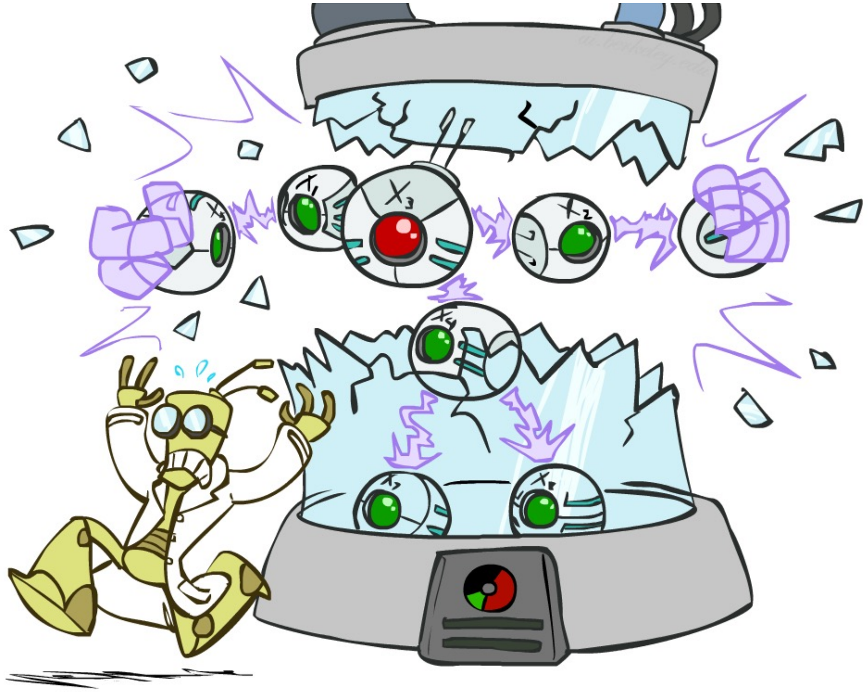
CS 3317: Artificial Intelligence

Bayes Nets: Independence



Instructor: Panpan Cai

[Slides adapted from UC Berkeley CS188]

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:

$$\forall x,y,z : P(x,y|z) = P(x|z)P(y|z) \qquad X \perp\!\!\!\perp Y | Z$$

Independence

- Two variables are *independent* if:

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

  - This says that their joint distribution *factors* into a product of two simpler distributions
  - Another form:

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

  - We write: $X \perp\!\!\!\perp Y$

- Independence is a simplifying *modeling assumption*
  - *Empirical* joint distributions: at best "close" to independent

Conditional Independence

- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.

- (X is conditionally independent of Y) given Z

$$X \perp\!\!\!\perp Y | Z$$

if and only if:

