Artificial Intelligence

CS4365 --- Fall 2022

Bayesian Networks: Conditional Independencies

Instructor: Yunhui Guo

Probability Recap

- Conditional probability $P(x|y) = \frac{P(x,y)}{P(y)}$
- Product rule $P(x,y) = P(x|y)P(y)$
- Chain rule
$$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1,X_2)\ldots \\ &= \prod_{i=1}^{n} P(X_i|X_1,\ldots,X_{i-1}) \end{aligned}$$
- X, Y independent if and only if: $\forall x, y : P(x,y) = P(x)P(y)$
- X and Y are conditionally independent given Z if and only if:

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

A Bayes Net

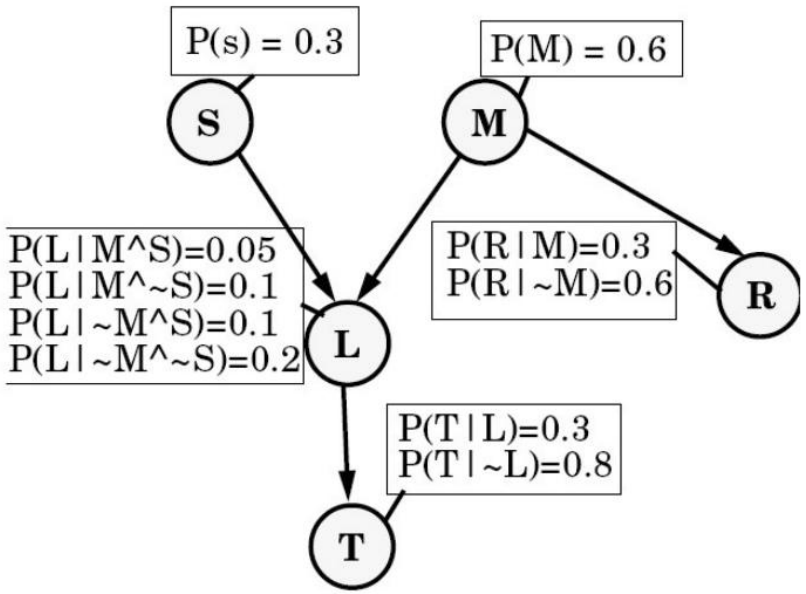
- A directed, acyclic graph, one node per random variable

- A conditional probability table (CPT) for each node
  - A collection of distributions over X, one for each combination of parents' values $P(X|a_1 \ldots a_n)$
- 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:

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

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

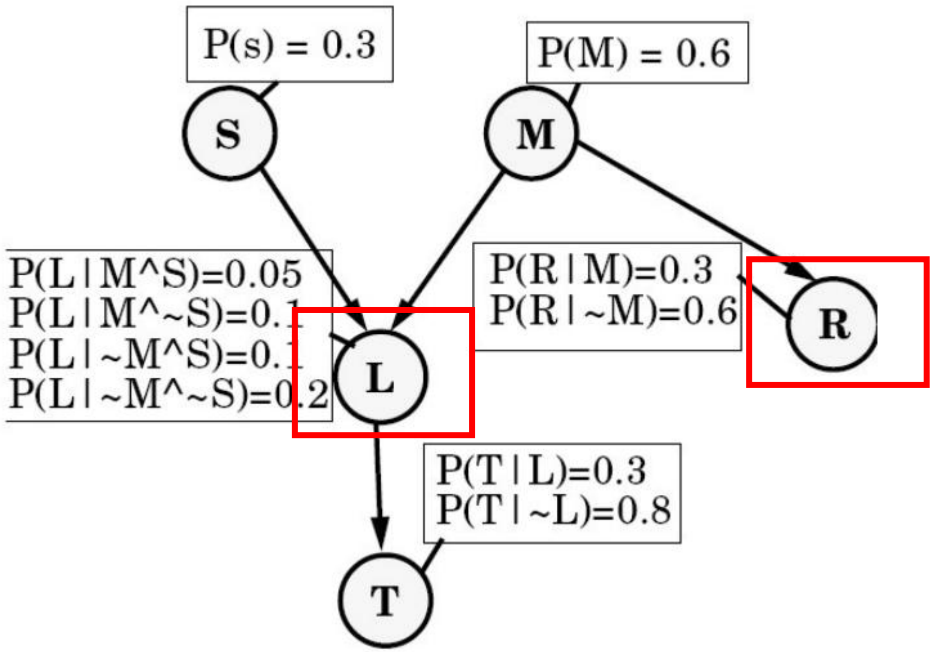


T only directly influenced by L (i.e. T is conditionally independent of R,M,S given L)

