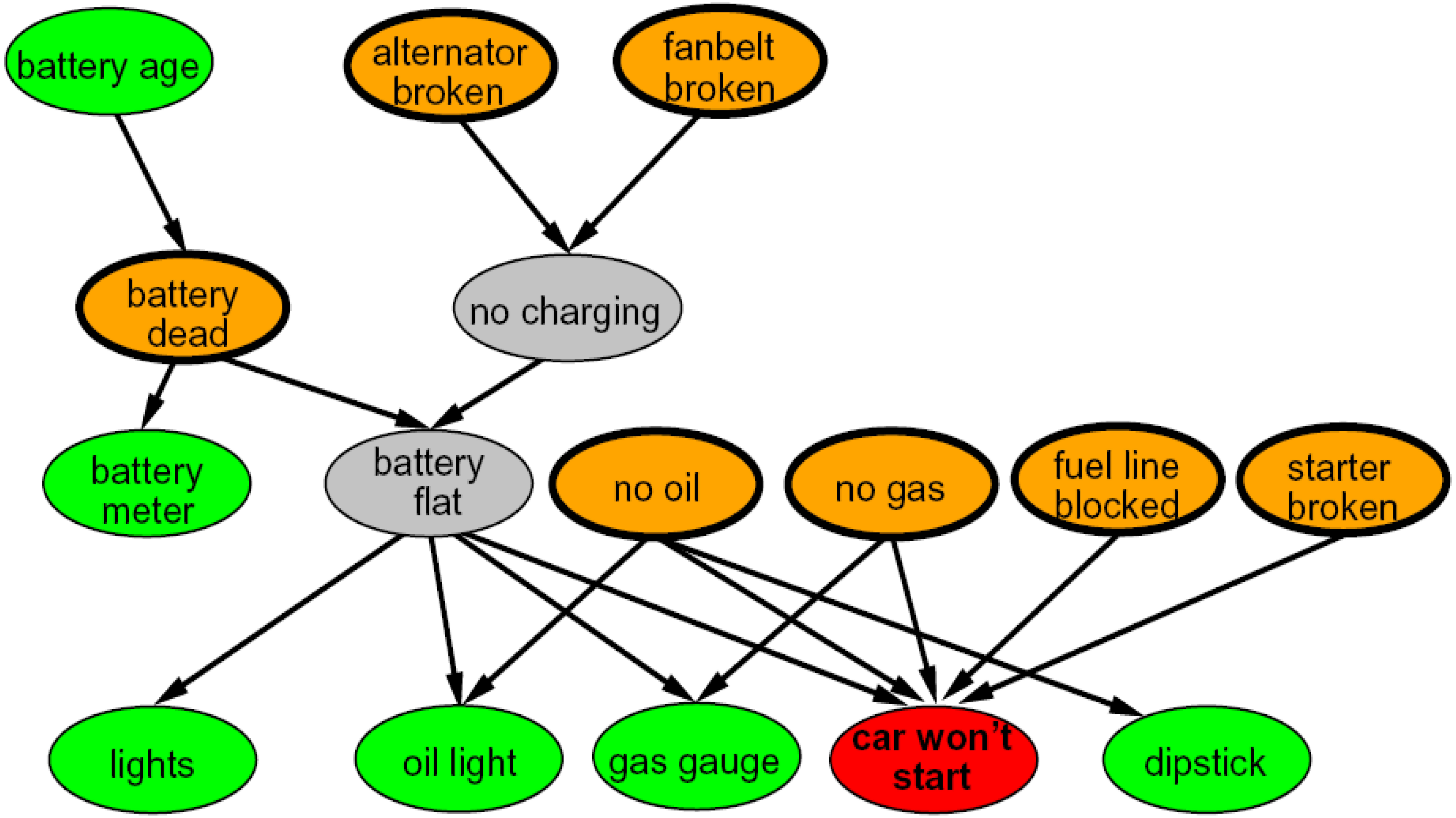
A Bayes Net

- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions

A Bayes Net


SocioEcon
Age
GoodStudent
ExtraCar
Mileage
RiskAversion
VehicleYear
SeniorTrain
DrivingSkill
MakeModel
DrivingHist
Antilock
DrivQuality
Airbag
CarValue
HomeBase
AntiTheft
Ruggedness
Accident
Theft
OwnDamage
Cushioning
OtherCost
OwnCost
MedicalCost
LiabilityCost
PropertyCost