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

or, equivalently, if and only if

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

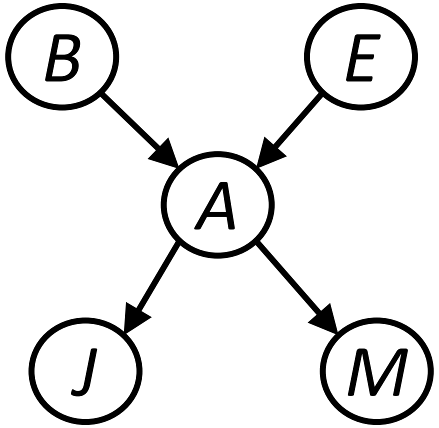
$$P(x|z,y) = \frac{P(x,z,y)}{P(z,y)}$$

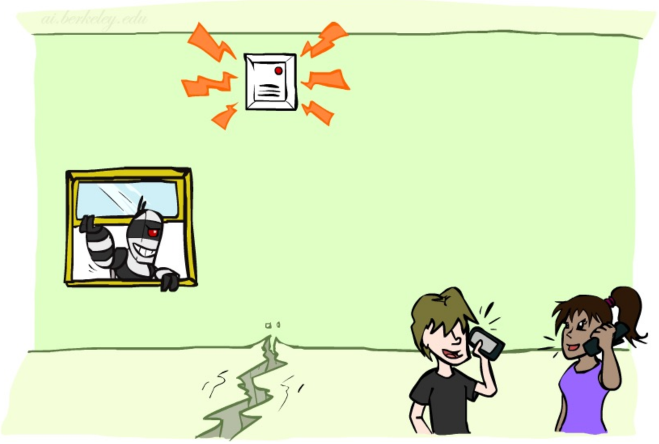
$$= \frac{P(x,y|z)P(z)}{P(y|z)P(z)}$$

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

Bayes Nets

- A Bayes net is a graphical model efficiently encoding an inference problem





- Local CPTs encode direct influences among variable

B	P(B)
+b	0.001
-b	0.999

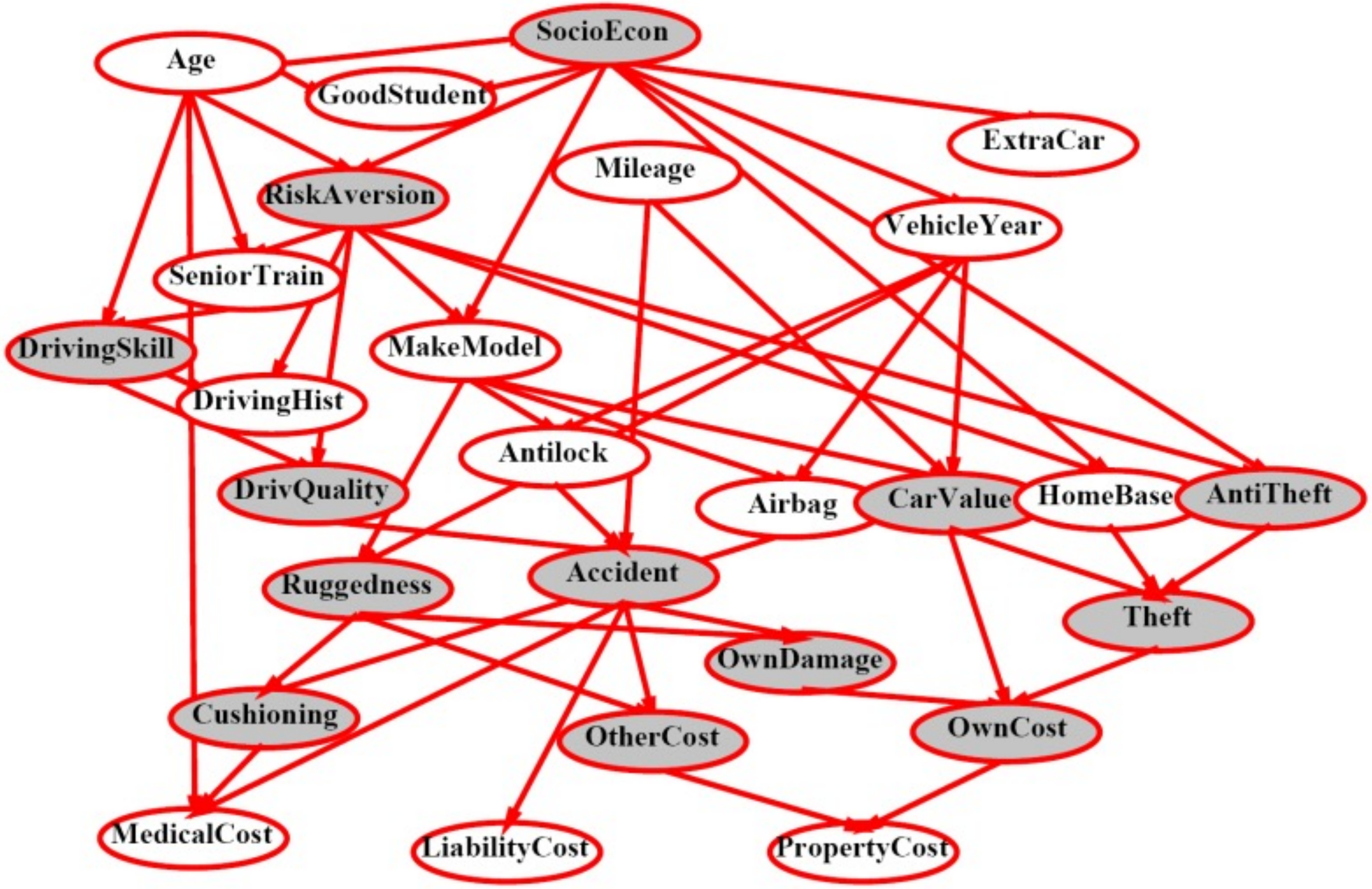
E	P(E)
+e	0.002
-e	0.998

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

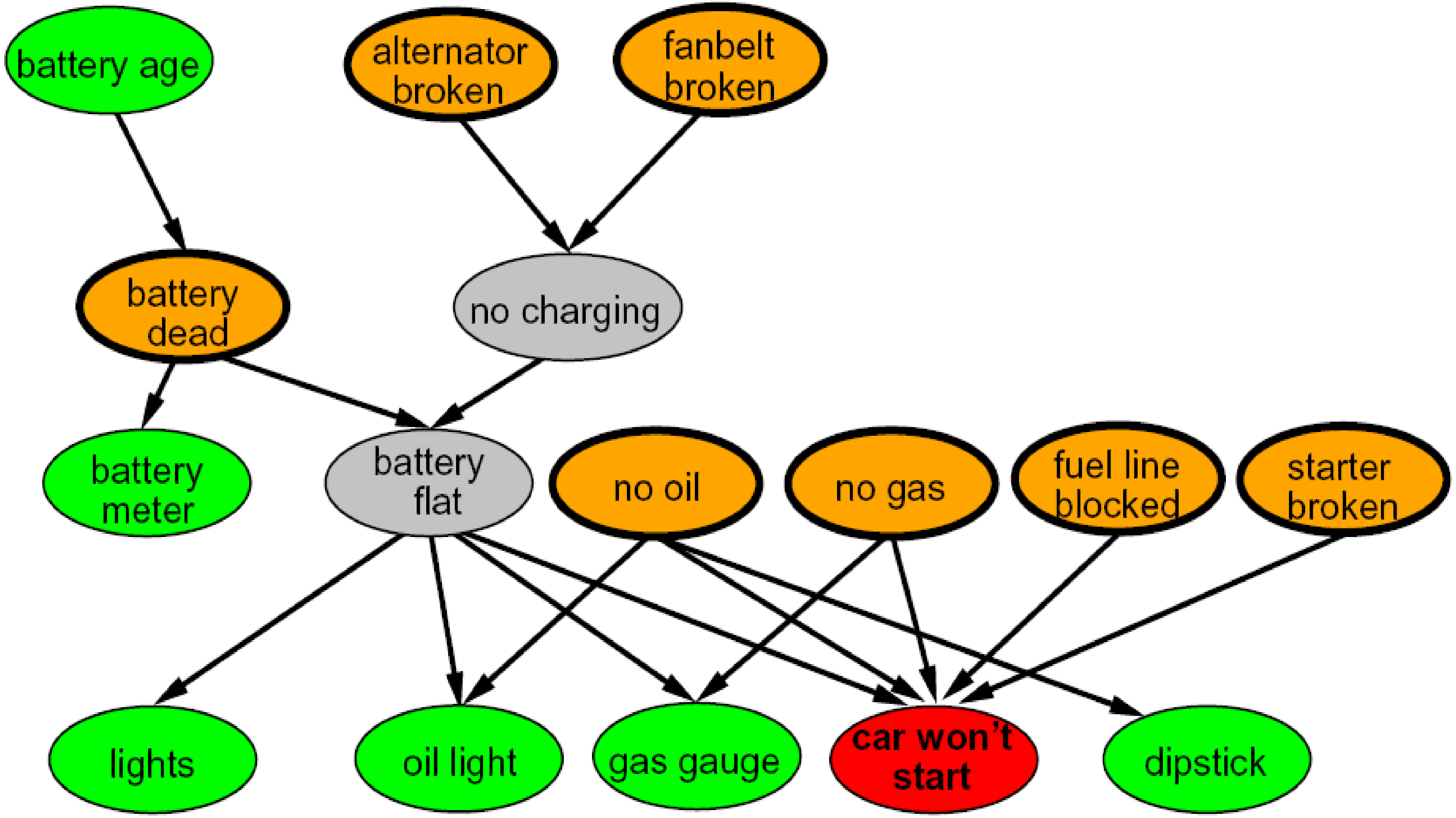
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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 Bayes Net: Insurance



Example Bayes Net: Car



Bayes Net Advantages

- How big is a joint distribution over N Boolean variables?

  2^N

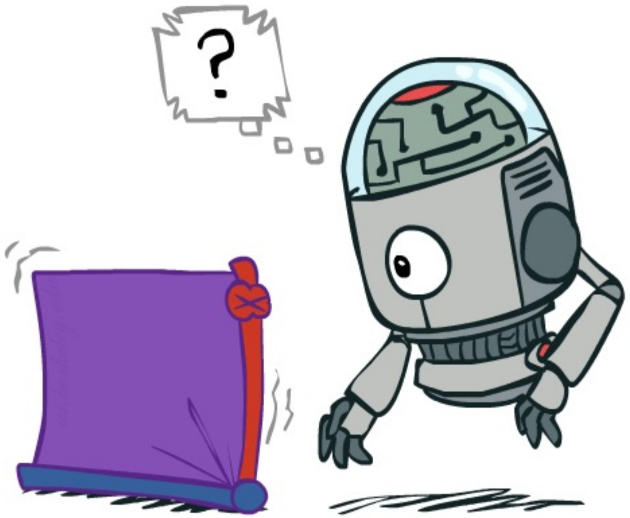
- How big is an N-node net if nodes have up to k parents?

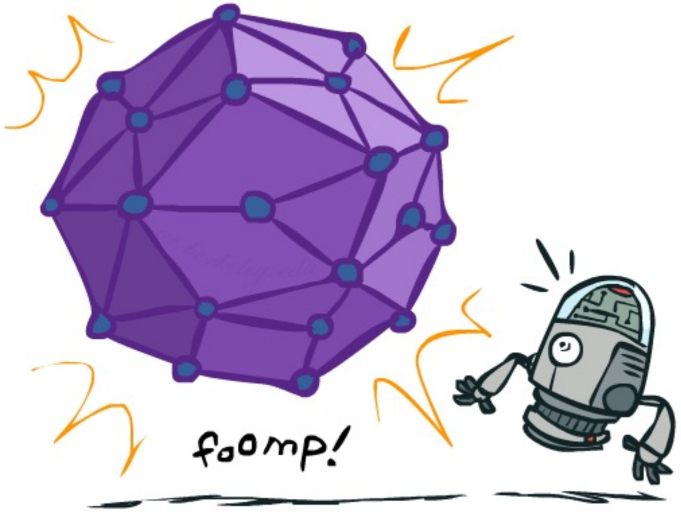
  $O(N * 2^{k+1})$

- Both give you the power to calculate

  $$P(X_1, X_2, \ldots X_n)$$

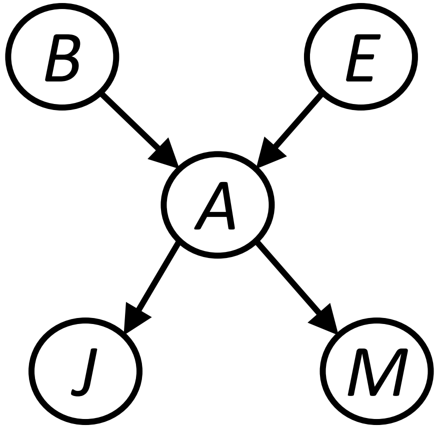
- BNs: Huge space savings!
- Also easier to gather local CPTs
- Also faster to answer queries (last lecture!)

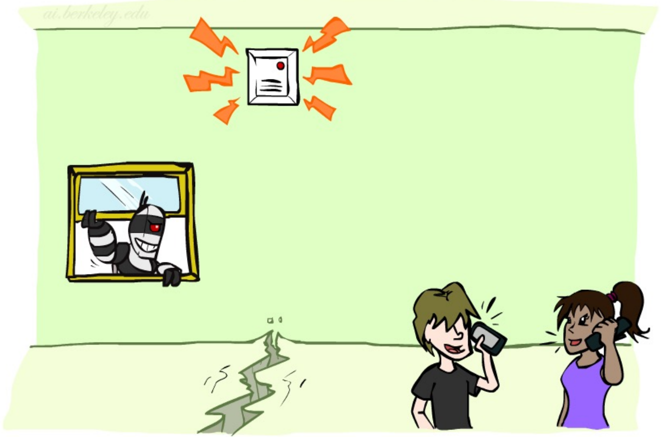




Core Questions in BNs

- A Bayes net is a graphical model efficiently encoding an inference problem





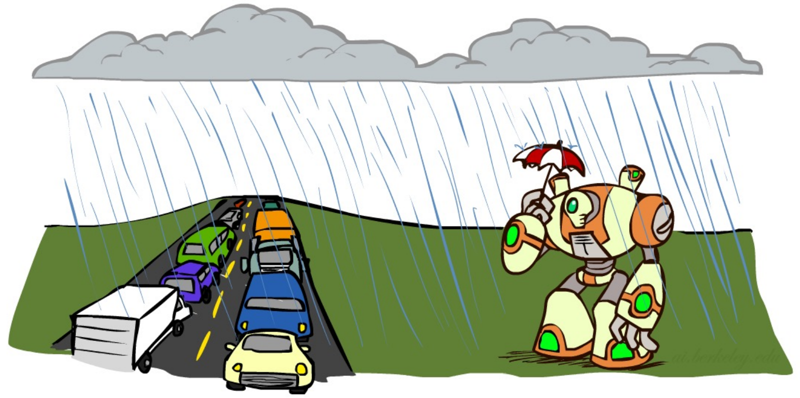
- Questions we can ask:
  - **Inference**: given a fixed BN, what is P(X | e)?
    - Sum over products
  - **Modeling**: what BN is most appropriate for a given domain?
    - Tradeoff expressiveness and efficiency
  - **Representation**: given a BN graph, what kinds of distributions can it encode?
    - Not all BN’s can represent all joint distributions,
    - as BN’s topology additionally assume conditional independences

Bayes Nets

- ✓ Representation
- ✓ Probabilistic Inference
  - Conditional Independence
  - Sampling
  - Learning Bayes' Nets from Data

Conditional independence in BNs

- Consider the traffic domain:
  - Traffic: there is heavy traffic
  - Rain: there is rain
  - Umbrella: someone holding umbrella



- Chain rule decomposition:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = \\ P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

- With conditional independence:

$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = \\ P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$

- This becomes a Bayes Net!