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,

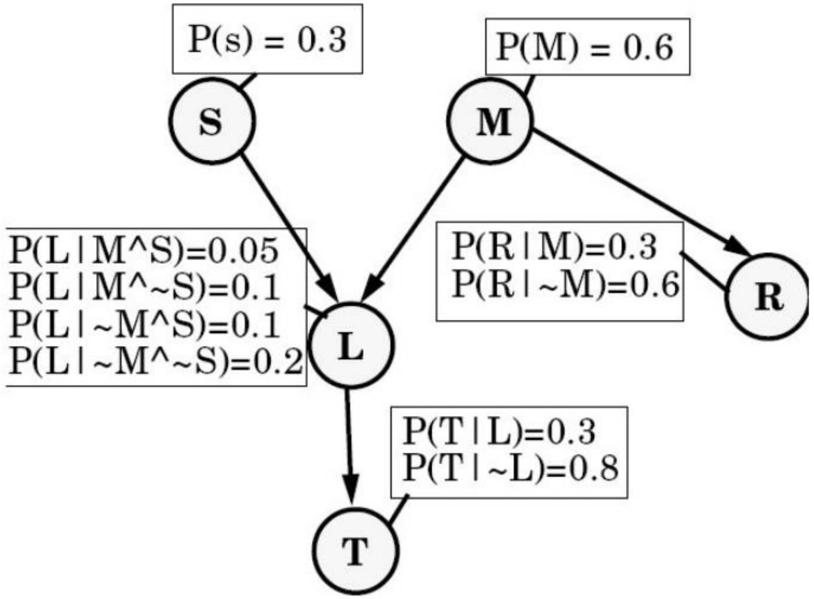
Parents(X_i) is a subset of $\{X_1, X_2, \ldots X_{i-1}\}$

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 $\{X_1 \ldots X_{i-1}\}$ given Parents(X_i)
    3. Define the conditional prob. table of $P(X_i=x \mid \text{Assignments of Parents}(X_i))$.

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

What Independencies does a Bayes Net Model?

- In order for a Bayesian network to model a probability distribution, the following must be true by definition:

  Each variable is conditionally independent of all its nondescendants in the graph given the value of all its parents.

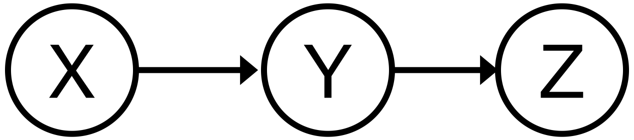
- This implies

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

- But what else does it imply?

Independence in a BN

- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counter example
  - Example:



  - Question: are X and Z necessarily independent?
    - Answer: no. Example: Easy exam causes Get A, which causes Get recommended.
    - X can influence Z, Z can influence X (via Y)
    - They could be independent: how?

Independence in a BN

- Given Y, does learning the value of Z tell us nothing new about X?

    I.e., is P(X|Y,Z) equal to P(X|Y)

    Yes. Since we know the value of all of X's parents (namely, Y), and Z is not a descendant of X, X is conditionally independent of Z.

Also, since independence is symmetric,

$$P(Z|Y,X) = P(Z|Y)$$

Quick proof that independence is symmetric

- Assume: $P(X|Y, Z) = P(X|Y)$
- Then:

Quick proof that independence is symmetric

- Assume: P(X|Y, Z) = P(X|Y)
- Then:

$$P(Z|X,Y) = \frac{P(X,Y|Z)P(Z)}{P(X,Y)}$$

(Bayes Rule)

Quick proof that independence is symmetric

- Assume: P(X|Y, Z) = P(X|Y)
- Then:

$$P(Z|X,Y) = \frac{P(X,Y|Z)P(Z)}{P(X,Y)} \quad \text{(Bayes Rule)}$$

$$= \frac{P(Y|Z)P(X|Y,Z)P(Z)}{P(X|Y)P(Y)} \quad \text{(Chain Rule)}$$

Quick proof that independence is symmetric

- Assume: P(X|Y, Z) = P(X|Y)
- Then:

$$P(Z|X,Y) = \frac{P(X,Y|Z)P(Z)}{P(X,Y)} \qquad \text{(Bayes Rule)}$$

$$= \frac{P(Y|Z)P(X|Y,Z)P(Z)}{P(X|Y)P(Y)} \qquad \text{(Chain Rule)}$$

$$= \frac{P(Y|Z)P(X|Y)P(Z)}{P(X|Y)P(Y)} \qquad \text{(By Assumption)}$$

Quick proof that independence is symmetric

- Assume: P(X|Y, Z) = P(X|Y)
- Then:

$$
\begin{aligned}
P(Z|X,Y) &= \frac{P(X,Y|Z)P(Z)}{P(X,Y)} && \text{(Bayes Rule)} \\
&= \frac{P(Y|Z)P(X|Y,Z)P(Z)}{P(X|Y)P(Y)} && \text{(Chain Rule)} \\
&= \frac{P(Y|Z)P(X|Y)P(Z)}{P(X|Y)P(Y)} && \text{(By Assumption)} \\
&= \frac{P(Y|Z)P(Z)}{P(Y)} = P(Z|Y) && \text{(Bayes Rule)}
\end{aligned}
$$

D-separation

Theorem [Verma & Pearl, 1998]:

If a set of evidence variables E d-separates X and Z in a Bayesian network's graph, then $(X \perp\!\!\!\perp Z \mid E)$.

