CS 188: Artificial Intelligence

Bayes' Nets: Independence

[Many of these slides were originally created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley.]

Announcements

- Midterm
  - Wednesday March 19, 7-9pm
  - Check Ed and Calendar for more midterm logistics/prep sessions, and see exam logistics page near top of course web site for more info.
- HW6
  - Due on Wednesday 3/12/25 at 11:59 PT
- Project 3
  - Due on Friday 3/7/25 at 11:59 PT

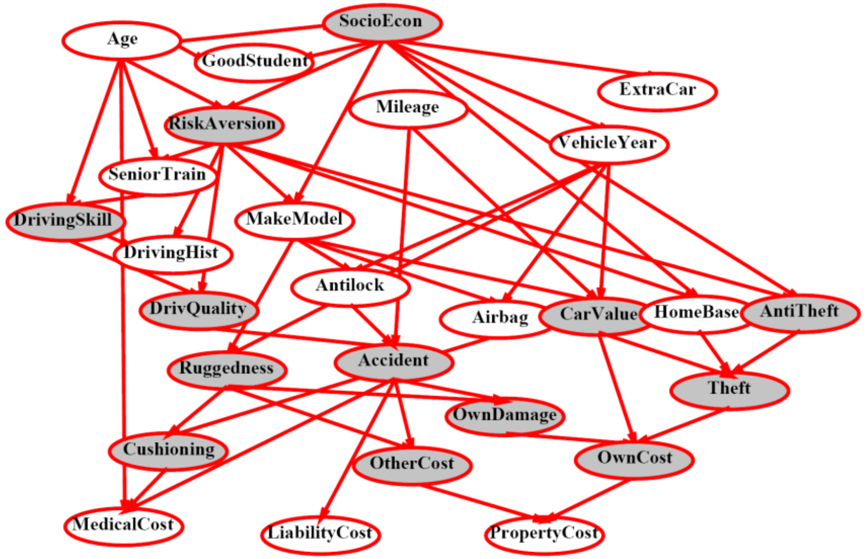
Probability Recap

- 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:

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

Bayes' Nets

- A Bayes' net is an efficient encoding of a probabilistic model of a domain



- Questions we can ask:
  - Modeling: what BN is most appropriate for a given domain?
  - Representation: given a BN graph, what kinds of distributions can it encode?
  - Inference: given a fixed BN, what is P(X | e)?

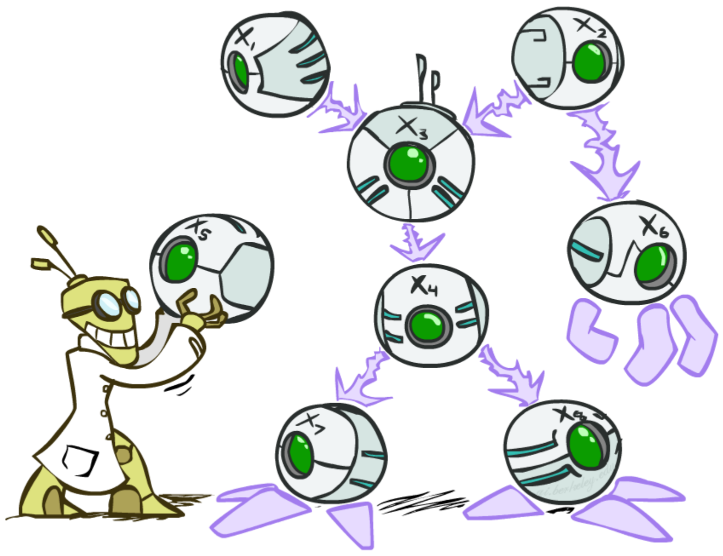
Bayes' Net Semantics

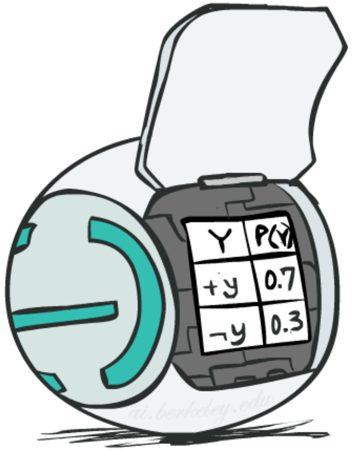
- A directed, acyclic graph, one node per random variable
- A conditional probability table (CPT) for each node
  - A collection of distributions over X, one for each combination of parents' values

$$P(X|a_1 \ldots a_n)$$

- 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:

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



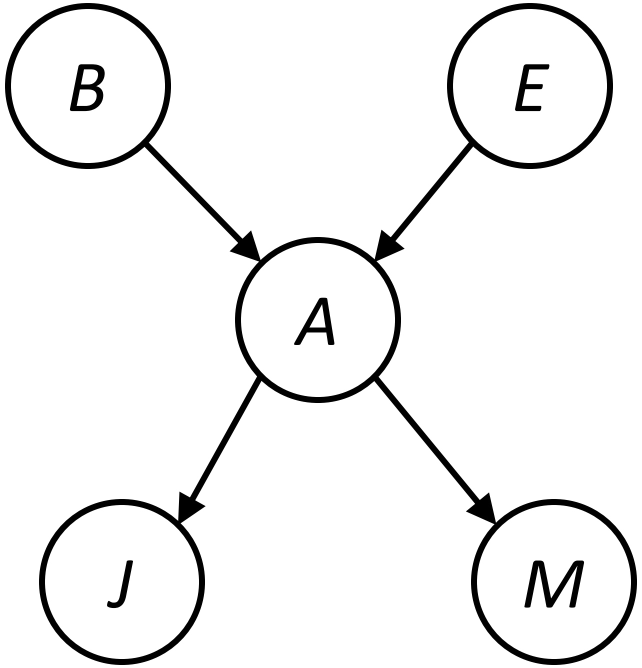


Recall, general case:

$$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots \\ &= \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}) \end{aligned}$$

Example: Alarm Network

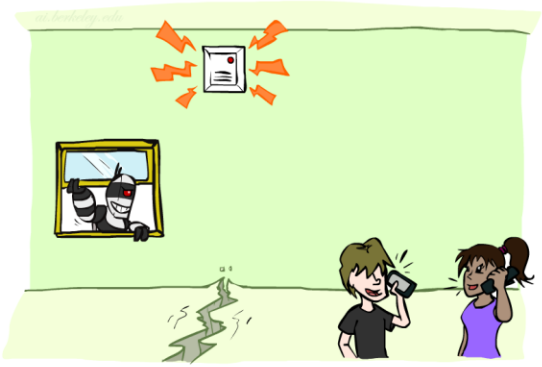
B	P(B)
+b	0.001
-b	0.999



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

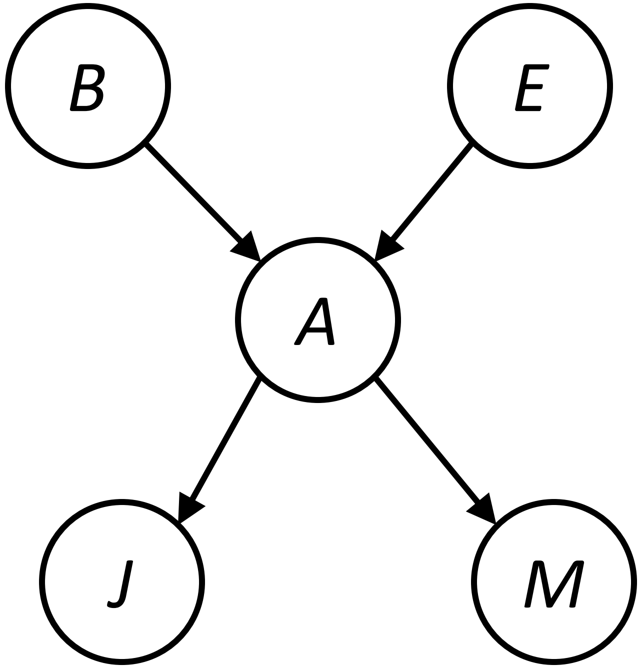


B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

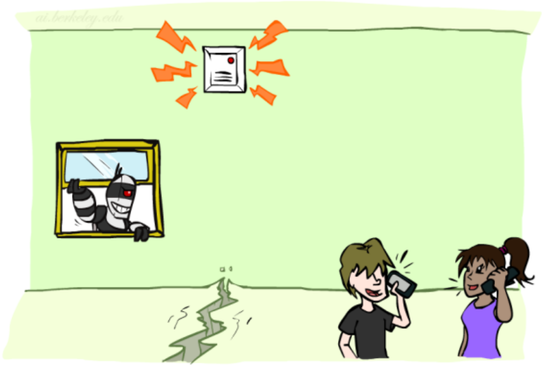
$$P(+b, -e, +a, -j, +m) =$$

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

$$P(+b, -e, +a, -j, +m) =$$
$$P(+b)P(-e)P(+a|+b, -e)P(-j|+a)P(+m|+a) =$$
$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

Size of a Bayes' Net

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

  2^N

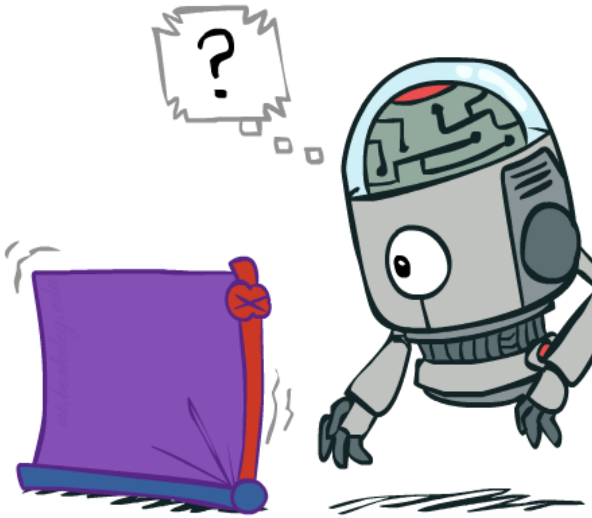
- How big is an N-node net if nodes have up to k parents?