Conditional independence in BNs

- Bayes net probabilities:

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

Chain rule (valid for all distributions)

$$P(x_1, x_2, \dots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \dots x_{i-1})$$

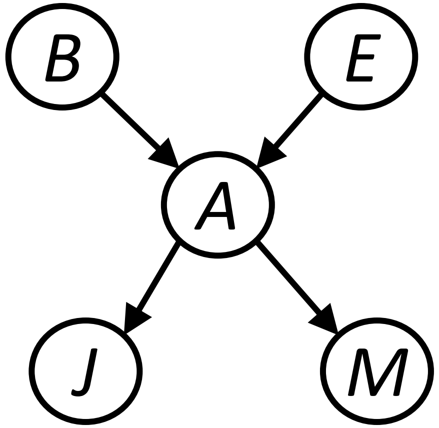
- They must be equivalent! So we must have:

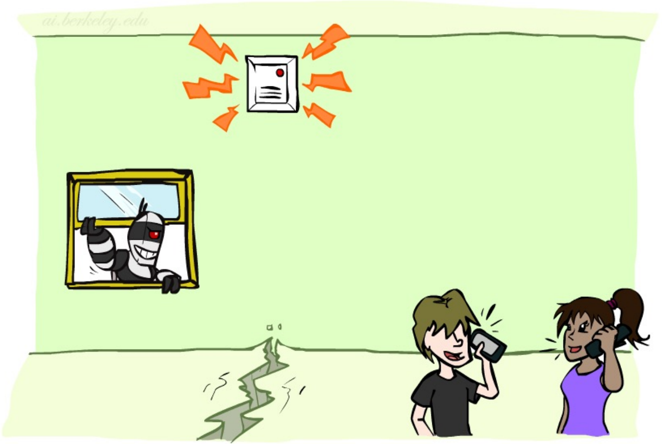
$$P(x_i | x_1, \dots x_{i-1}) = P(x_i | \mathit{parents}(X_i))$$

  - X_i is independent of all non-parent variables coming before it, given its parents

Example: Alarm Network

- Consider the alarm domain:





- Chain rule decomposition:

$$P(b,e,a,j,m) =$$
$$P(b)P(e|b)P(a|b,e)P(j|b,e,a)P(m|b,e,a,j)$$

- Bayes Net:

$$P(b,e,a,j,m) =$$
$$P(b)P(e)P(a|b,e)P(j|a)P(m|a)$$

- Independences to spot out?

Bayes Nets: Assumptions

- The following conditional independence assumptions are made immediately, when given a BN graph:

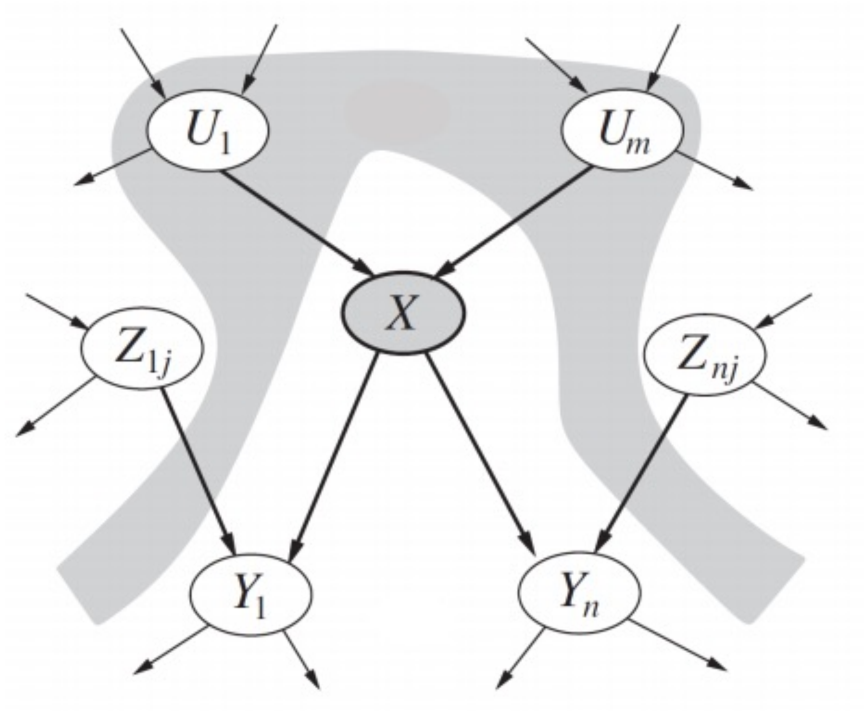
$$P(x_i|x_1 \cdots x_{i-1}) = P(x_i|parents(X_i))$$

- Beyond above "chain rule → Bayes net" conditional independence assumptions
  - Often more conditional independences
  - They can be read from the graph

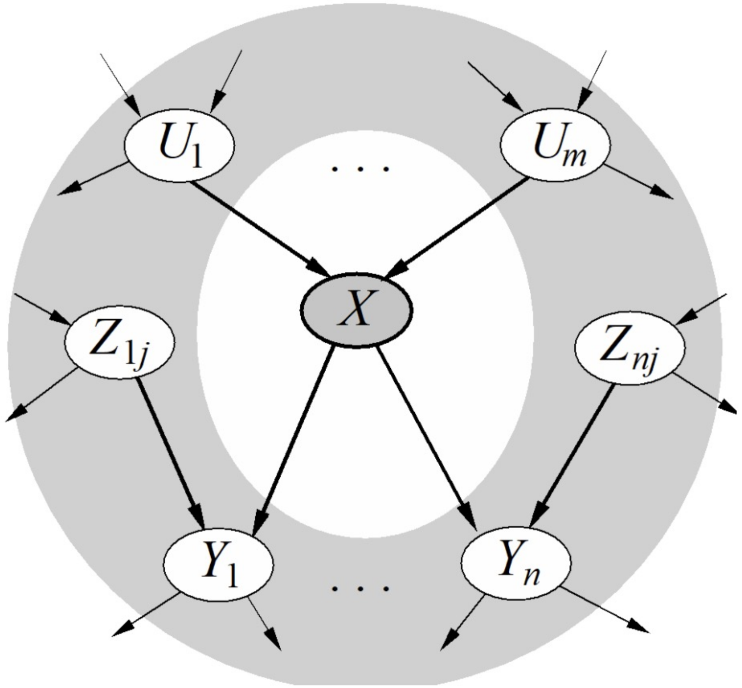
- Important for modeling: examine the expressiveness and improve efficiency
  - Known to have strong correlation -> fix the graph!
  - Known to be conditionally independent -> simplify the graph!

Example: More Conditional Independences

- Each node, given its parents, is conditionally independent of all its non-descendants in the graph

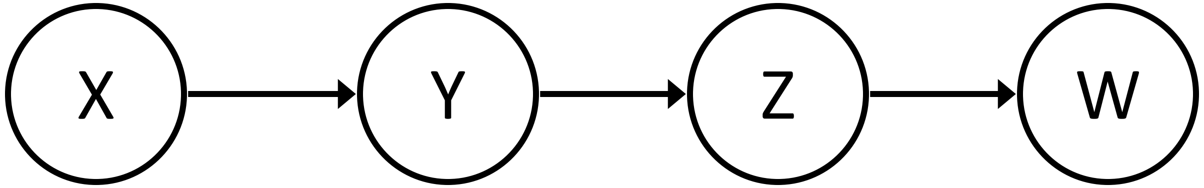


Each node, given its MarkovBlanket, is conditionally independent of all other nodes in the graph



MarkovBlanket refers to the parents, children, and children's other parents.