D-separation: a condition / algorithm for automatically inferring whether learning the value of one variable might give us any additional hints about some other variable, given what we already know.

Study independence properties for triples

Causal Chains

- This configuration is a "causal chain"

$$X \rightarrow Y \rightarrow Z \qquad P(x,y,z) = P(x)P(y|x)P(z|y)$$

X: Exam easy Y: Get A Z: Get recommended

- Guaranteed X independent of Z ? *No!*
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
  - Example:
    - Eaxm easy causes Get A causes Get recommended.
    - In numbers:

      P(+y | +x) = 1, P(-y | -x) = 1,
      P(+z | +y) = 1, P(-z | -y) = 1

Causal Chains

- This configuration is a "causal chain"

$$X \rightarrow Y \rightarrow Z \qquad P(x, y, z) = P(x)P(y|x)P(z|y)$$

X: Exam easy   Y: Get A   Z: Get recommended

- Guaranteed X independent of Z given Y?

$$\begin{aligned} P(z|x, y) &= \frac{P(x, y, z)}{P(x, y)} \\ &= \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)} \\ &= P(z|y) \end{aligned}$$

Yes!

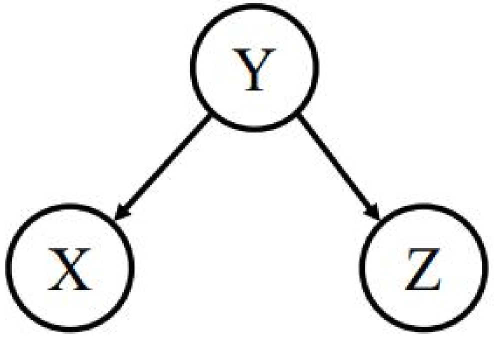
Evidence along the chain "blocks" the influence

Common Cause

- This configuration is a "common cause"

Y: Get A

X: Get reward



Z: Get recommended

- Guaranteed X independent of Z ? *No!*

$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

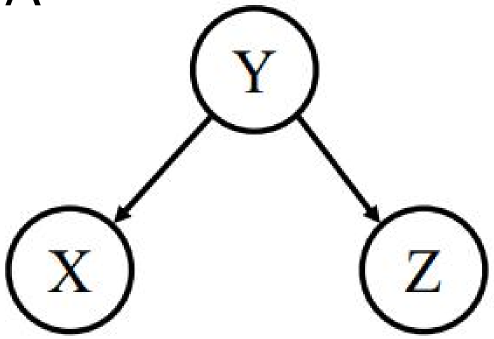
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
  - Example:
    - Get A causes both Get reward and Get recommended
    - In numbers:
      - P(+x | +y) = 1, P(-x | -y) = 1,
      - P(+z | +y) = 1, P(-z | -y) = 1

Common Cause

- This configuration is a "common cause"

Y: Get A

X: Get reward



Z: Get recommended

$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

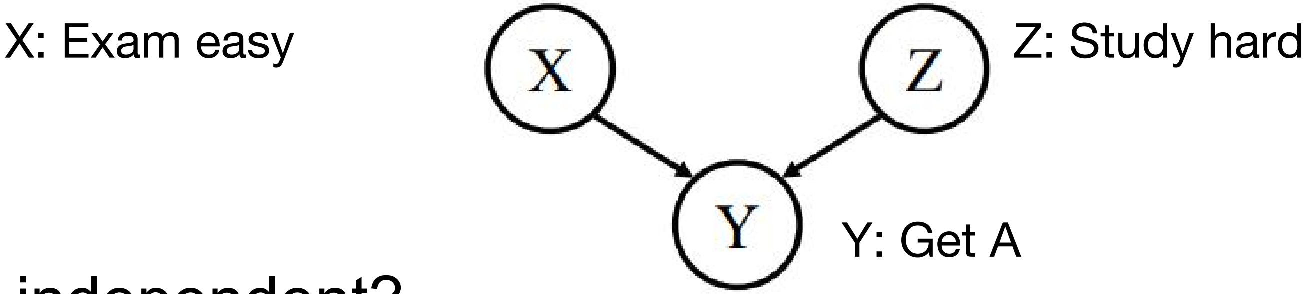
- Guaranteed X and Z independent given Y?

$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)} = P(z|y)$$

- Observing the cause blocks influence between effects.

Common Effect

- Last configuration: two causes of one effect (v-structures)



- Are X and Z independent?
  - Yes: Exam easy and Study hard cause Get A, but they are not correlated
- Are X and Z independent given Y?
  - No: seeing Get A puts Exam easy and Study hard in competition as explanation.
- This is backwards from the other cases
  - Observing an effect activates influence between possible causes.