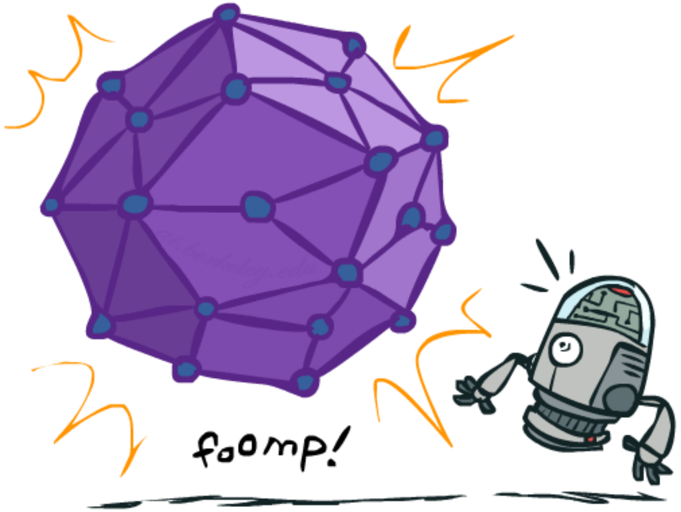
  $O(N * 2^{k+1})$

- Both give you the power to calculate

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

- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also faster to answer queries (coming)





Bayes' Nets

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

Conditional Independence

- X and Y are independent if

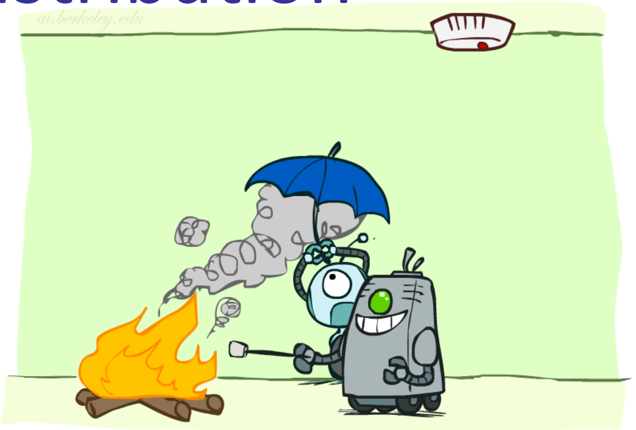
$$\forall x, y \;\; P(x,y) = P(x)P(y) \dashrightarrow X \perp\!\!\!\perp Y$$

- X and Y are conditionally independent given Z

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

- (Conditional) independence is a property of a distribution

- Example: $Alarm \perp\!\!\!\perp Fire | Smoke$

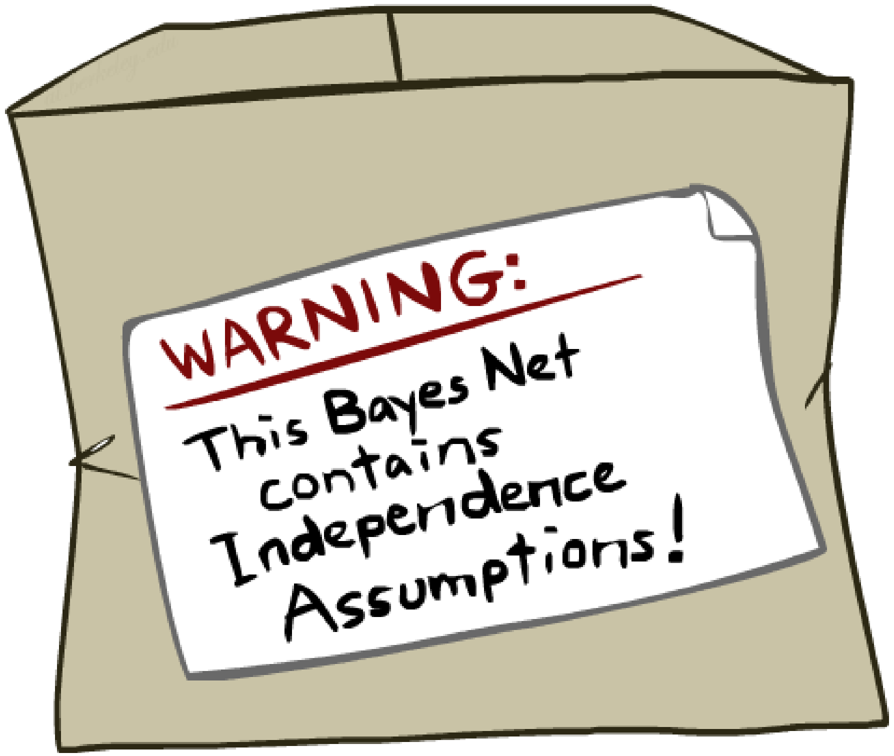


Bayes Nets: Assumptions

- Assumptions we make with Bayes net graph:

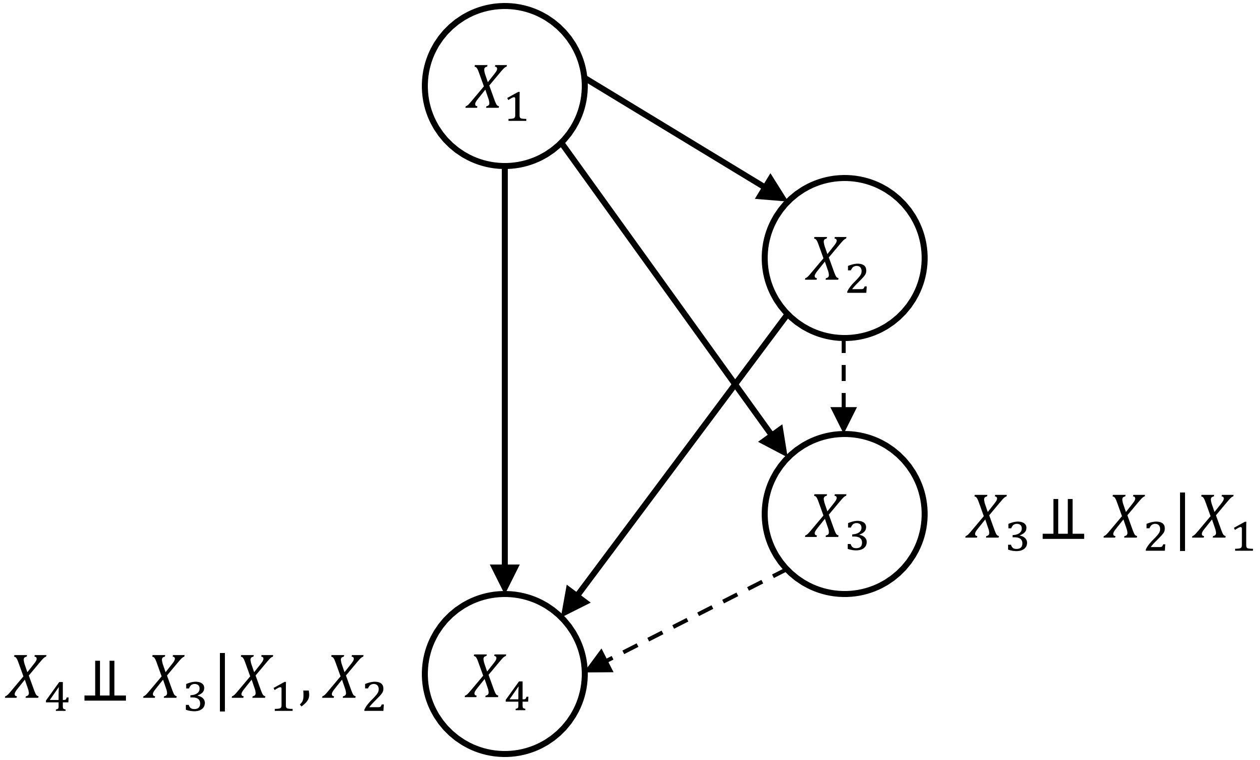
$$P(x_i|x_1 \cdots x_{i-1}) = P(x_i|parents(X_i))$$