Example



- Conditional independence assumptions directly from simplifications in chain rule:

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

$$X \perp\!\!\!\perp Z | Y$$

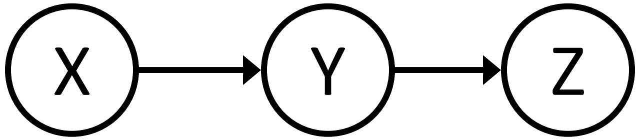
$$P(x,y,z,w) = P(x)P(y|x)P(z|y)P(w|z)$$

$$W \perp\!\!\!\perp \{X,Y\} | Z$$

- Additional implied conditional independence assumptions?

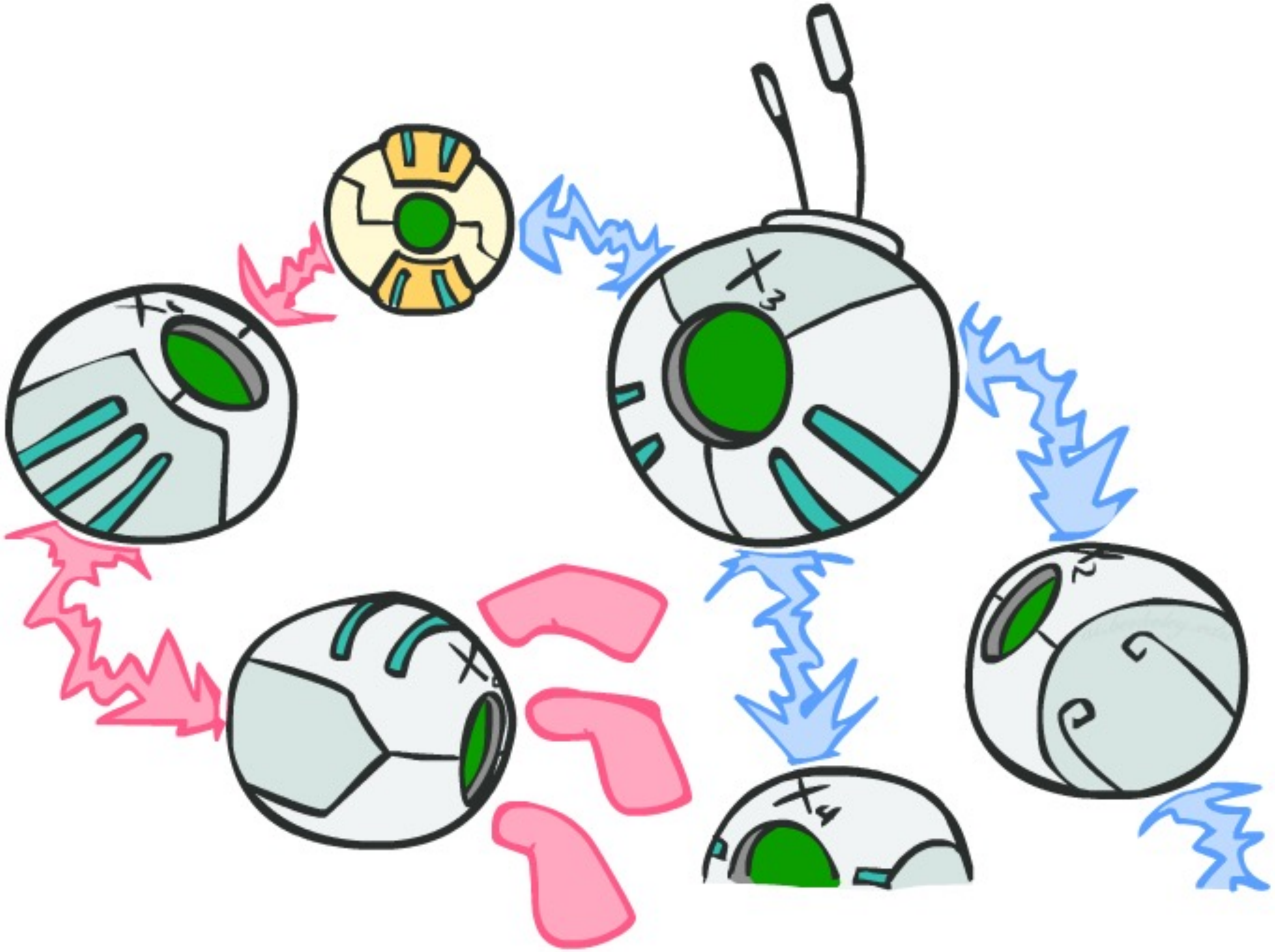
Independence Questions

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



  - Question: are X and Z independent?
    - Answer: no.
      - Example: low pressure causes rain, which causes traffic.
    - X can influence Z, Z can influence X (via Y)
    - Note: they *could* be independent: how?

D-separation: Outline

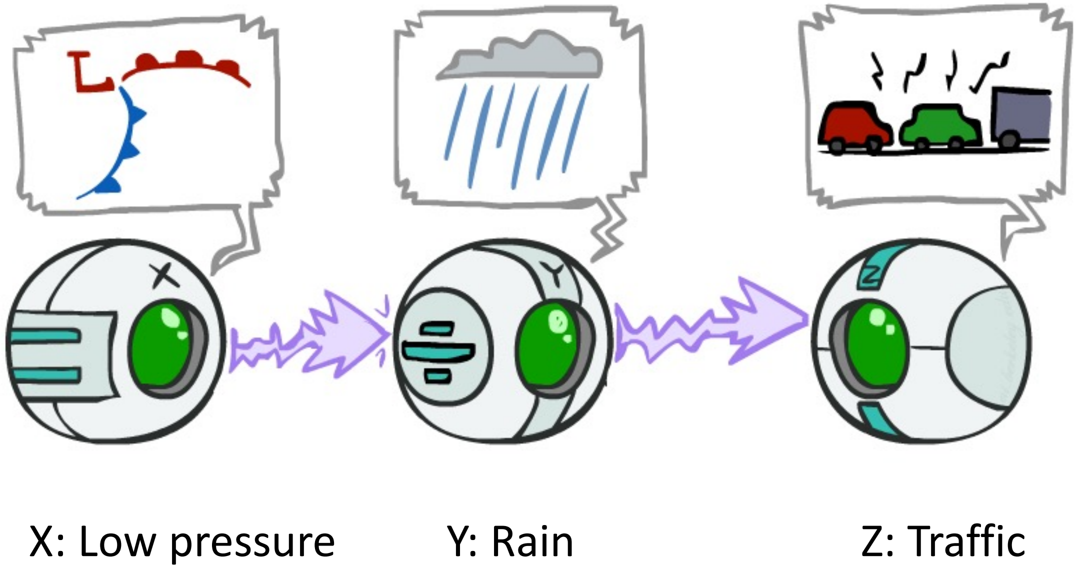


D-separation: Outline

- Study independence properties for triples
  - Why triples?
- Analyze complex cases in terms of local triples
- Answer independence questions – the D-separation algorithm

Causal Chains

- This configuration is a “causal chain”



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

- 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:
    - Low pressure causes rain causes traffic, high pressure causes no rain causes no traffic
    - 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: Low pressure Y: Rain Z: Traffic

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

- Guaranteed X independent of Z given Y?

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

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

$$= P(z|y)$$

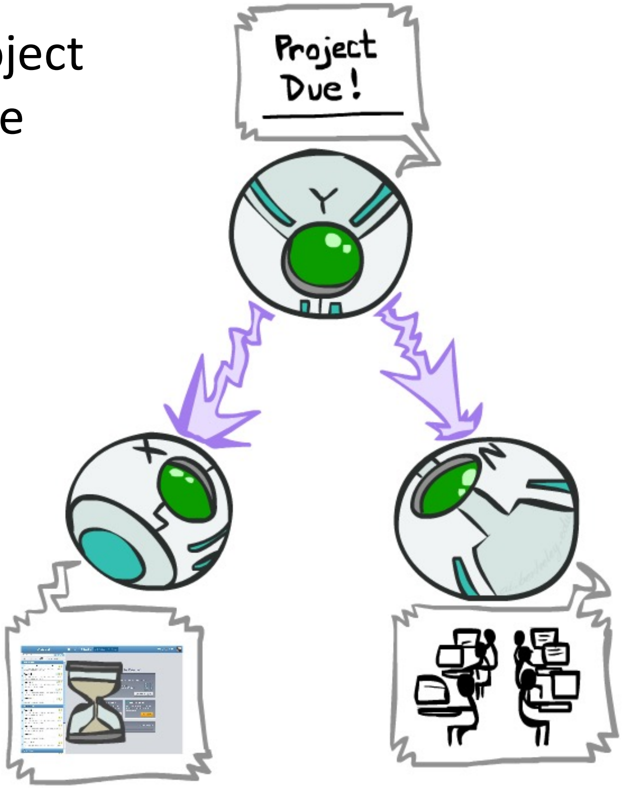
Yes!