The General Case

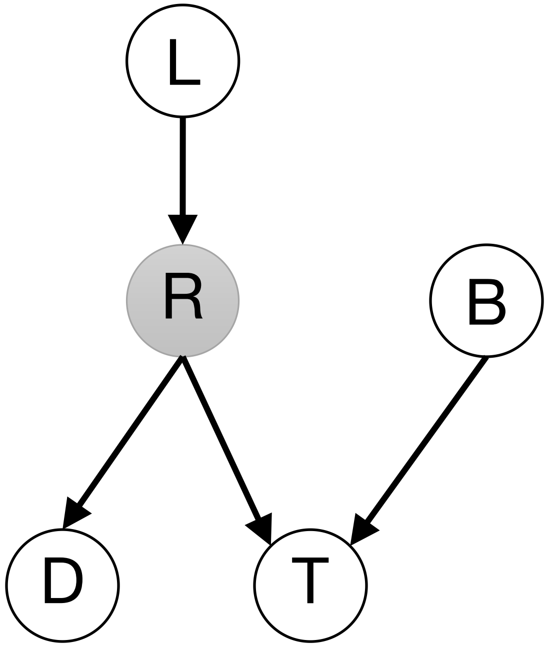
- General question: in a given BN, are two variables independent (given evidence)?
- Solution: analyze the graph
- Any complex example can be broken into repetitions of the three canonical cases

Reachability

- Recipe: shade evidence nodes, look for paths in the resulting graph

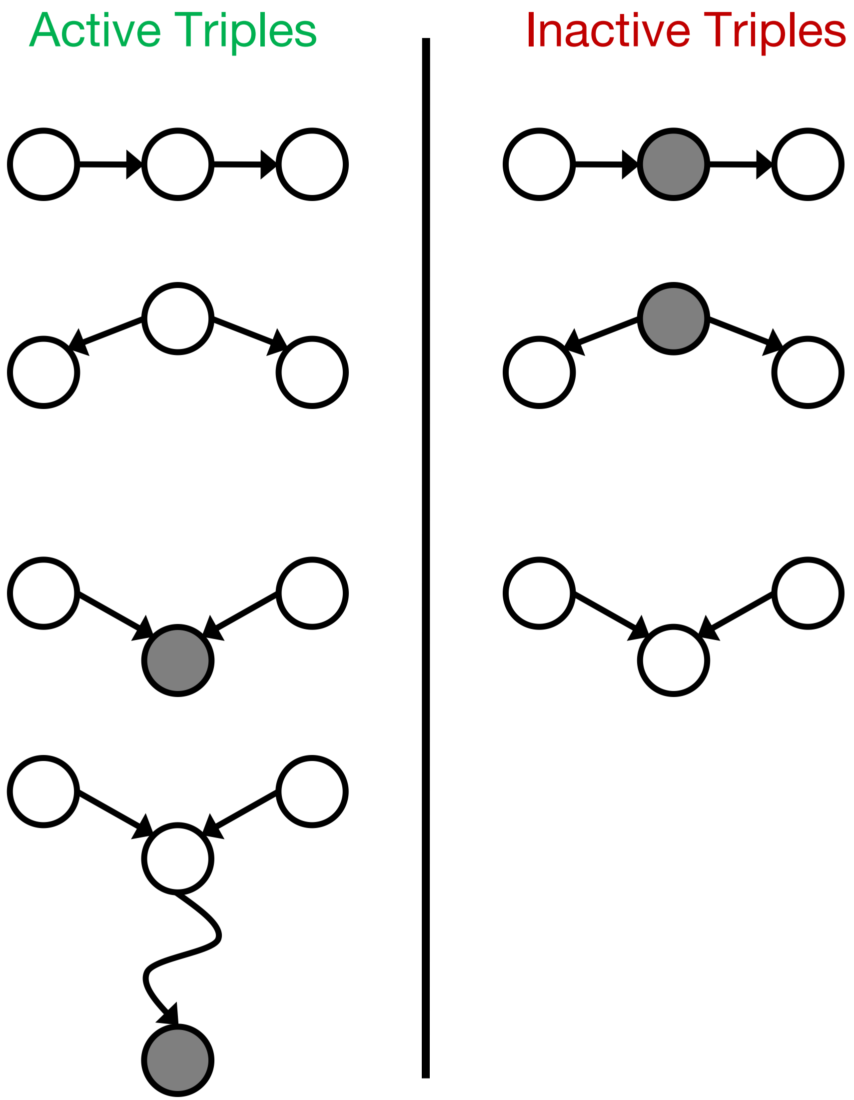
- Place balls on one of the variables

- If any ball can reach another random variable, then they are not conditionally independent, otherwise they are