- Beyond above "chain rule → Bayes net" conditional independence assumptions
  - Often additional conditional independences
  - Many can also be determined from just the graph
- Important for modeling: understand assumptions made when choosing a Bayes net graph

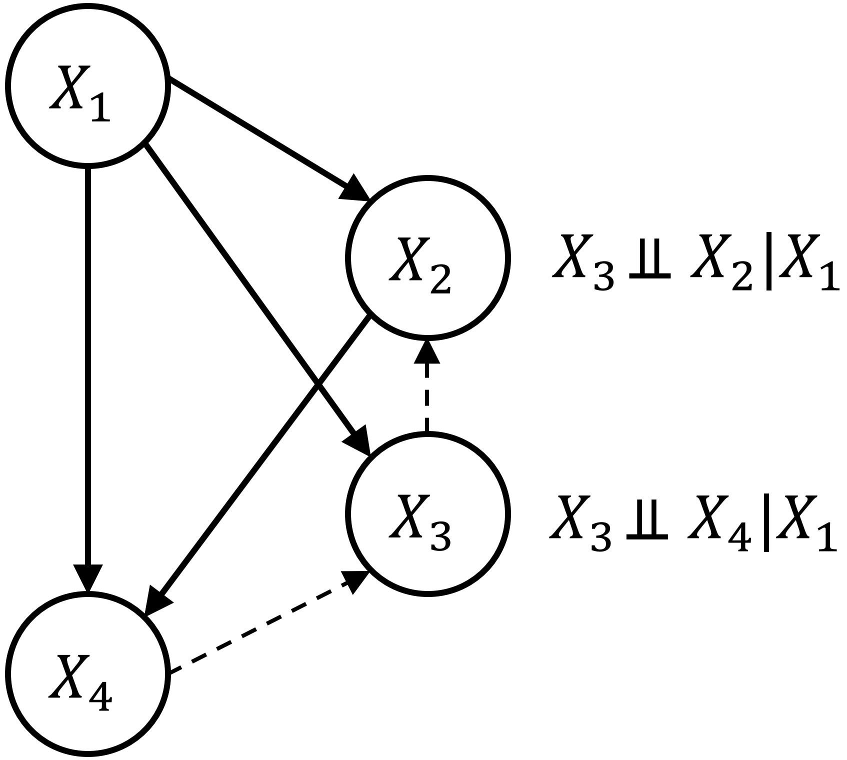


Conditional Independence and the Chain Rule

- In general a node is conditionally independent from all its non-descendents in the graph, given its parents.

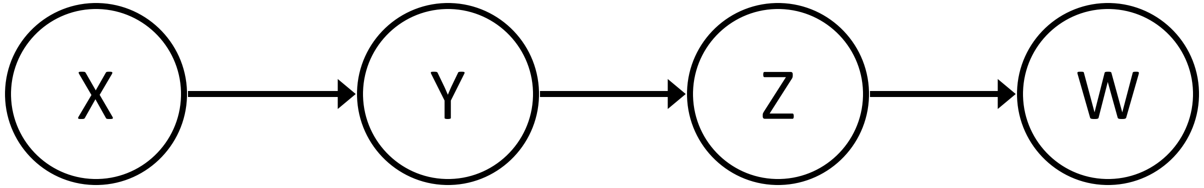


Missing edges from complete graph…



Missing edges under any ordering of nodes

Example

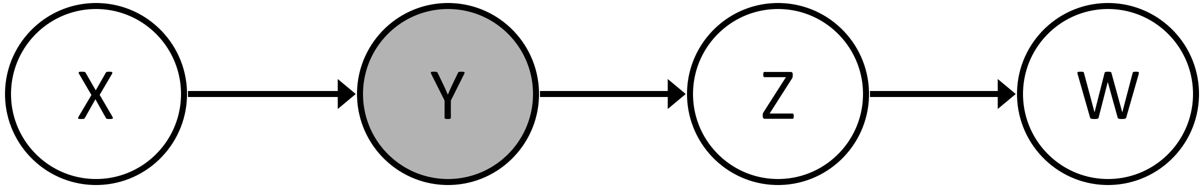


- Conditional independence assumptions directly from simplifications in chain rule:

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

- Additional implied conditional independence assumptions?

Example



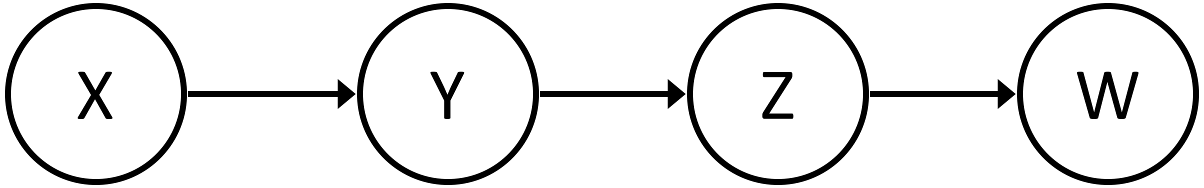
- Conditional independence assumptions directly from simplifications in chain rule:

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

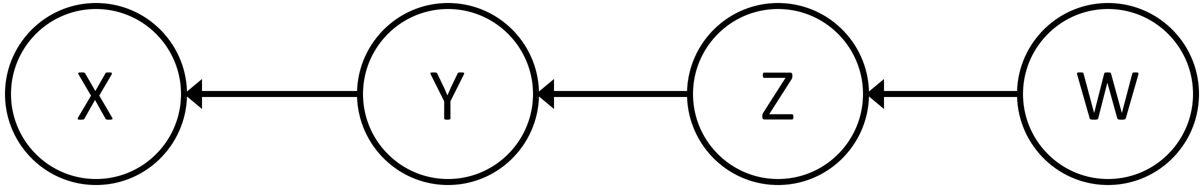
- Additional implied conditional independence assumptions?

$$W \perp\!\!\!\perp X|Y$$

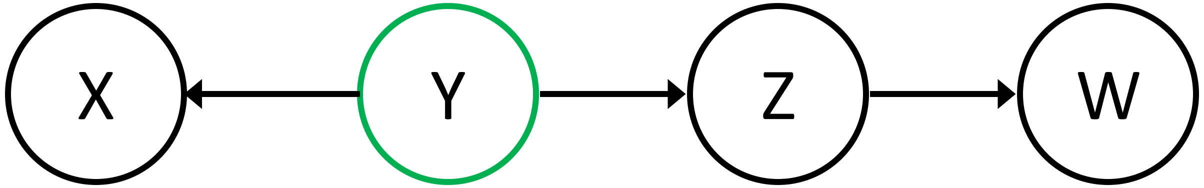
Chain Probabilities



$$\frac{\cancel{P(X)}}{1}\frac{P(Y,X)}{\cancel{P(X)}}\frac{P(Z,Y)}{P(Y)}\frac{P(W,Z)}{P(Z)}$$

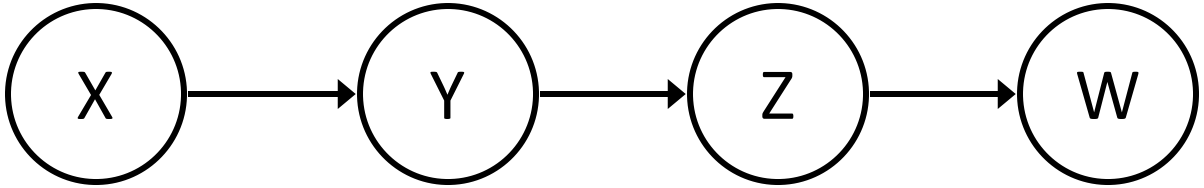


$$\frac{P(Y,X)}{P(Y)}\frac{P(Z,Y)}{P(Z)}\frac{P(W,Z)}{\cancel{P(W)}}\frac{\cancel{P(W)}}{1}$$

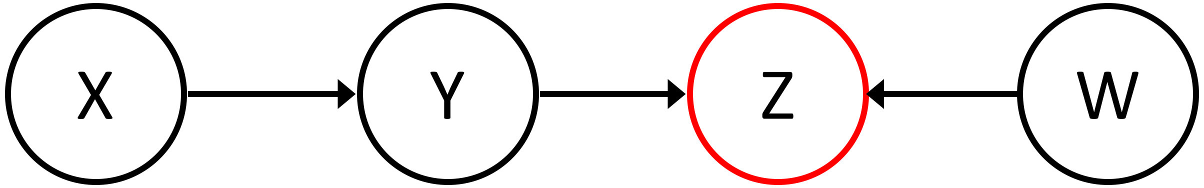


$$\frac{P(Y,X)}{\cancel{P(Y)}}\frac{\cancel{P(Y)}}{1}\frac{P(Z,Y)}{P(Y)}\frac{P(W,Z)}{P(Z)}$$