- Evidence along the chain "blocks" the influence

Common Causes

- This configuration is a "common cause"

Y: Project due



X: Forums busy

Z: Lab full

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

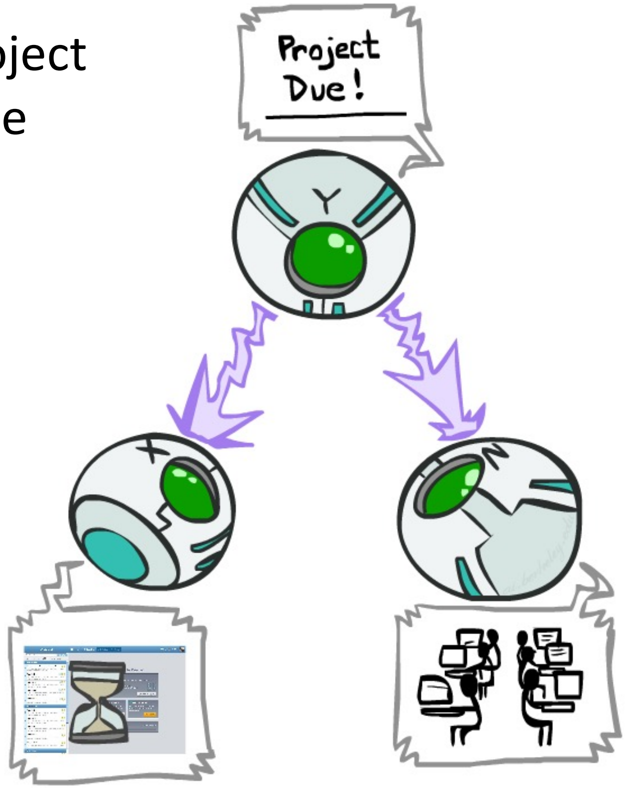
- Guaranteed X independent of Z ?
- *No!*
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
  - Example:
    - Project due causes both forums busy and lab full
    - In numbers:

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: Project due



X: Forums busy

Z: Lab full

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

- Guaranteed X and Z independent given Y?

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

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

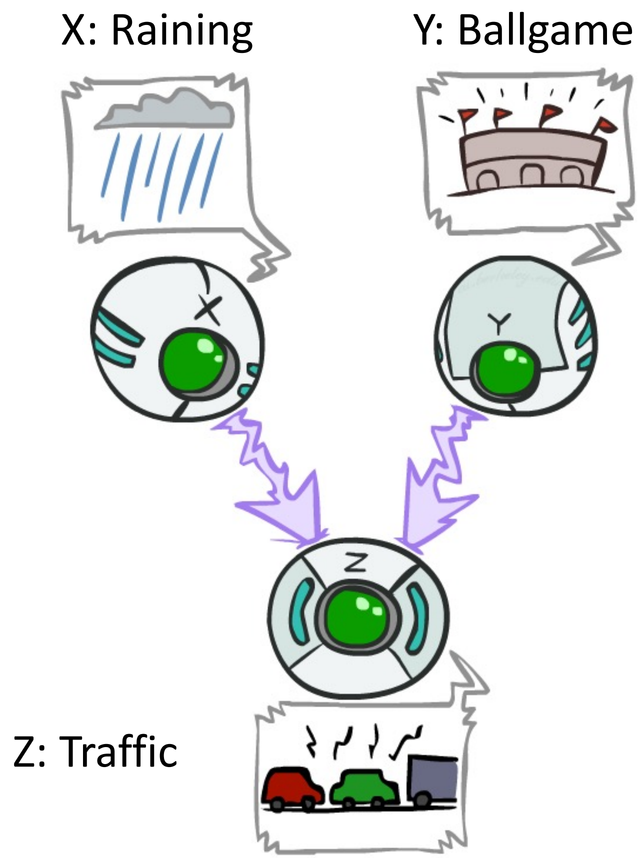
$$= P(z|y)$$

Yes!

- Observing the cause blocks influence between effects.

Common Effect

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

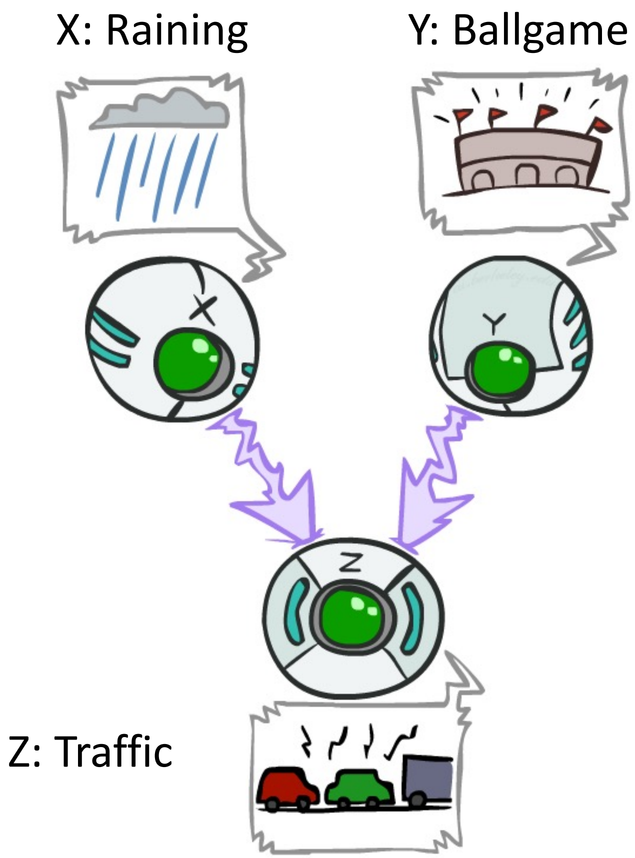


- Are X and Y independent?
  - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
- Proof:

$$P(x, y) = \sum P(x, y, z)$$

Common Effect

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



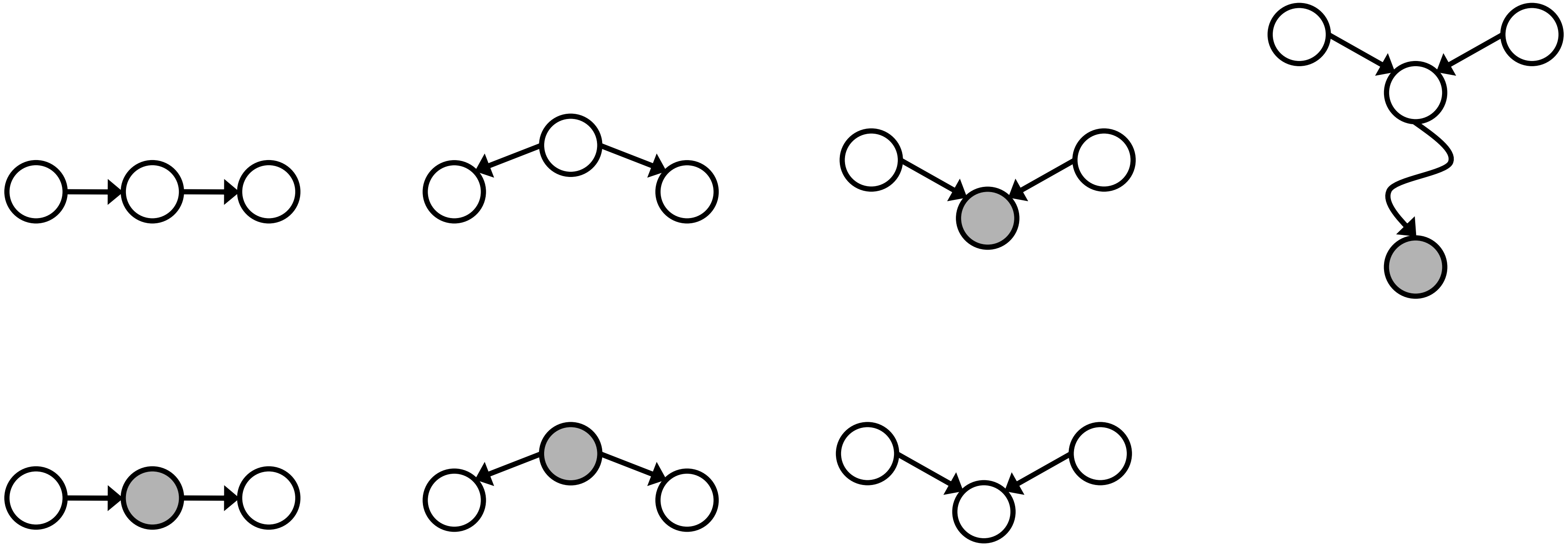
- Are X and Y independent?
  - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
  - (Proved previously)
- Are X and Y independent given Z?
  - *No*: seeing traffic puts the rain and the ballgame in competition as explanation.
- This is reverse of the previous cases
  - Observing an effect activates influence between possible causes.