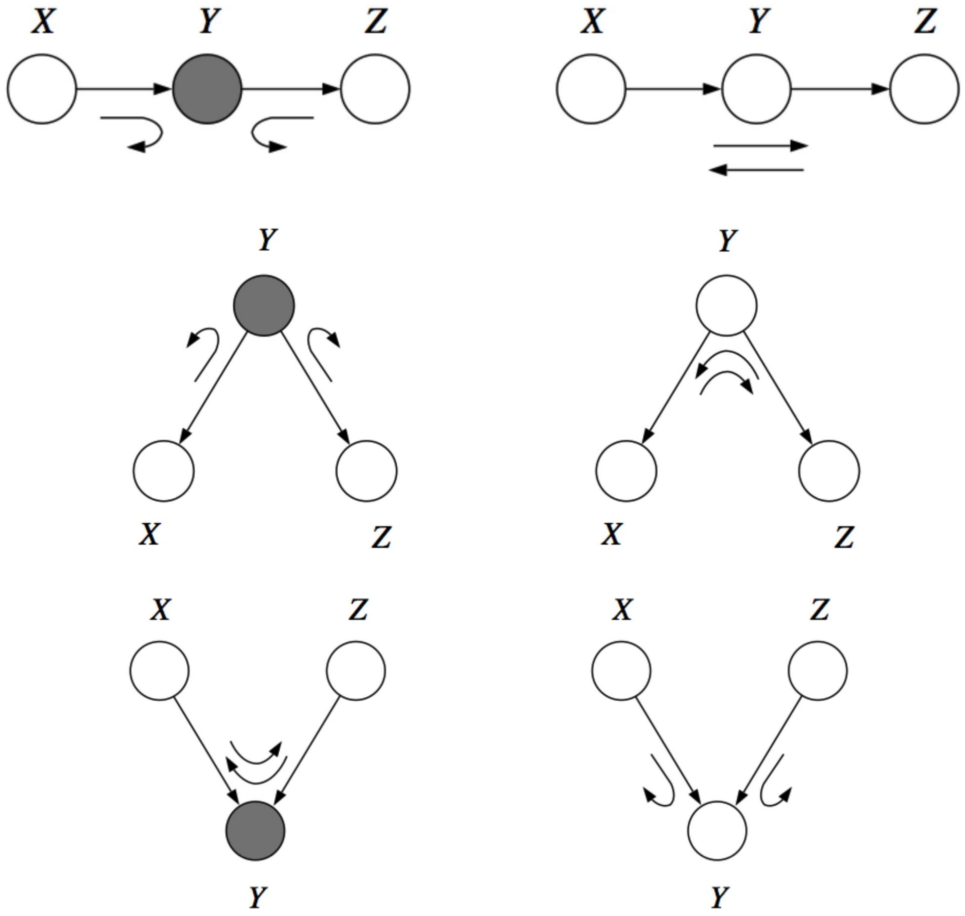


Active / Inactive Paths

- **Question**: Are X and Y conditionally independent given evidence variables {Z}?
  - Yes, if X and Y "d-separated" by Z
  - Consider all (undirected) paths from X to Y
  - No active paths = independence!
- A path is active if each triple is active:
  - Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
  - Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
  - Common effect (aka v-structure)
    - $A \rightarrow B \leftarrow C$ where B or one of its descendents is observed
- All it takes to block a path is a single inactive segment



Reachability



D-Separation

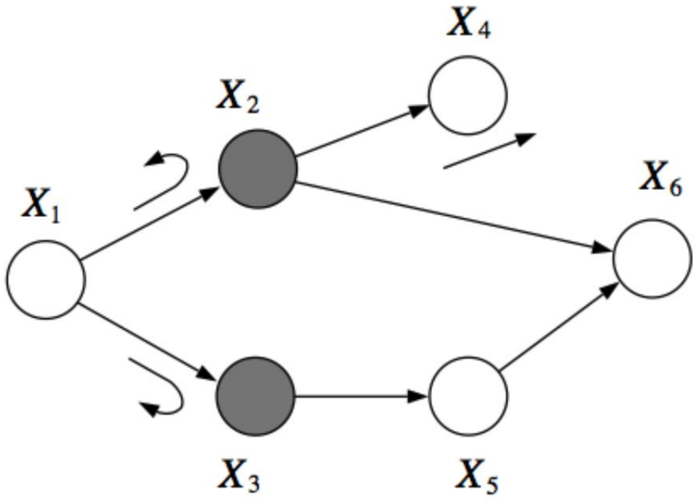
- Query: $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$
- Check all (undirected!) paths between X_i and X_j
  - If one or more active, then independence not guaranteed