Chain Probabilities

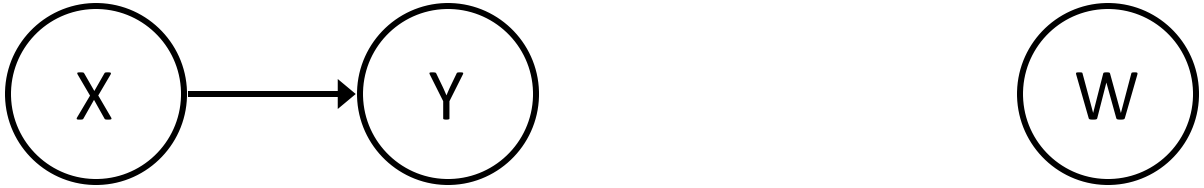


$$\frac{\cancel{P(X)}}{1}\frac{P(Y,X)}{\cancel{P(X)}}\frac{P(Z,Y)}{P(Y)}\frac{P(W,Z)}{P(Z)}$$



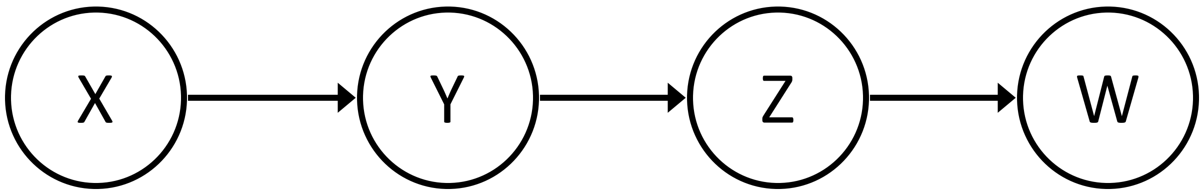
$$\frac{\cancel{P(X)}}{1}\frac{P(Y,X)}{\cancel{P(X)}}\frac{P(Y,W,Z)}{P(Y,W)}\frac{P(W)}{1}$$

Marginalize over Z:

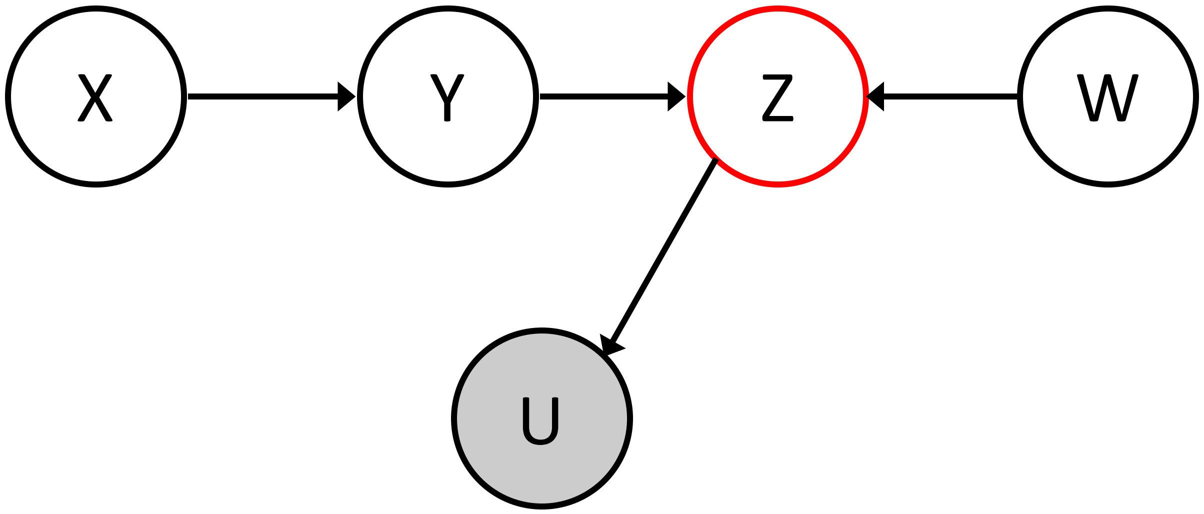


$$\frac{P(Y,X)}{1}\frac{\cancel{P(Y,W)}}{\cancel{P(Y,W)}}\frac{P(W)}{1}$$

Chain Probabilities



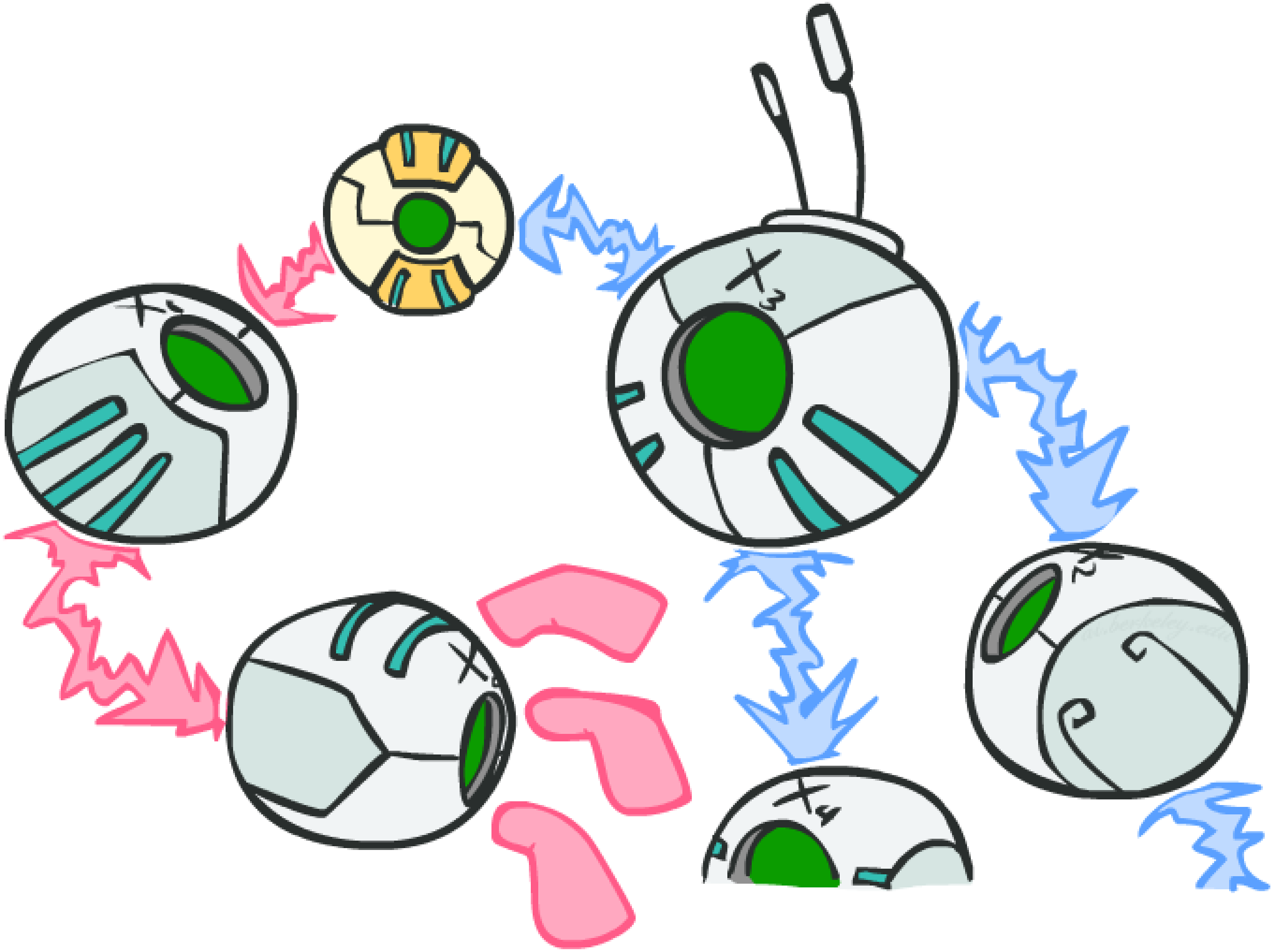
$$\frac{P(Y,X)}{1}\frac{P(Z,Y)}{P(Y)}\frac{P(W,Z)}{P(Z)}$$



$$\frac{P(Y,X)}{1}\frac{P(Y,W,Z)}{P(Y,W)}\frac{P(W)}{1}$$

Note: We can't marginalize over Z if Z is an evidence node, or if some other evidence node U depends on Z (since that changes the distribution of Z):

D-separation



D-separation: Overview

- D-separation:
  - a condition / algorithm for answering conditional independence queries from just studying the graph
- How:
  - Study independence properties for triples
  - Analyze complex cases as composition of triples

Triple Type 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 ? *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

Triple Type 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?