Putting Together

Causal Chain | Common Cause | Common Effect

Active Triples

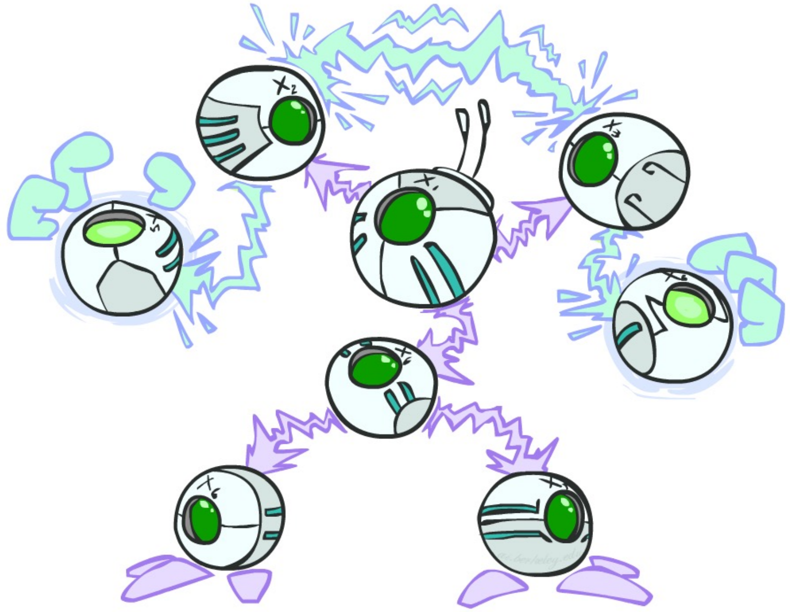
Inactive Triples



The General Case

The General Case

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

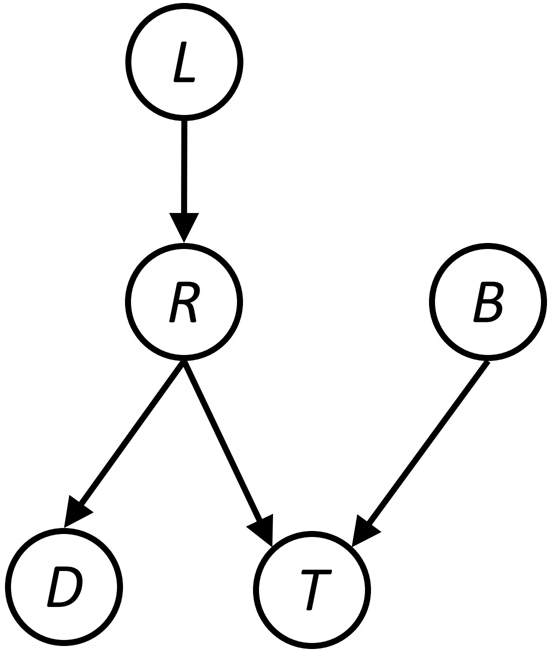


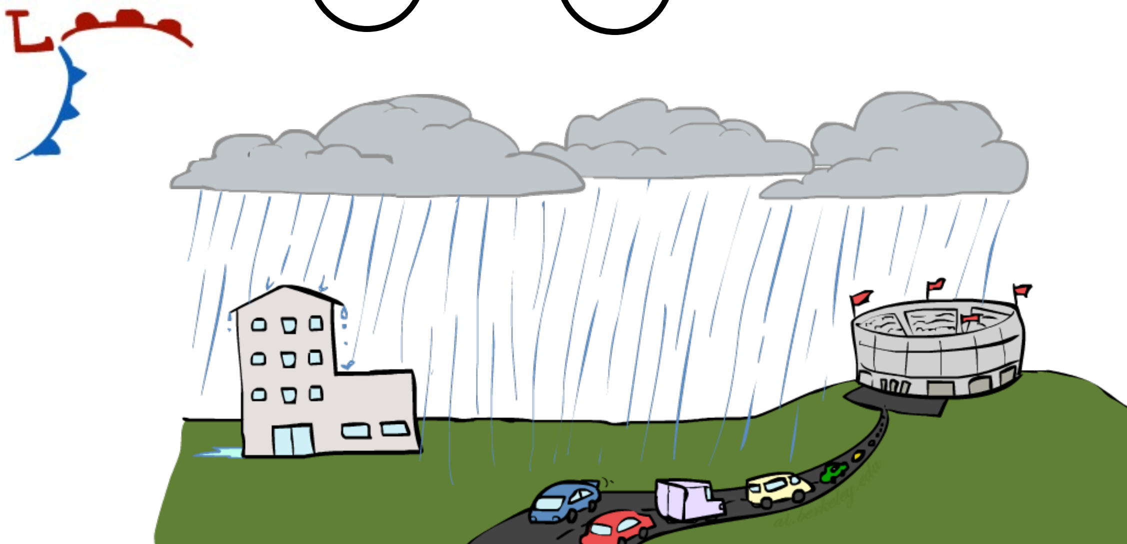
Reachability

- Recipe:
  1. Shade evidence nodes
  2. Look for active paths in the resulting graph

- Attempt 1: if all *undirected* paths between two nodes are blocked by a shaded node, they are conditionally independent

- Almost works, but not quite
  - Where does it break?
  - Answer: the v-structure at T doesn't count as a link in a path unless "active"



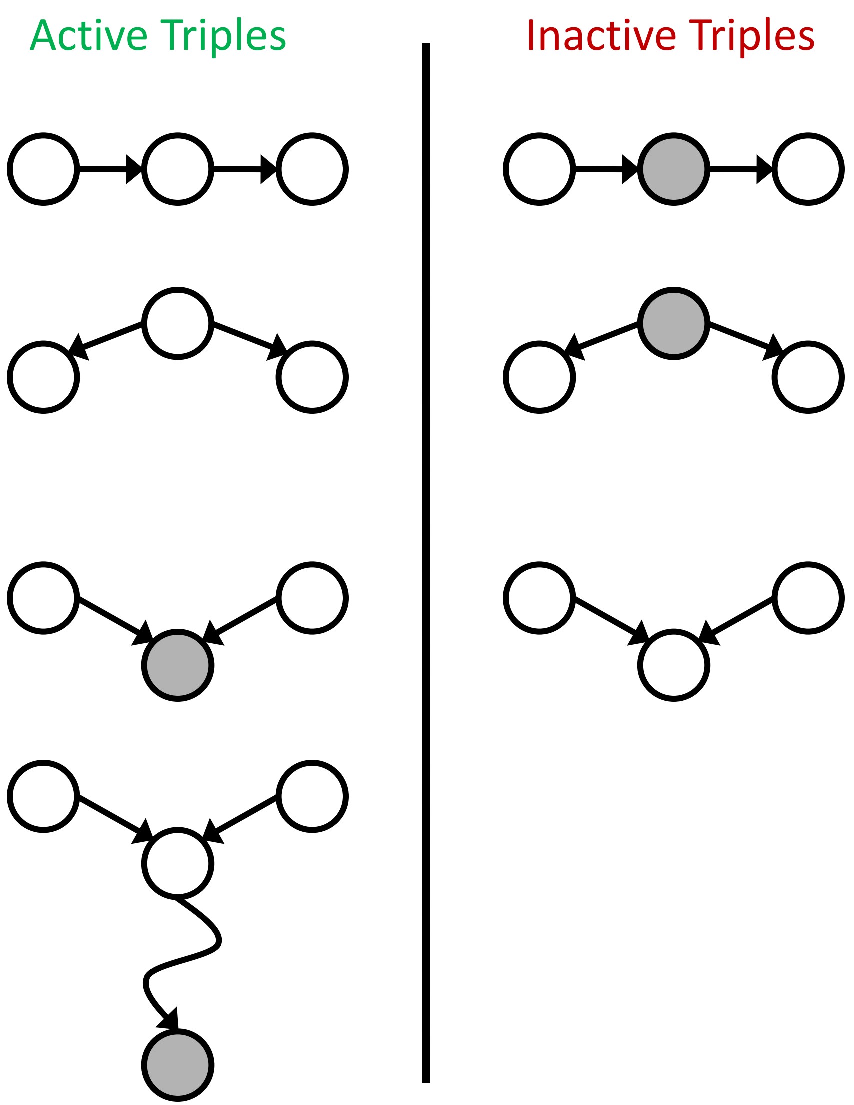


Active / Inactive Paths

- 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 -> B -> C where B is unobserved (either direction)
  - Common cause A <- B -> C where B is unobserved
  - Common effect (aka v-structure)
    A -> B <- C where B *or one of its descendants* is observed

- All it takes to block a path is a single inactive segment
- All it takes to break independence is a single active path



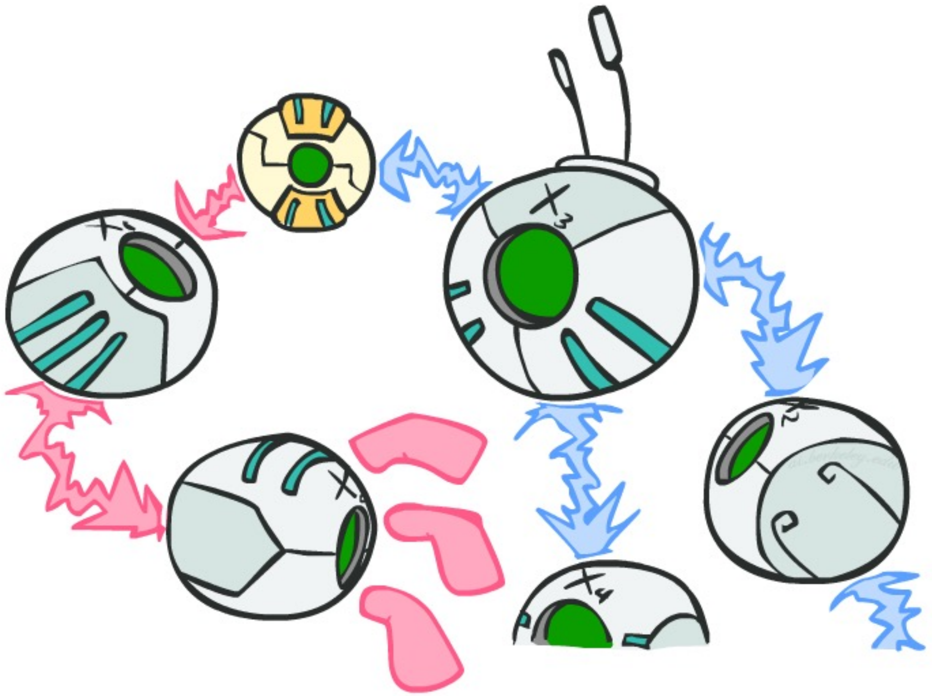
D-Separation

- 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

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

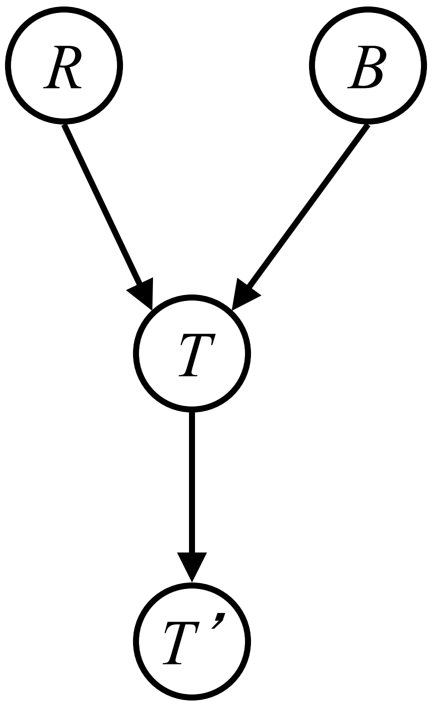


Example

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

$R \perp\!\!\!\perp B | T$

$R \perp\!\!\!\perp B | T'$



Example

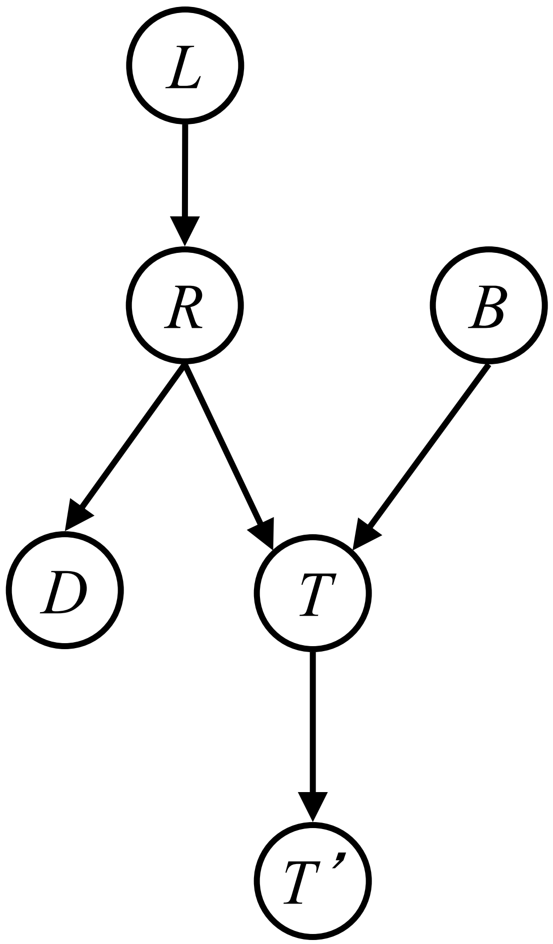
$L \perp\!\!\!\perp T' | T$ *Yes*

$L \perp\!\!\!\perp B$ *Yes*

$L \perp\!\!\!\perp B | T$

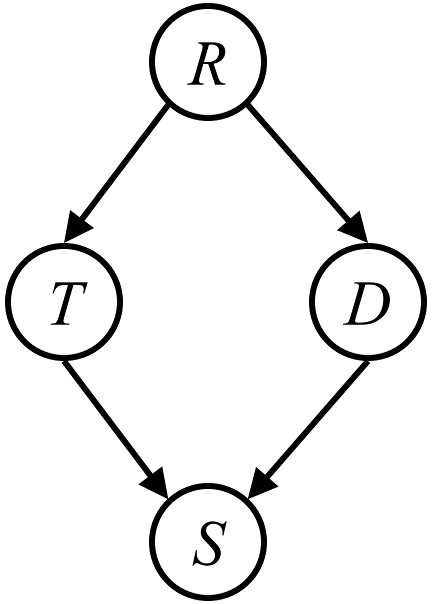
$L \perp\!\!\!\perp B | T'$

$L \perp\!\!\!\perp B | T, R$ *Yes*



Example

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



$$T \perp\!\!\!\perp D$$

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

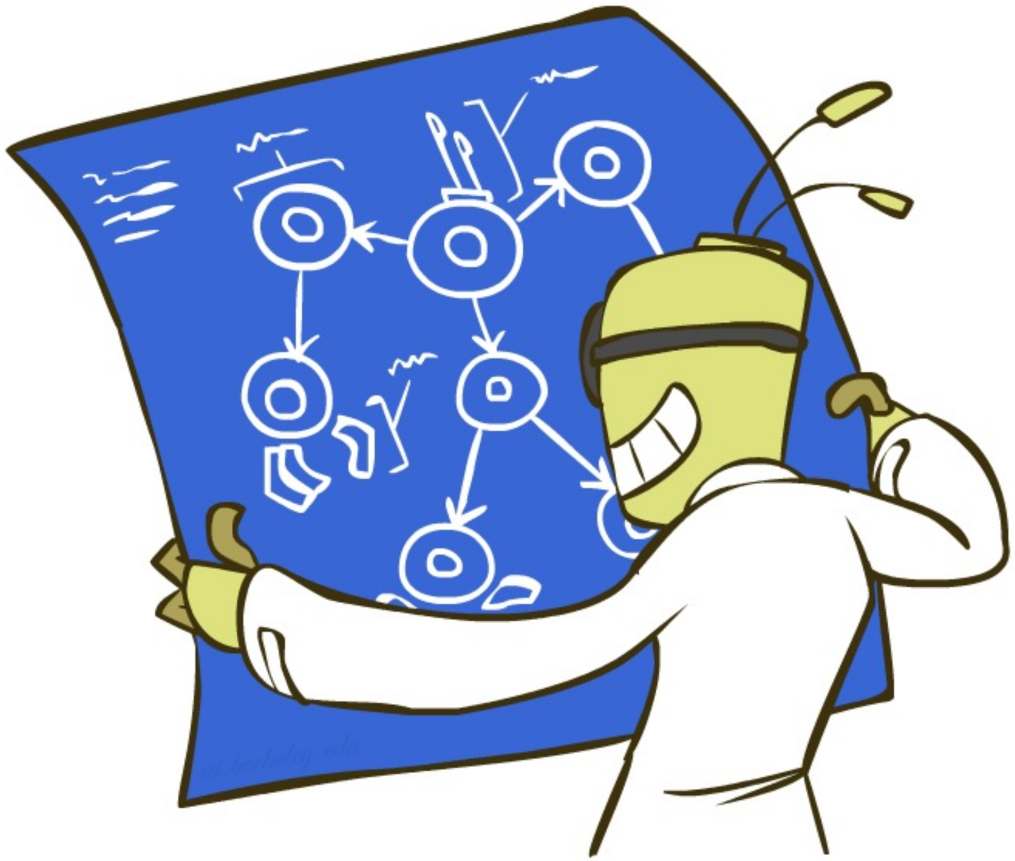
$$T \perp\!\!\!\perp D | R, S$$

"Read" Structure Implications

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

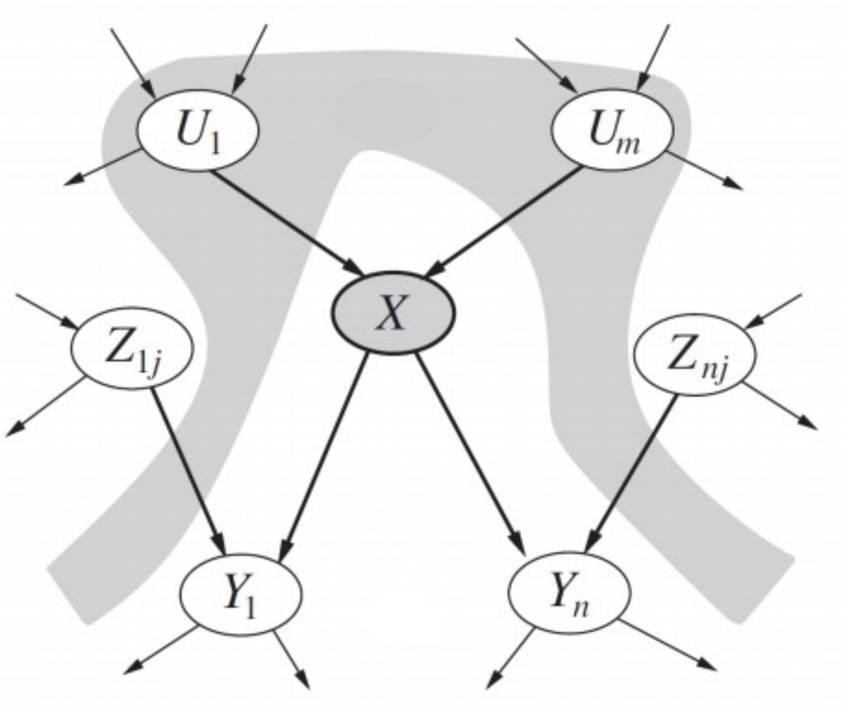
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