$$X_i \not\!\perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

  - Otherwise (i.e. if all paths are inactive), then independence is guaranteed

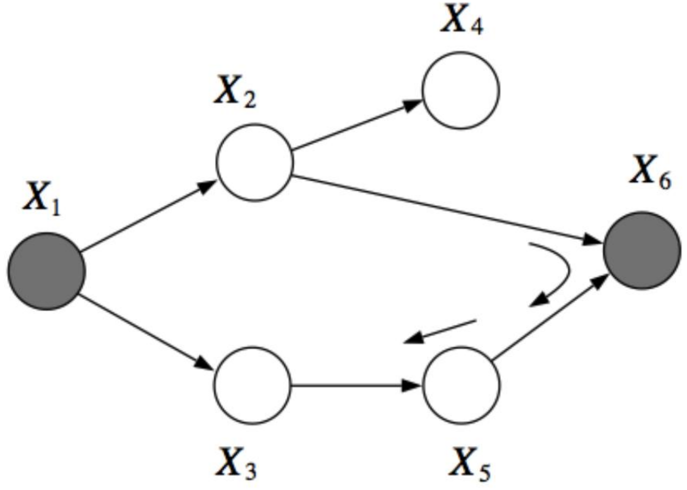
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

Reachability



- $(X_1 \perp\!\!\!\perp X_4 \mid X_2, X_3)$?

Reachability



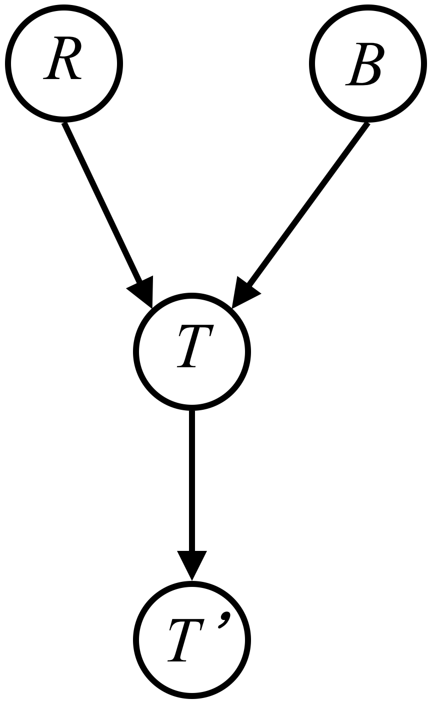
- $(X_2 \perp\!\!\!\perp X_3 \mid X_1, X_6)$?

Example

$R \perp\!\!\!\perp B$ Yes

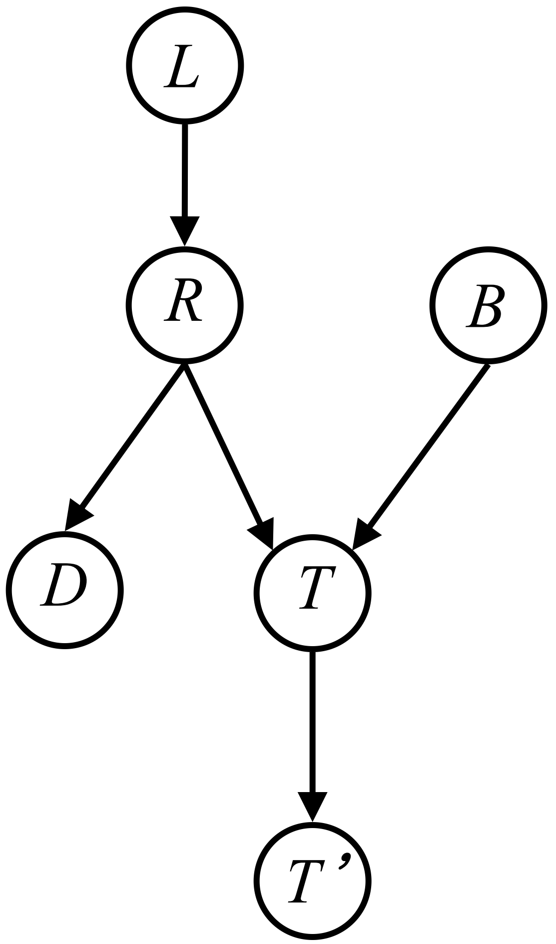
$R \perp\!\!\!\perp B | T$ No

$R \perp\!\!\!\perp B | T'$ No



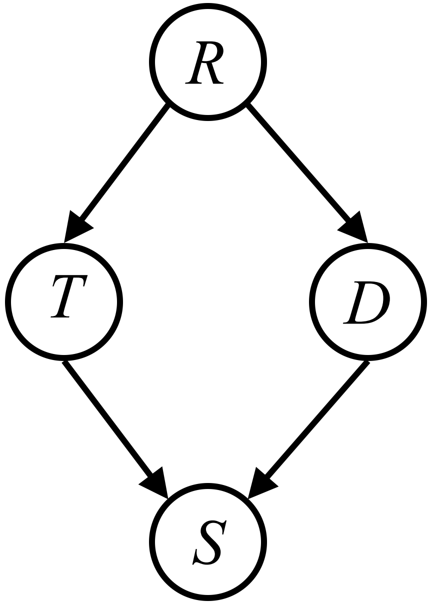
Example

$L \perp\!\!\!\perp T' \| T$	Yes
$L \perp\!\!\!\perp B$	Yes
$L \perp\!\!\!\perp B \| T$	No
$L \perp\!\!\!\perp B \| T'$	No
$L \perp\!\!\!\perp B \| T, R$	Yes