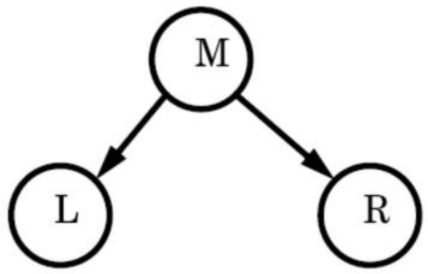
A Bayes Net


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

A Bayes Net

- Represents a set of variables and their conditional dependencies via a directed acyclic graph

- **Nodes** represent **variables**,
    - Can be assigned (observed) or unassigned (unobserved)
- **Edges** represent **conditional dependencies**
    - Similar to CSP constraints



- Nodes that are not connected represent variables that are conditionally independent of each other
- For example, a Bayes network could represent the probabilistic relationships between diseases and symptoms

Example: Coin Flips

- N independent coin flips

- No interactions between variables: absolute independence

Example: Traffic

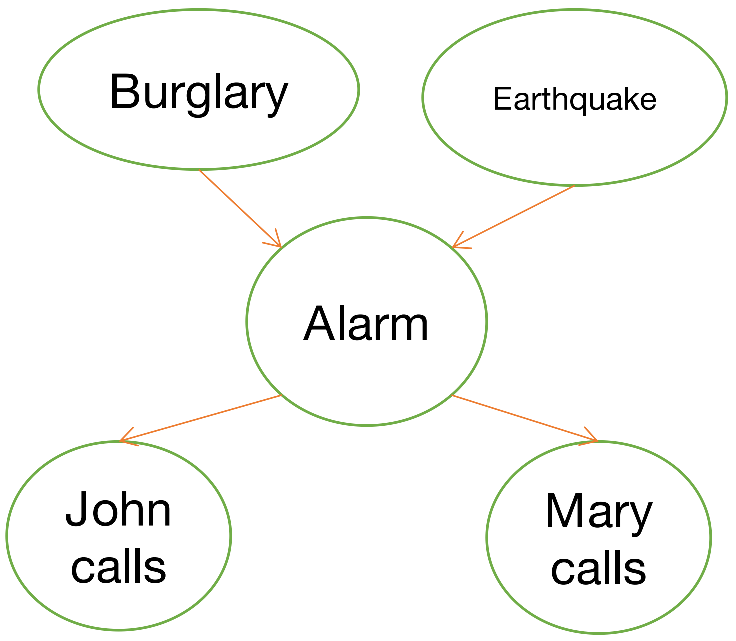
- **Variables**:
  - R: It rains
  - T: There is traffic
- Model 1: independence

- Model 2: rain causes traffic

- Why model 2 is better?

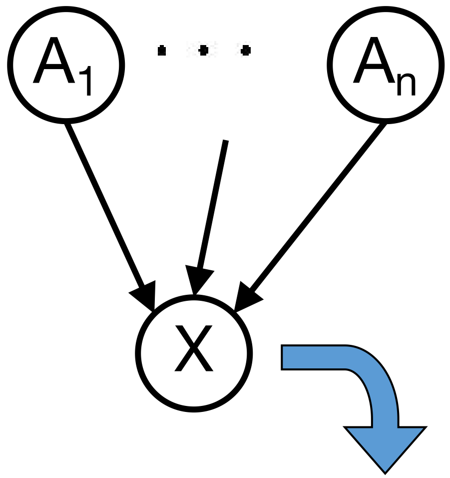
Example: Alarm Network

- **Variables**:
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!