- This list determines the set of probability distributions that a BN graph can represent



Explain it!

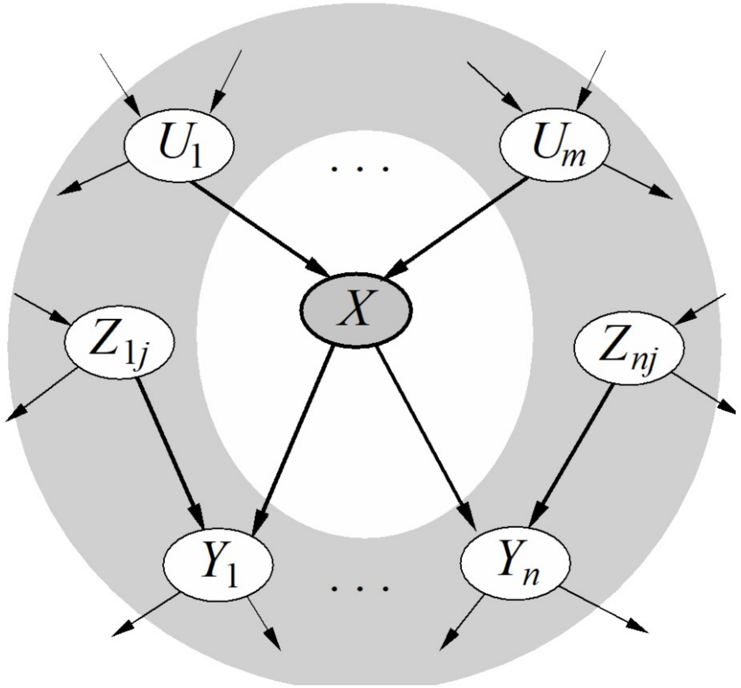
- Each node, given its parents, is conditionally independent of all its non-descendants in the graph



Why does this hold?

Explain it!

Each node, given its MarkovBlanket, is conditionally independent of all other nodes in the graph

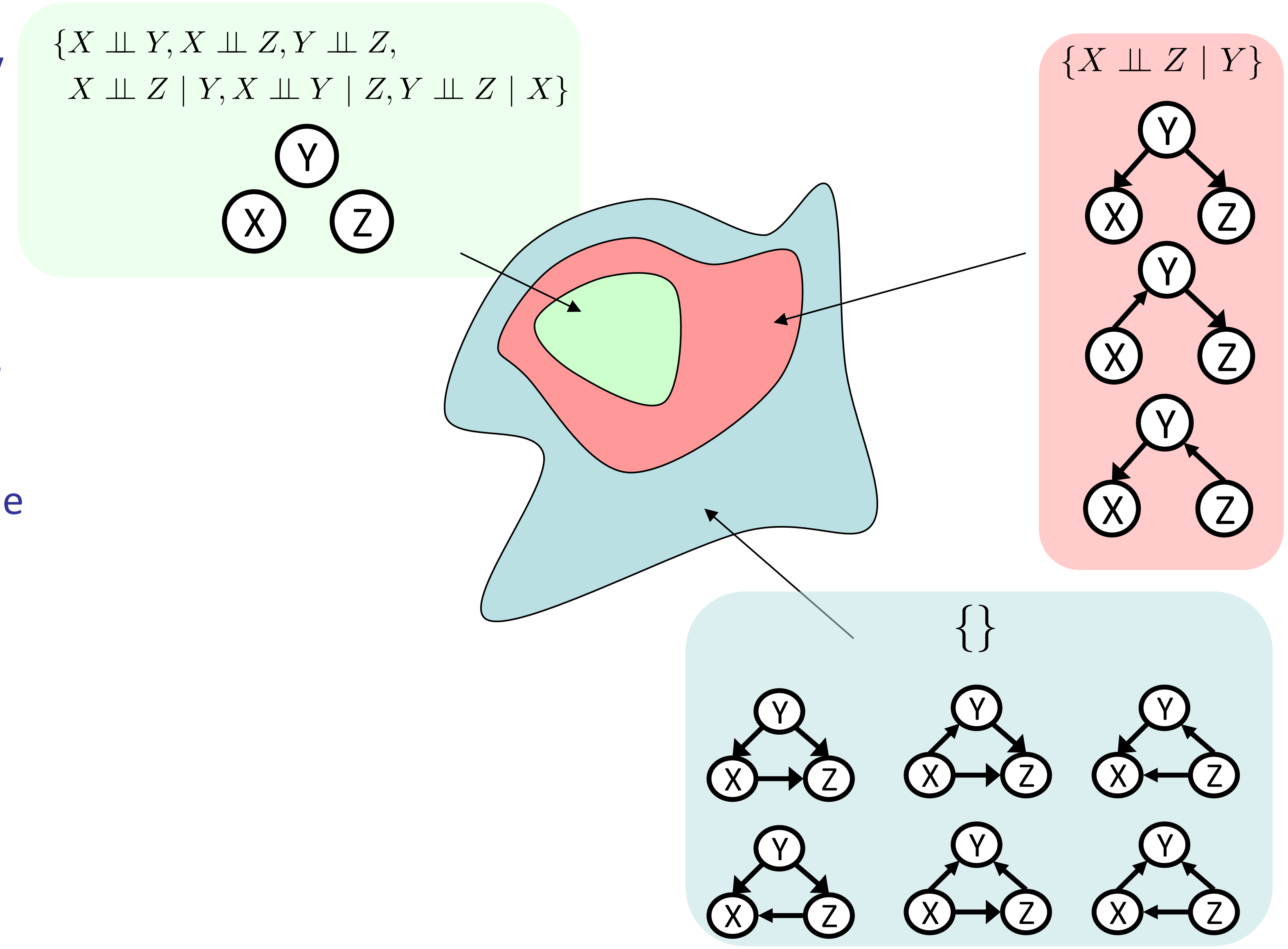


MarkovBlanket refers to the parents, children, and children's other parents.

Why does this hold?

Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded
- Adding arcs increases the set of distributions, but has several costs
- Full conditioning can encode any distribution



Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions (by making use of conditional independences!)
- 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

Bayes Nets

- ✓ Representation
- ✓ Probabilistic Inference
  - Enumeration (exact, exponential complexity)
  - Variable elimination (exact, worst-case exponential complexity, often better)
  - Probabilistic inference is NP-complete
- ✓ Conditional Independences
- Sampling
- Learning from data