Example

- Variables:
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I'm sad



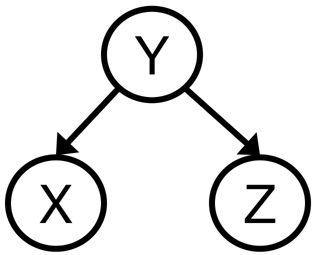
- Questions:

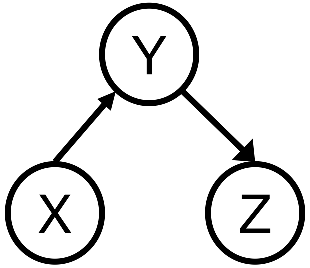
$T \perp\!\!\!\perp D$ No

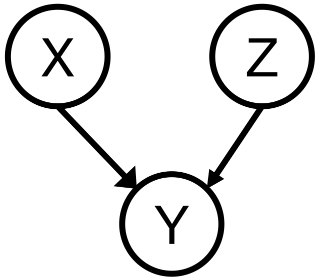
$T \perp\!\!\!\perp D | R$ Yes

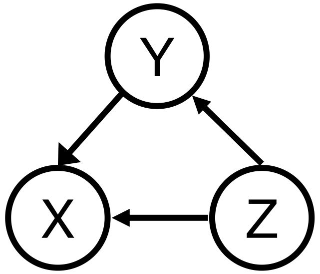
$T \perp\!\!\!\perp D | R, S$ No

Computing All Independences









Structure Implications

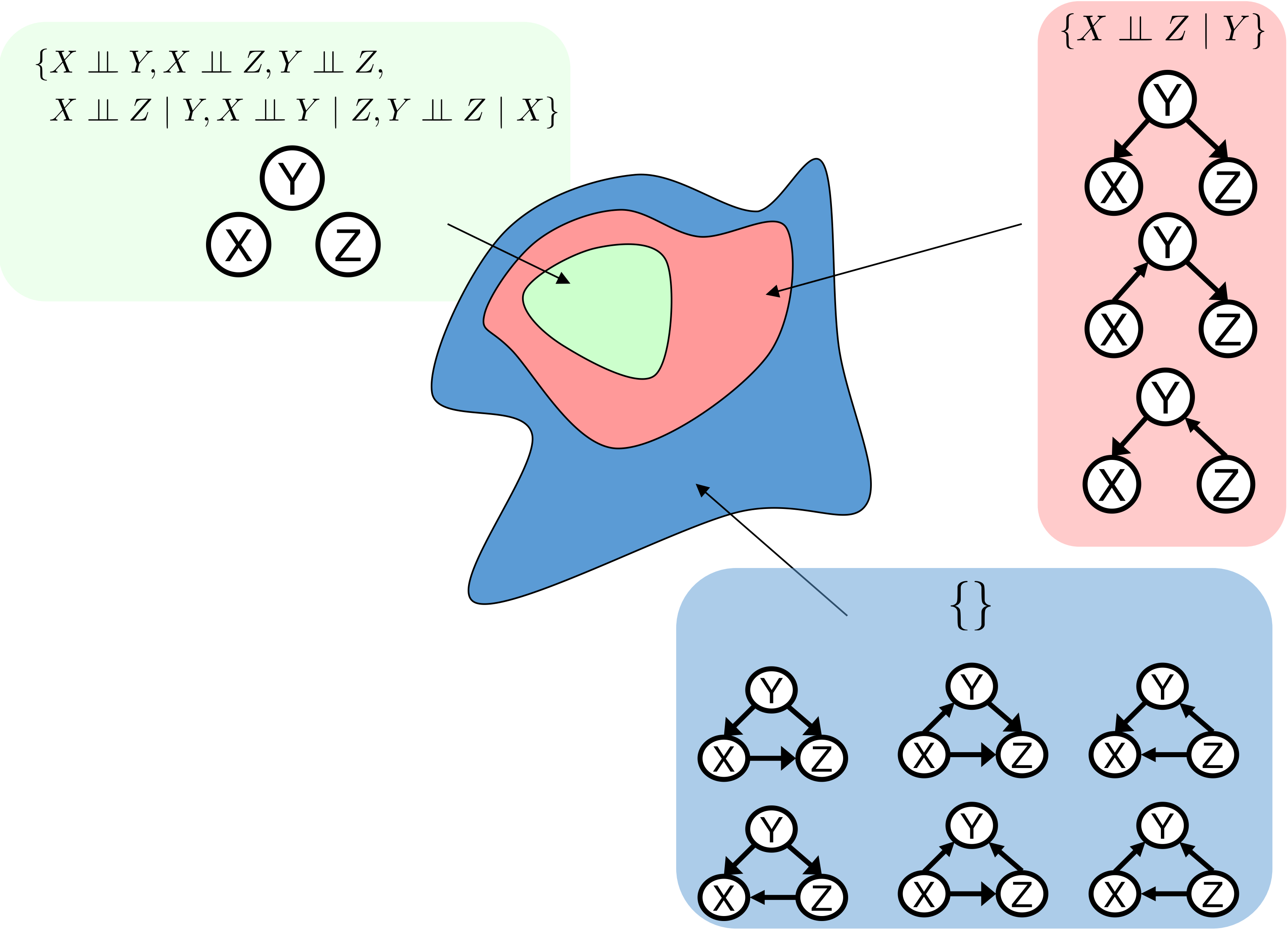
- Given a Bayes net structure, can run d-separation algorithm to build a complete list of conditional independences that are necessarily true of the form