Bayes' Net Semantics

- **Variables**: a set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents' values
  - CPT: conditional probability table
  - Description of a noisy "causal" process



$$P(X|A_1 \ldots A_n)$$

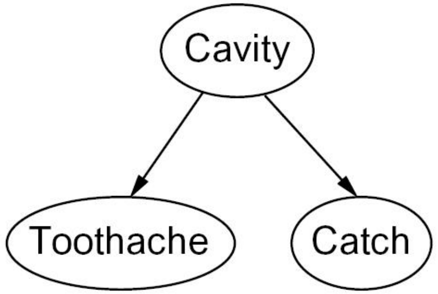
- *A Bayes net = Topology (graph) + Local Conditional Probabilities*

Probabilities in BNs

- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

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

- Example:



$$P(\mathit{+cavity}, \mathit{+catch}, \mathit{-toothache})$$

Probabilities in BNs

- Why are we guaranteed that setting

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

results in a proper joint distribution?

- Chain rule (valid for all distributions): $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$
- Assume conditional independences: $P(x_i | x_1, \ldots x_{i-1}) = P(x_i | \mathit{parents}(X_i))$
    - Consequence:

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

Example: Coin Flips

X_1 X_2 $\cdots$ X_n

$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

$\cdots$

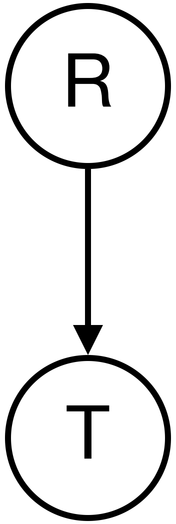
$P(X_n)$

h	0.5
t	0.5

$$P(h, h, t, h) =$$

Only distributions whose variables are absolutely independent can be represented by a Bayes' net with no arcs.

Example: Traffic



$P(R)$

+r	1/4
-r	3/4

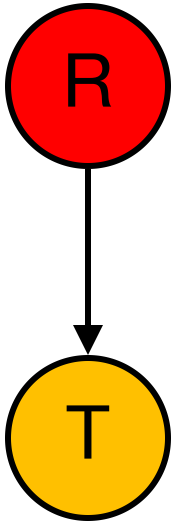
$P(T|R)$

+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$$P(+r, -t) =$$

Example: Traffic

- Causal direction



$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

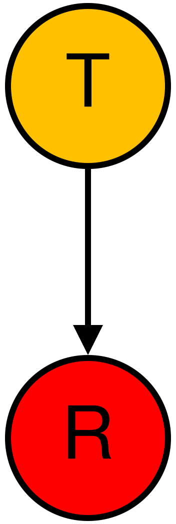
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$P(T,R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

Example: Traffic

- Reverse causality?



$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

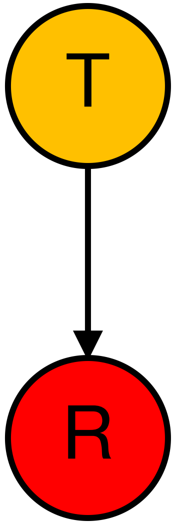
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$P(T,R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

Example: Traffic

- Reverse causality?



$P(T)$

+t	9/16
-t	7/16

$P(R|T)$

+t	+r	1/3
	-r	2/3

-t	+r	1/7
	-r	6/7

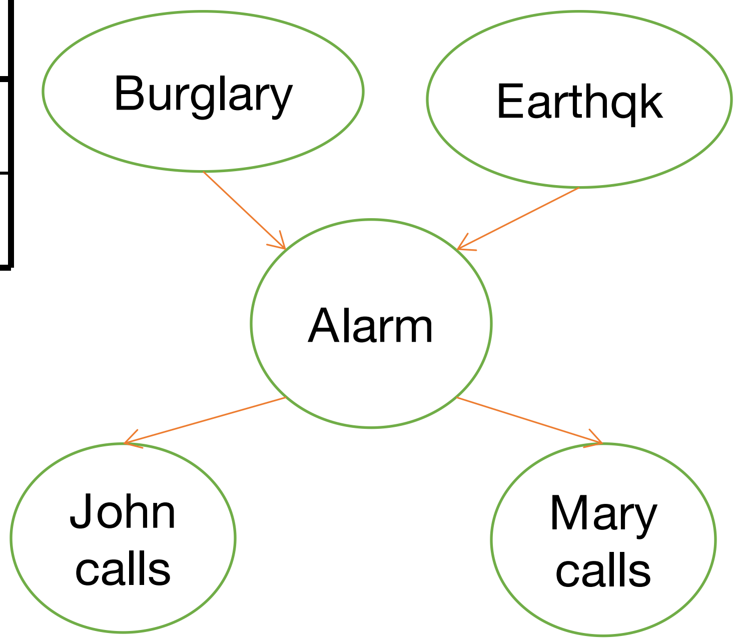
$P(T, R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

- BNs need not actually be causal

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



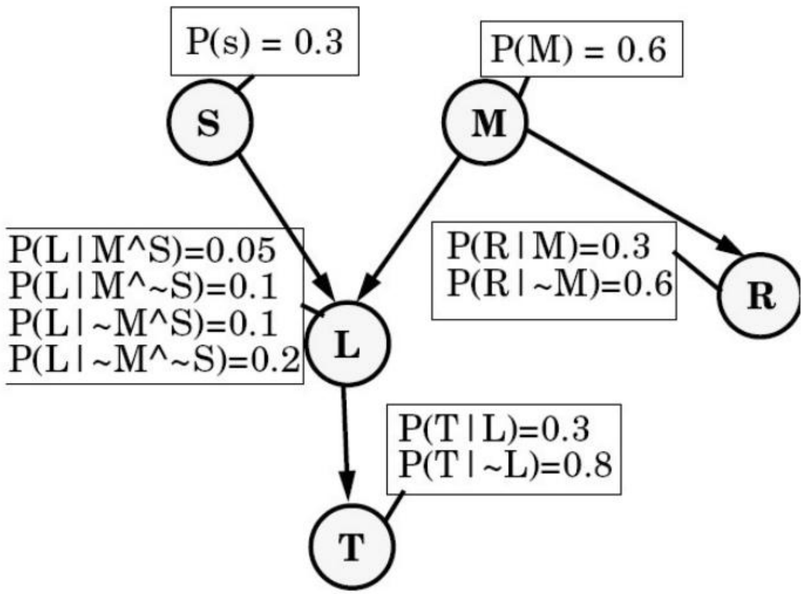
E	P(E)
+e	0.002
-e	0.998

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

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

A Bayes Net



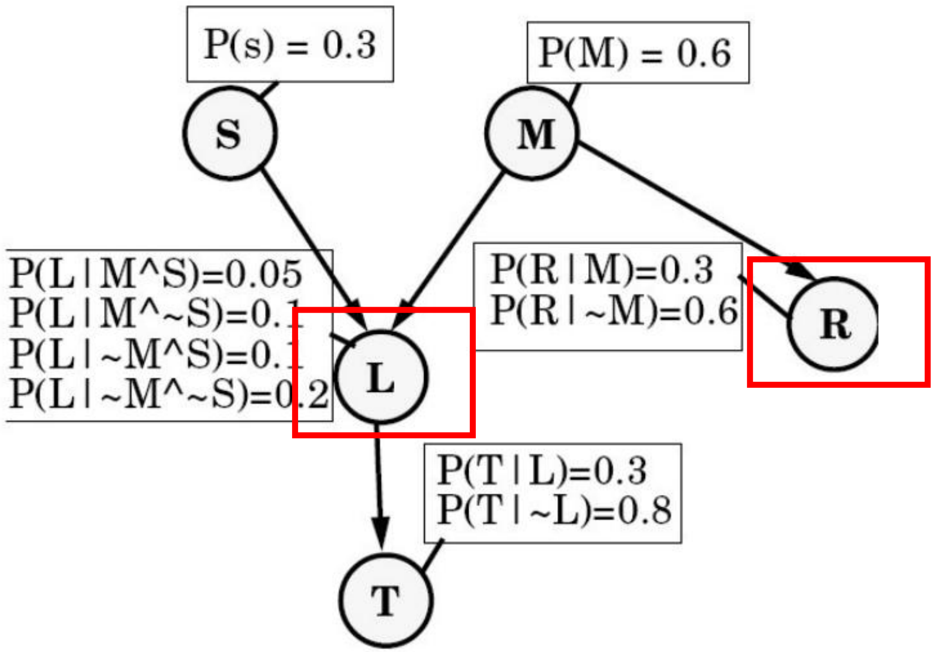
Suppose:
T : The lecture started by 10:35
L : The lecturer arrives late
R : The lecture concerns robots
M : The lecturer is Mr. M
S : It is sunny

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

A Bayes Net



- Two unconnected variables still can affect each other.
- Each node is conditionally independent of anyone earlier in the tree, given its parents
- You can deduce many other conditional independence relations from a Bayes net.

Building a Bayes Net

We will place an ordering on nodes (call them X_1 , X_2 ... X_n), such that:

X_1 has no parents,

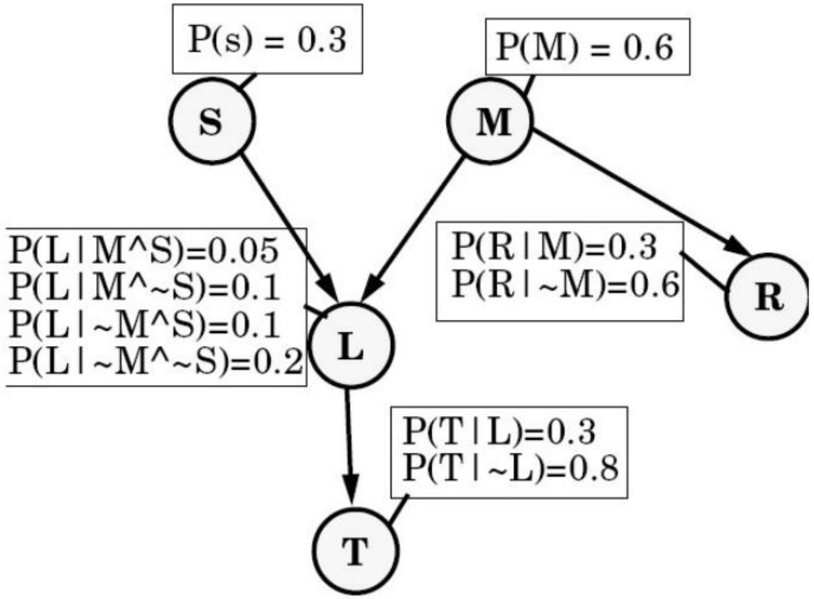
Parents(X_i) is a subset of {X_1, X_2 , ... X_{i-1}}

This can always be done for any acyclic graph. There are usually multiple solutions.

To make a Bayes Net, follow these rules

1. Choose a set of relevant variables.
2. Choose an ordering for them X_1 .. X_n
3. While there are variables left:
    1. Pick X_i and add a node to the network
    2. Set Parents(X_i) to be a minimal set of already-added nodes such that we have conditional independence of X_i and all other members of {X_1 ... X_{i-1}} given Parents(X_i)
    3. Define the conditional prob. table of P(X_i=x | Assignments of Parents(X_i)).

Computing with a Bayes Net



- The first thing we might want to do is compute an entry in a joint probability table
- Given an assignment of truth values to our variables, what is the probability? E.g., What is P(S, ~M, L, ~R, T)?

What you should know

- The meanings of independence and conditional independence

- The definition of a Bayes net

- Computing probabilities of assignments of variables (i.e. members of the joint p.d.f) with a bayes net

- The method for computing arbitrary conditional probabilities