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

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

$$= P(z|y)$$

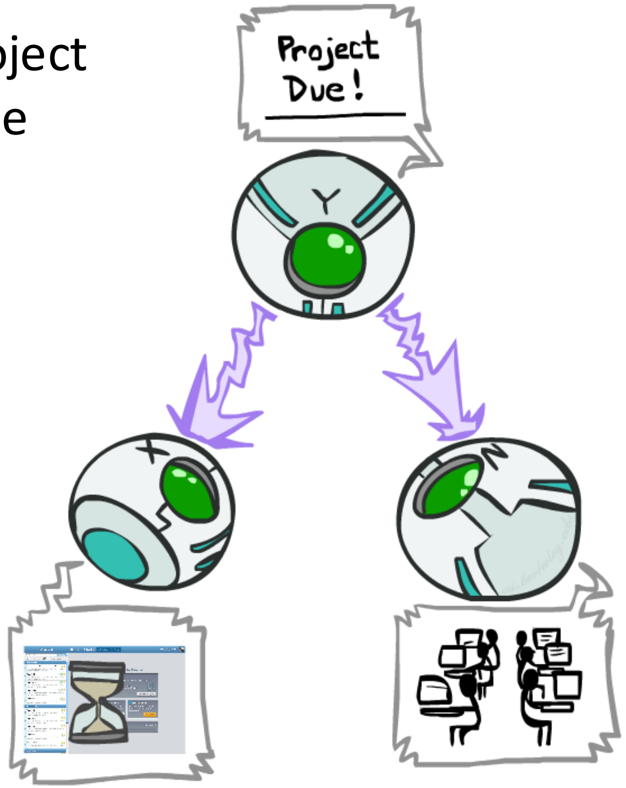
Yes!

- Evidence along the chain "blocks" the influence

Triple Type 2: Common Cause

- This configuration is a “common cause”

Y: Project due



X: Forums busy

Z: Lab full

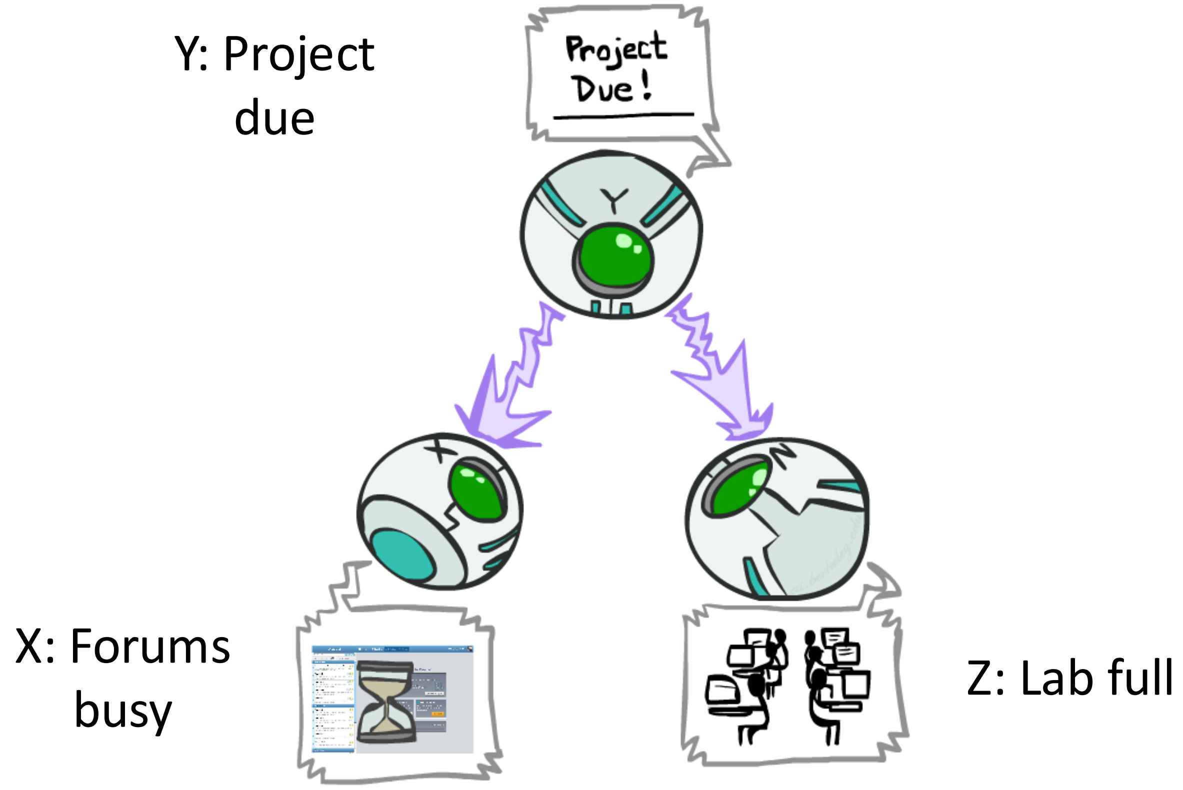
$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

- Guaranteed X 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

Triple Type 2: Common Cause

- This configuration is a "common cause"



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

- Guaranteed X and Z independent given Y?

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

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

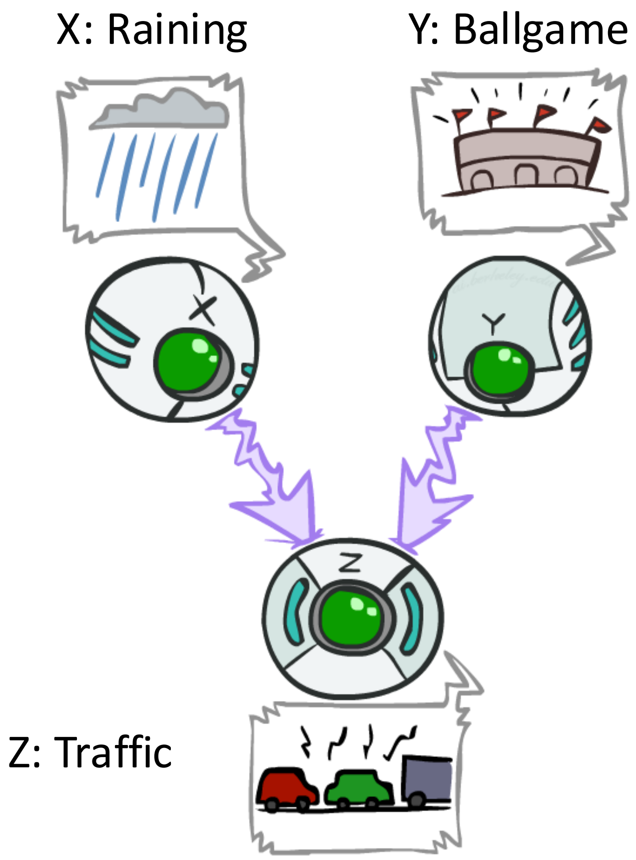
$$= P(z|y)$$

Yes!

- Observing the cause blocks influence between effects.

Triple Type 3: Common Effect

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

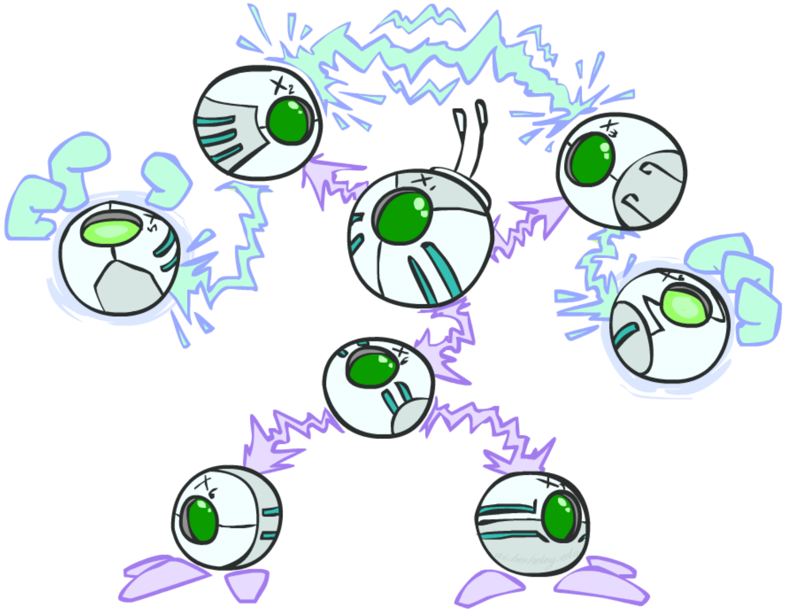


- Are X and Y independent?
  - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
  - Still need to prove they must be (try it!)
- Are X and Y independent given Z?
  - *No*: seeing traffic puts the rain and the ballgame in competition as explanation.
- This is backwards from the other cases
  - Observing an effect activates influence between possible causes.

The General Case

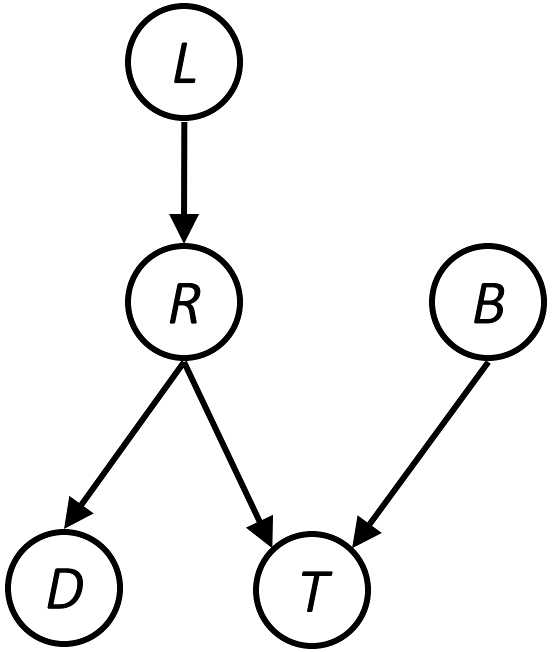
The General Case

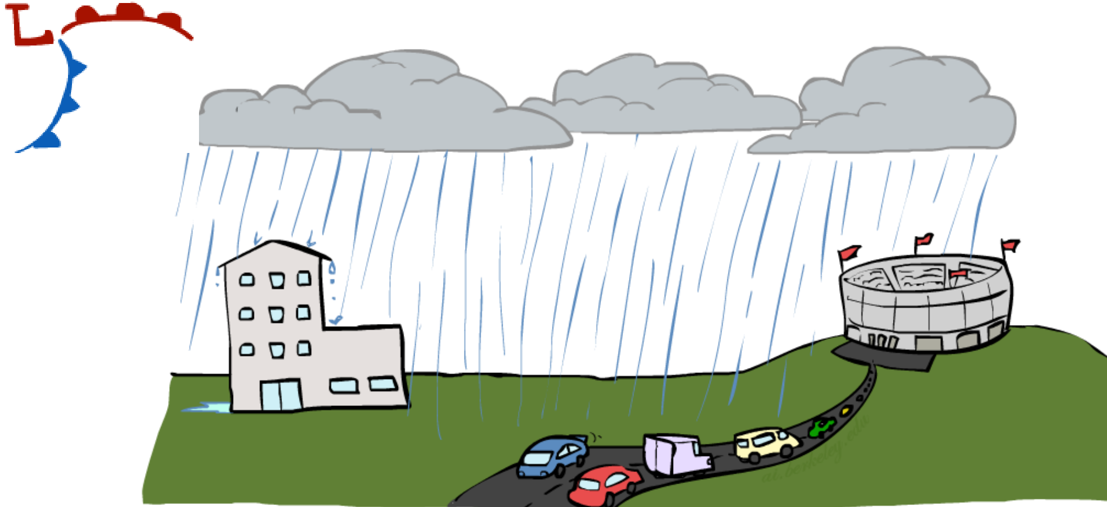
- General question: in a given BN, are two variables independent given some evidence?
- Last time we only considered evidence at parents of a single node.
- General Solution: analyze the graph
- Each path can be seen as repetitions of the three canonical cases



Reachability

- Recipe: shade evidence nodes, look for paths in the resulting graph
- Attempt 1: if two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent given the evidence (shaded nodes).
- Almost works, but not quite
  - Where does it break?
  - Answer: the structure at T doesn't count as a link in a path unless "active"





From Triples to Paths to D-Separation

- A path is active if each (overlapping) triple is active:

  Note: e.g. for a path A – B – C – D – E, the triples are:

  A – B – C, B – C – D, C – D – E

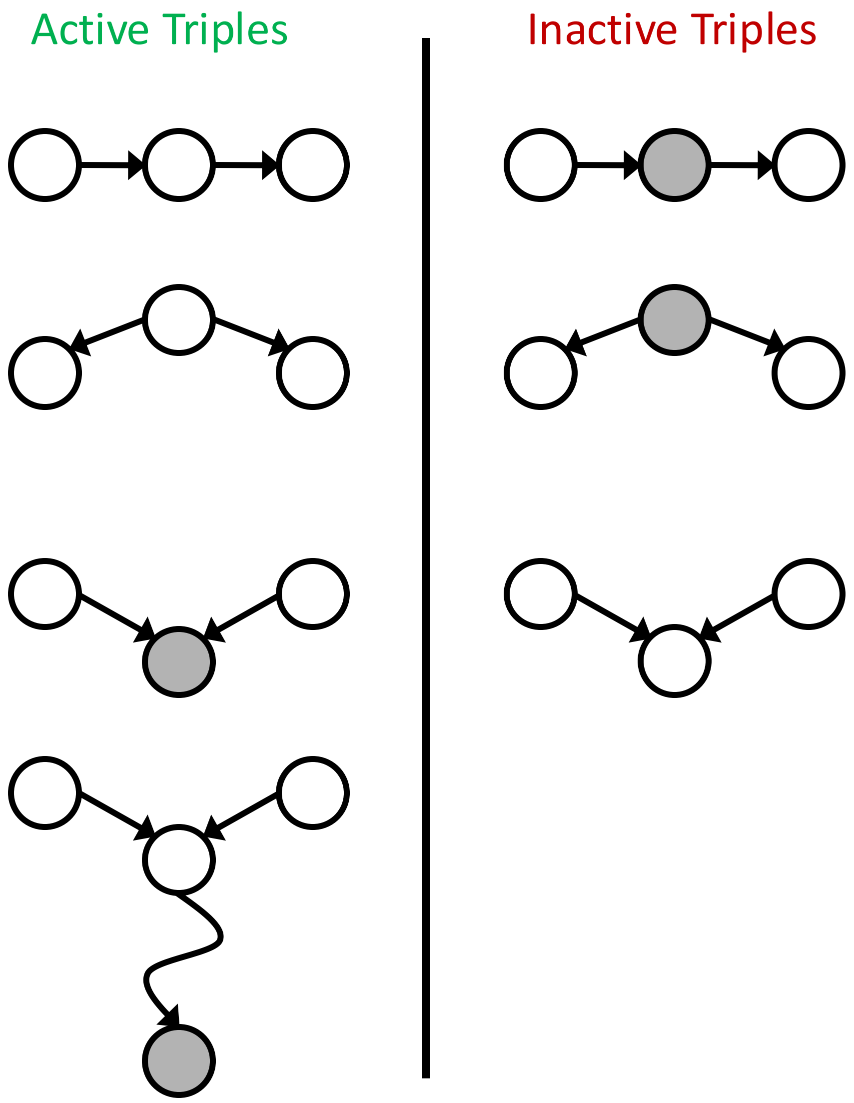
  Note: all it takes to block a path is a single inactive segment

- Are X and Y "D-separated" given evidence variables {Z}?
  - Consider all (undirected) paths from X to Y
  - If none of the paths are active, then X and Y are D-separated given {Z}
  - On the other hand, if there is at least one active path, then X and Y are not D-separated given {Z}

- Independence and D-separation:

  X and Y are guaranteed conditionally independent given {Z}
  IF AND ONLY IF
  X and Y are d-separated given {Z}

  → just need to check the graph



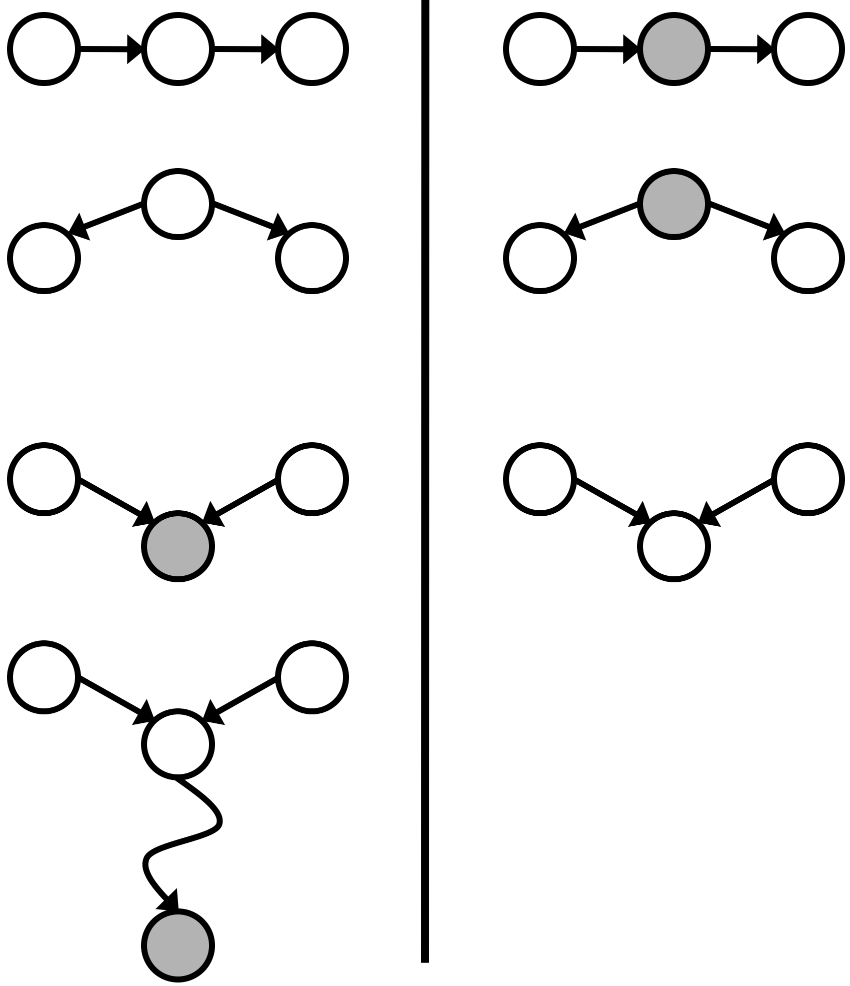
Recap of Triples

Active Triples | Inactive Triples

Causal Chain:

Common Cause:

Common Effect (“v-structure”)

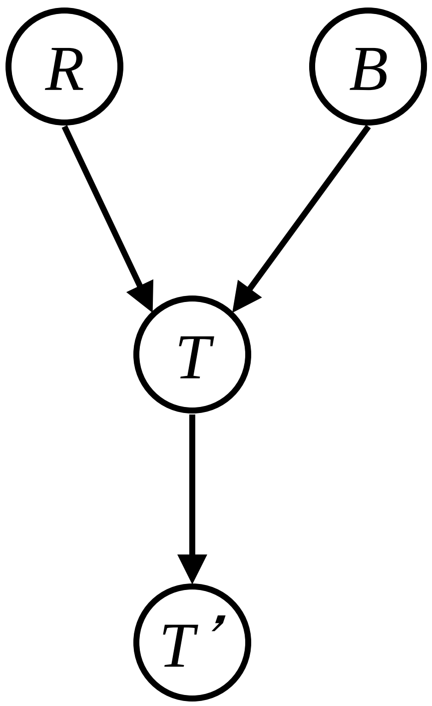


Example

Red = Nodes are conditionally independent given the evidence

Blue = Nodes are d-separated given the evidence

$R \perp\!\!\!\perp B$	*Yes*	*Yes*
$R \perp\!\!\!\perp B \mid T$	*No*	*??*
$R \perp\!\!\!\perp B \mid T'$	*No*	*??*



Example

Red = Nodes are conditionally independent given the evidence

Blue = Nodes are d-separated given the evidence

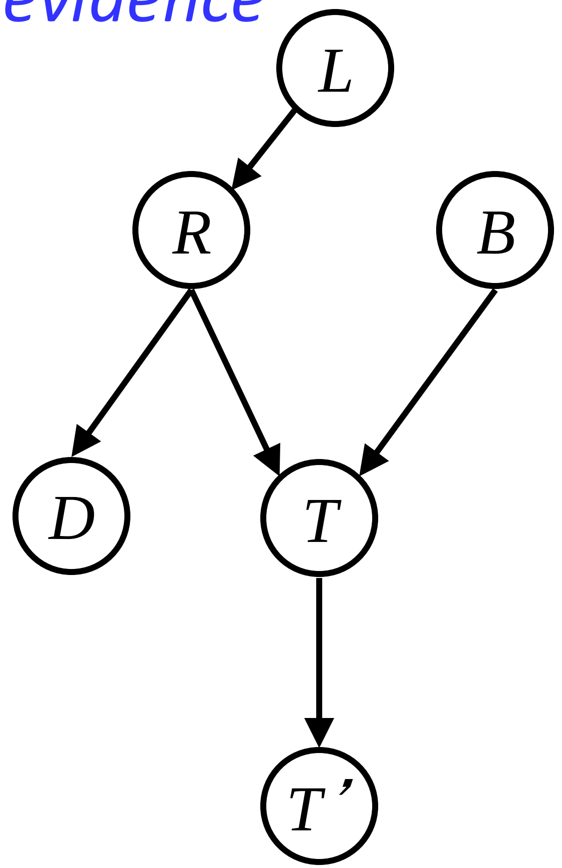
$L \perp\!\!\!\perp T' | T$ *Yes*   *Yes*

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

$L \perp\!\!\!\perp B | T$ *No*   *??*

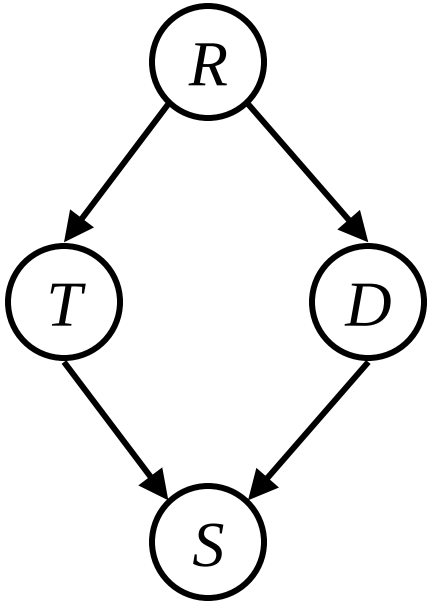
$L \perp\!\!\!\perp B | T'$ *No*   *??*

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



Example

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



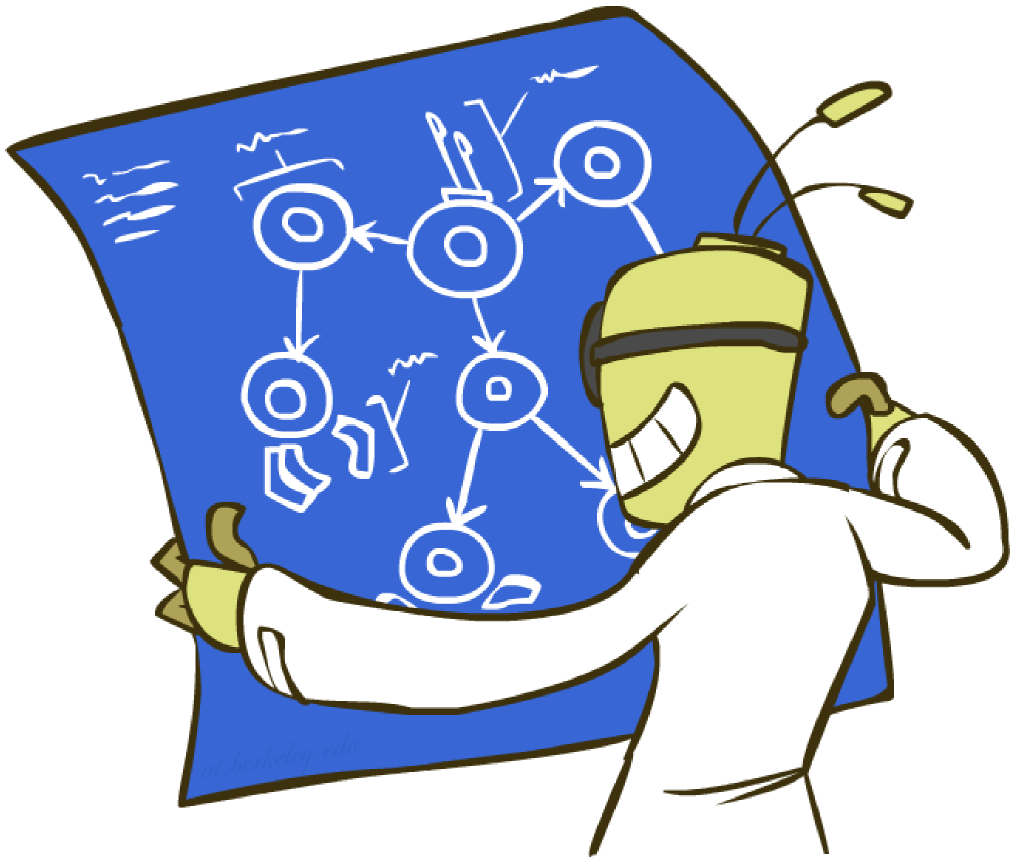
$T \perp\!\!\!\perp D$	*No*	*??*
$T \perp\!\!\!\perp D \mid R$	*Yes*	*Yes*
$T \perp\!\!\!\perp D \mid R, S$	*No*	*??*

Structure Implications

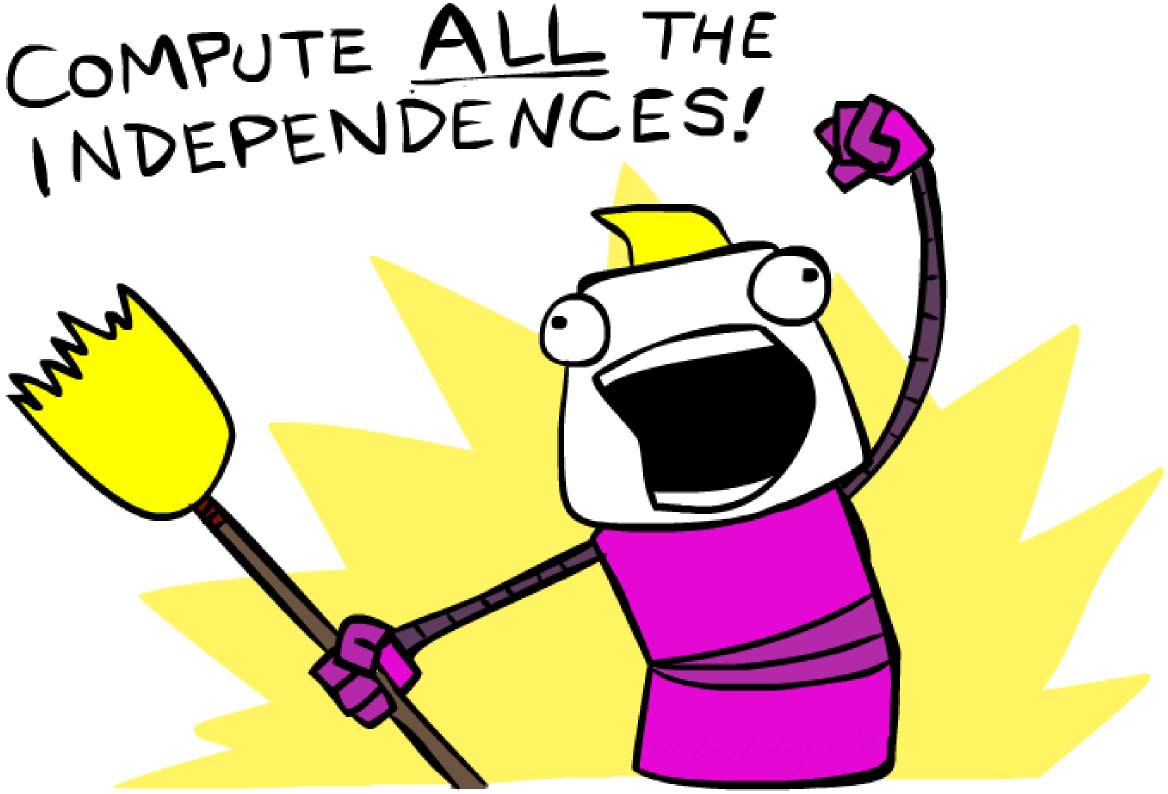
- Given a Bayes net structure, can run d-separation algorithm to build a complete list of conditional independences that are necessarily true of the form

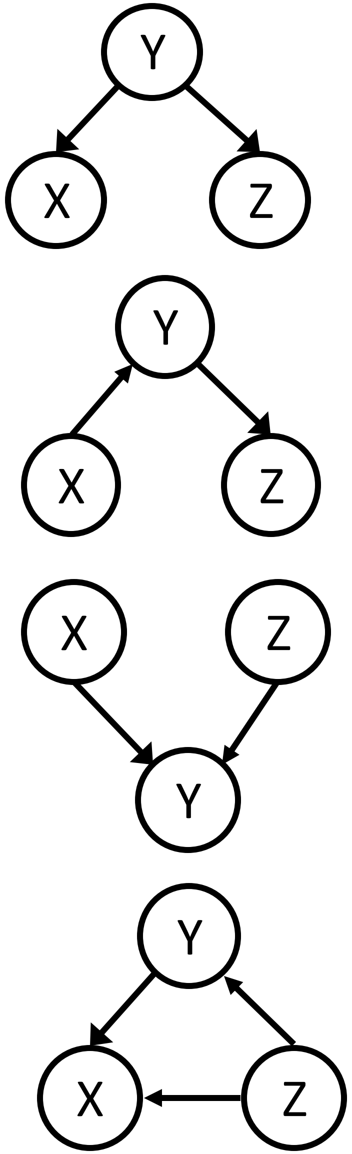
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

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



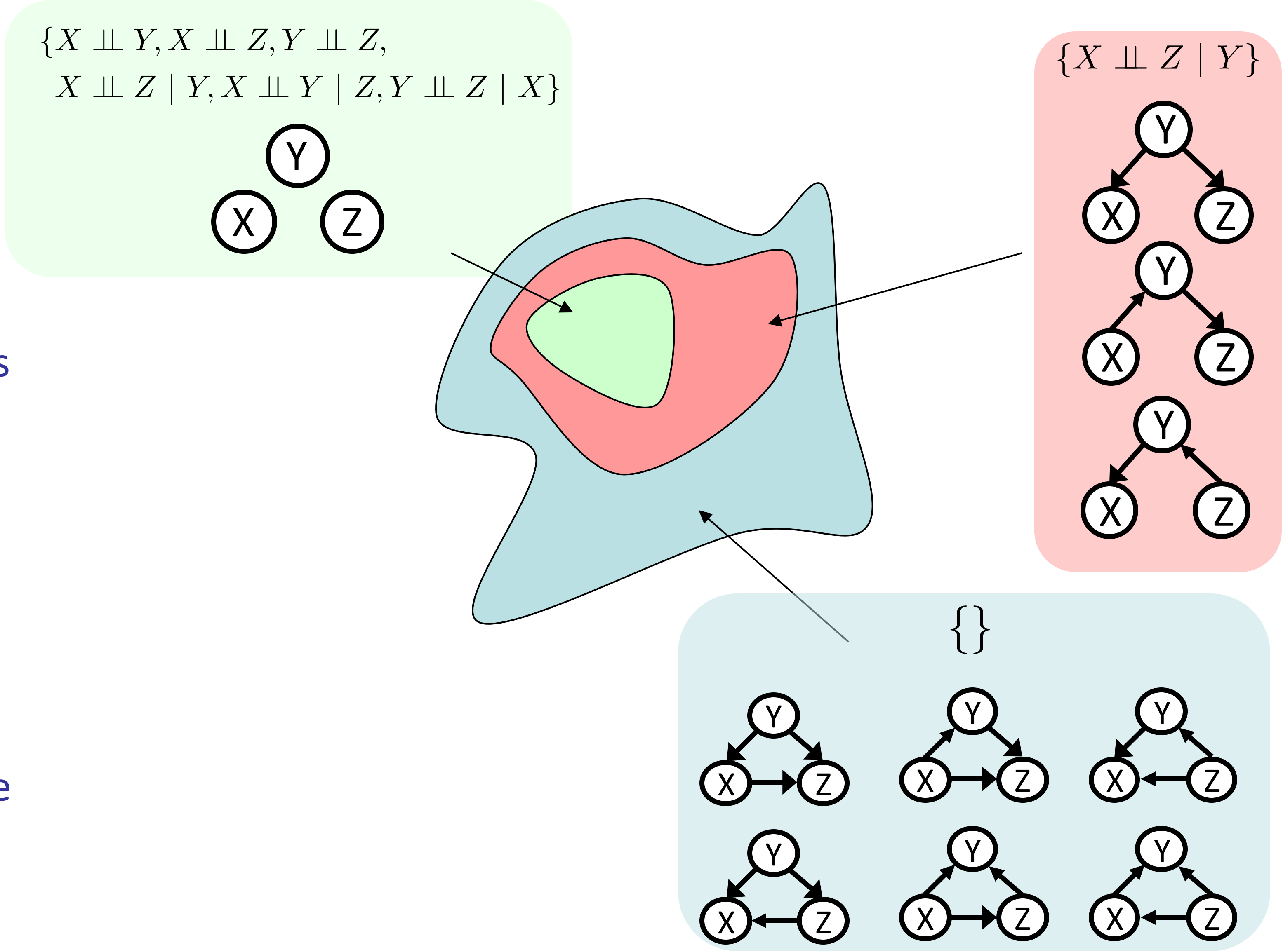
Computing All Independences





Topology Limits Distributions

- Given some graph 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



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

Bayes' Nets

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