$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, \ldots, X_{k_n}\}$$

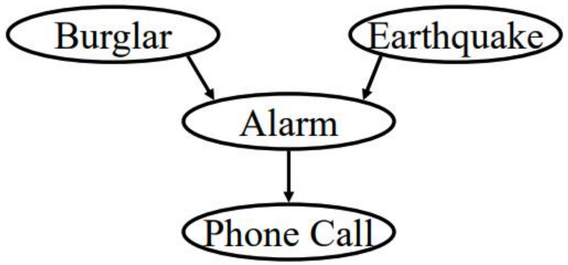
- This list determines **the set of probability distributions** that can be represented

Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded
- **The graph structure guarantees certain (conditional) independences**
- (There might be more independence)
- Adding arcs increases the set of distributions, but has several costs
- Full conditioning can encode any distribution



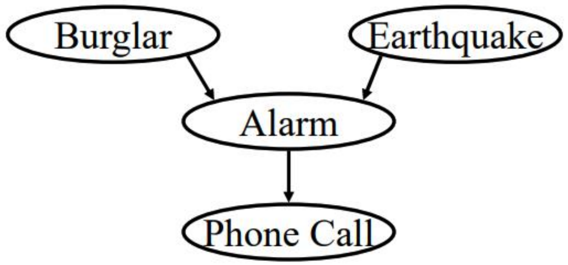
The “Burglar Alarm” example



- Your house has a burglar alarm that is also sometimes triggered by earthquakes

- While you are on vacation, one of your neighbors calls and tells you your home’s burglar alarm is ringing.

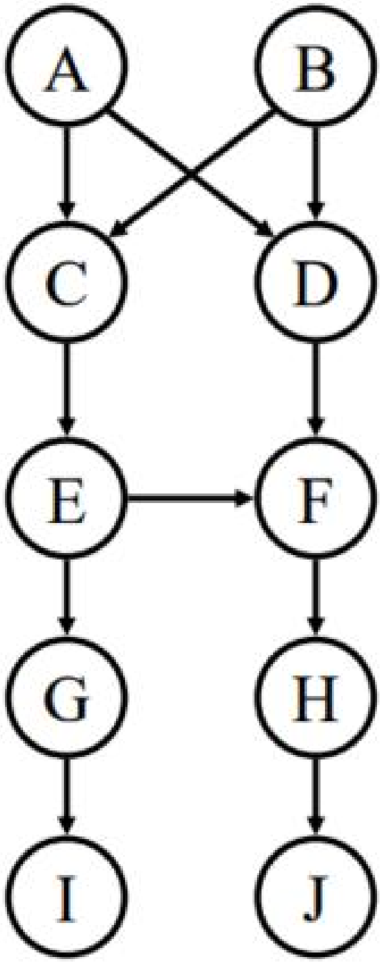
The “Burglar Alarm” example



- But now suppose you learn that there was a medium-sized earthquake in your neighborhood. Probably not a burglar after all.
- Earthquake “explains away” the hypothetical burglar.
- But then it must not be the case that

(Burglar $\perp\!\!\!\perp$ Earthquake | Phone Call), even though (Burglar $\perp\!\!\!\perp$ Earthquake)

More examples



1. $C \perp\!\!\!\perp D$ No
2. $(C \perp\!\!\!\perp D \mid A)$ No
3. $(C \perp\!\!\!\perp D \mid A, B)$ Yes
4. $(C \perp\!\!\!\perp D \mid A, B, J)$ No
5. $(C \perp\!\!\!\perp D \mid A, B, J, E)$ Yes

Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions
- Guaranteed independencies of distributions can be deduced from BN graph structure
- D-separation gives precise conditional independence guarantees from graph alone
- A Bayes' net's joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution