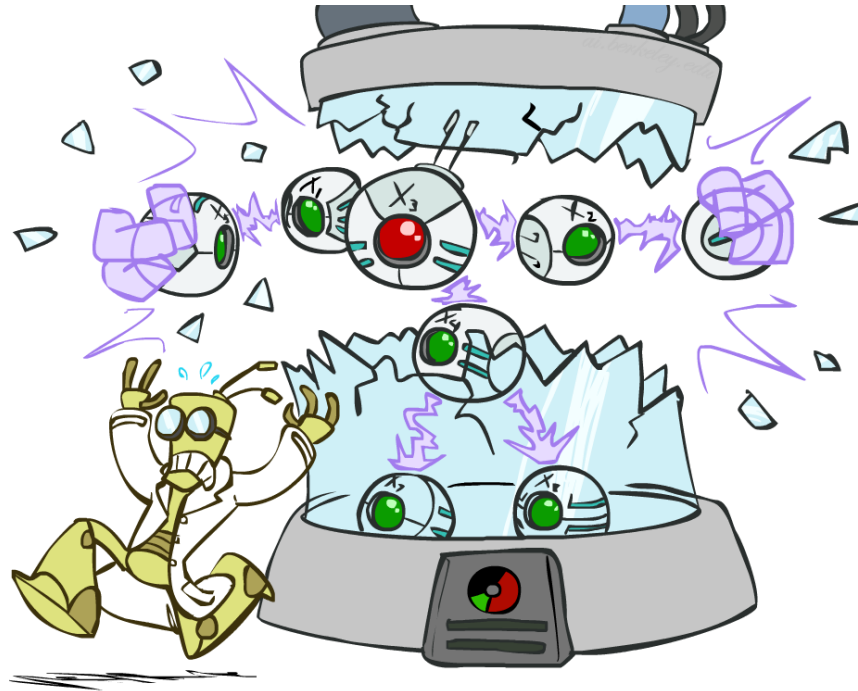


Announcements

- Midterm Exam #2 will be on Tuesday Nov 6 (same day as the US Midterm Elections).
- Topics covered: **Expectimax and Utilities, MDPs, Reinforcement Learning, Probability, Markov Models and HMMs, Bayes' Nets.**
- You should review: ALMA chapters 5 and 13-17, plus Sutton and Barto chapters 3 and 6
- I'll try to release a practice exam by the weekend.
- Extra credit: vote before you come, and bring in an "I voted sticker" AND/OR take a friend to the polls and get their sticker. More EC for more GOTV. Polls open at 7AM.

Bayes' Nets part 2



Slides courtesy of Dan Klein and Pieter Abbeel --- University of California, Berkeley

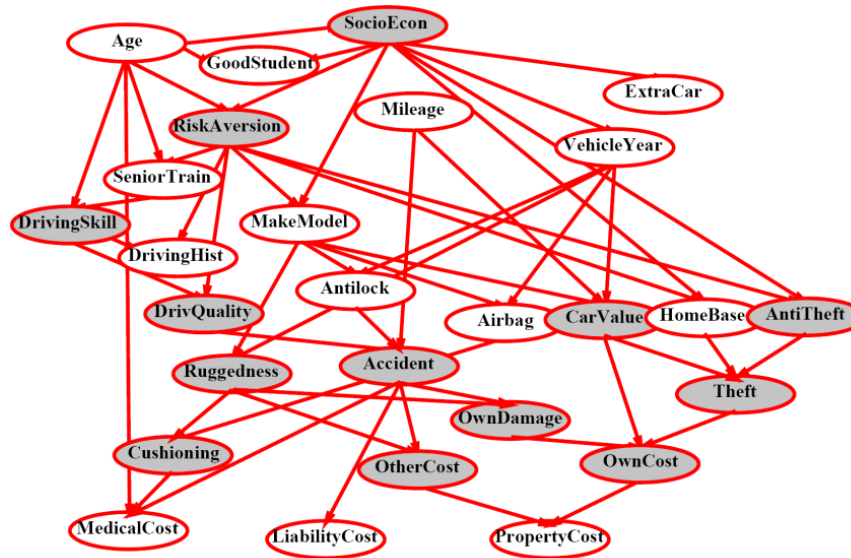
[These slides were created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley. All CS188 materials are available at <http://ai.berkeley.edu>.]

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
$$\begin{aligned} P(X_1, X_2, \dots, X_n) &= P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots \\ &= \prod_{i=1}^n P(X_i|X_1, \dots, X_{i-1}) \end{aligned}$$
- 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: $X \perp\!\!\!\perp Y | Z$
$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

Bayes' Nets

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



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

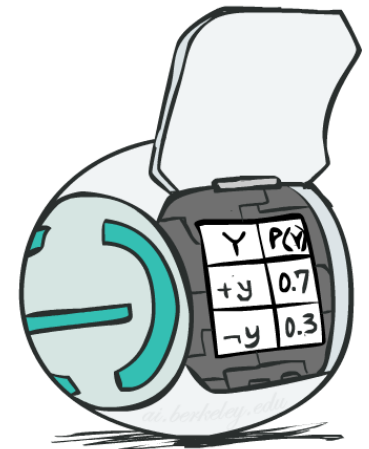
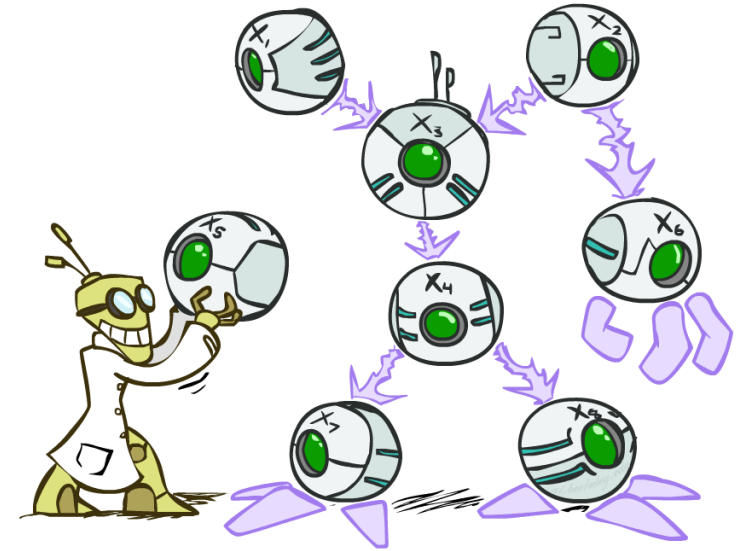
Bayes' Net Semantics

- 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

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

- 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:

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



Probabilities in BNs



- Why are we guaranteed that setting

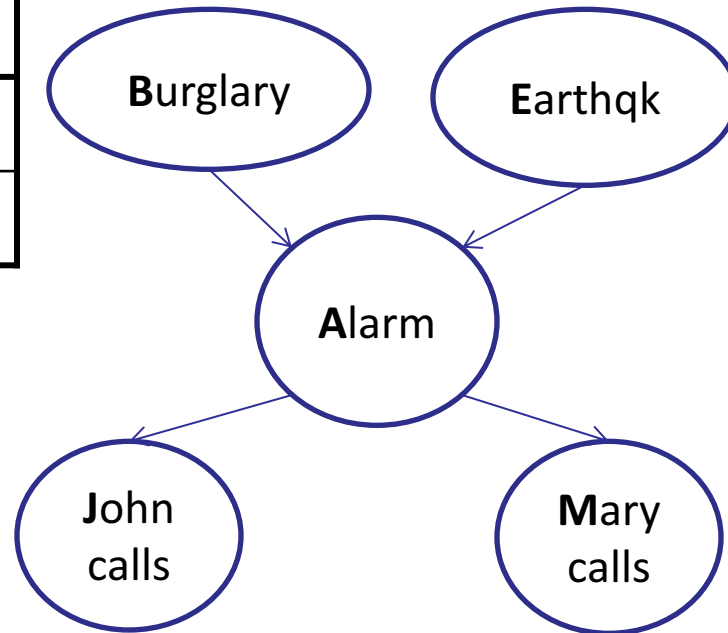
$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i))$$

results in a proper joint distribution?

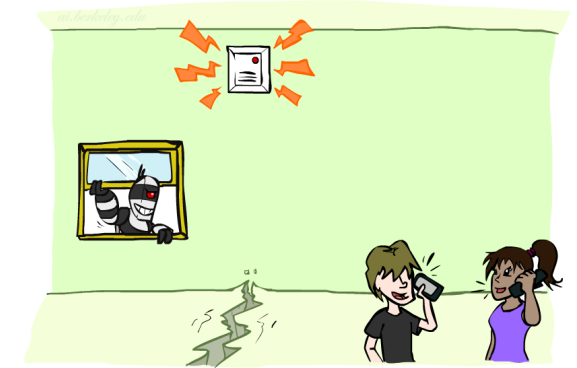
- Chain rule (valid for all distributions): $P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | x_1 \dots x_{i-1})$
- Assume conditional independences: $P(x_i | x_1, \dots, x_{i-1}) = P(x_i | \text{parents}(X_i))$
 - Consequence: $P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(X_i))$
- Not every BN can represent every joint distribution
 - The topology enforces certain conditional independencies

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

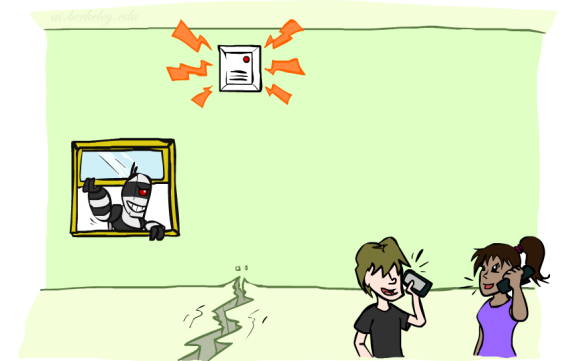
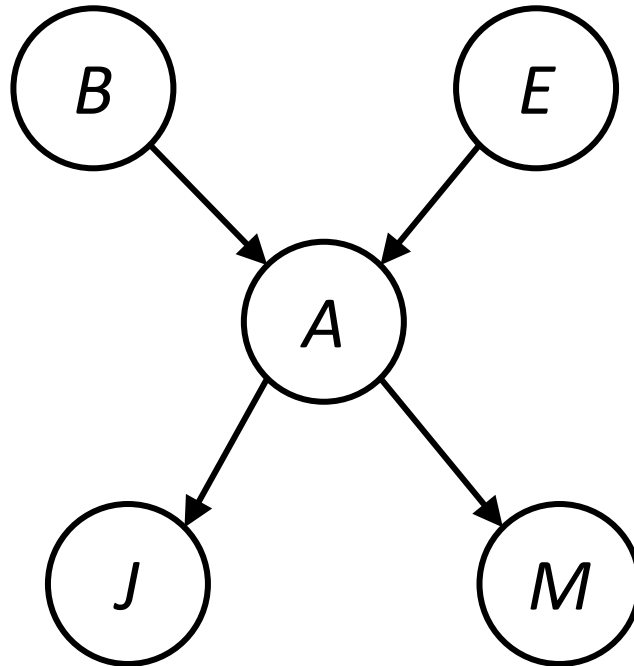
Example: Alarm Network

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

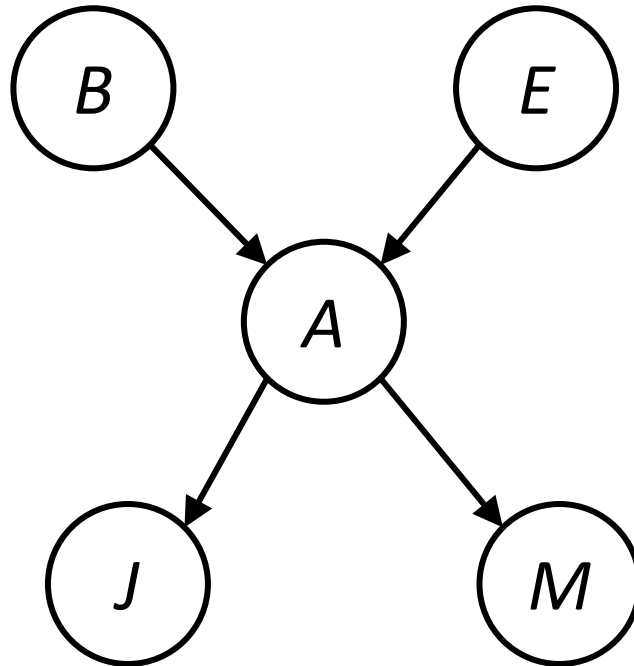


$$P(+b, -e, +a, -j, +m) =$$

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

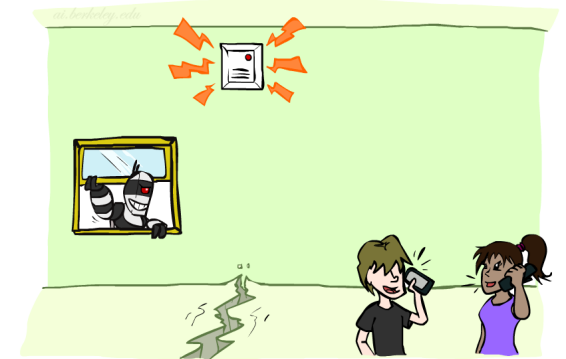
Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

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



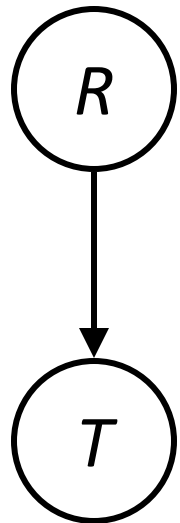
A	J	P(J A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

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

$$\begin{aligned}
 &P(+b, -e, +a, -j, +m) = \\
 &P(+b)P(-e)P(+a|+b, -e)P(-j|+a)P(+m|+a) = \\
 &0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7
 \end{aligned}$$

Example: Traffic

- Causal direction



$P(R)$

+r	1/4
-r	3/4

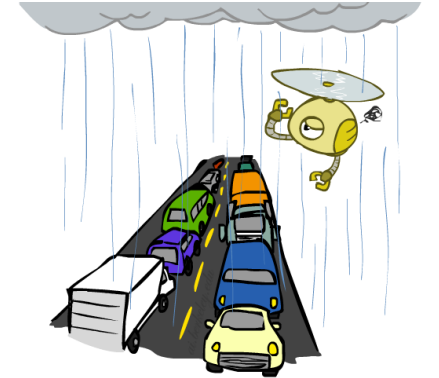
$P(T|R)$

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

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

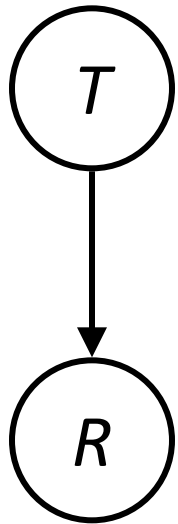
$P(T, R)$

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



Example: Reverse 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

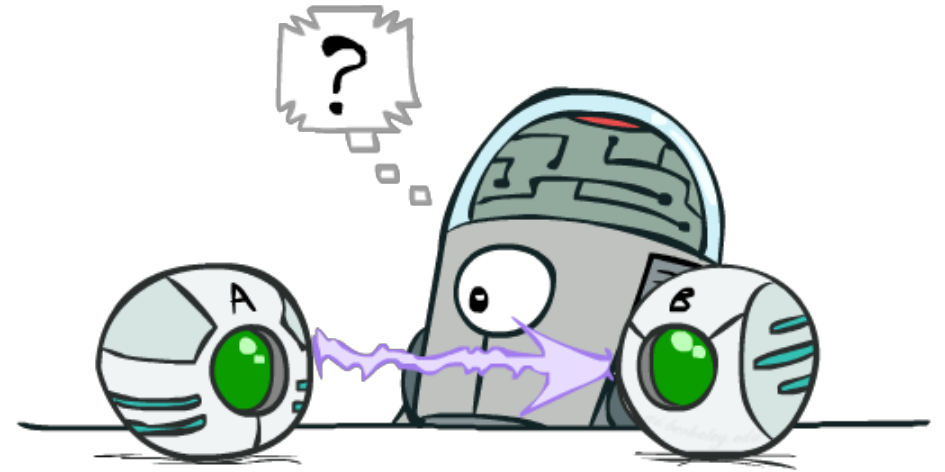


$P(T, R)$

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

Causality?

- When Bayes' nets reflect the true causal patterns:
 - Often simpler (nodes have fewer parents)
 - Often easier to think about
 - Often easier to elicit from experts
- BNs need not actually be causal
 - Sometimes no causal net exists over the domain (especially if variables are missing)
 - E.g. consider the variables *Traffic* and *Drips*
 - End up with arrows that reflect correlation, not causation
- What do the arrows really mean?
 - Topology may happen to encode causal structure
 - **Topology really encodes conditional independence**
$$P(x_i | x_1, \dots, x_{i-1}) = P(x_i | \text{parents}(X_i))$$



Size of a Bayes' Net

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

$$2^N$$

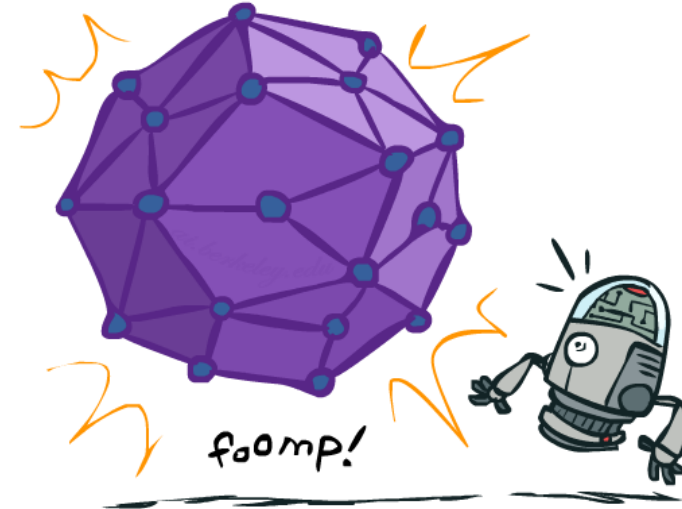
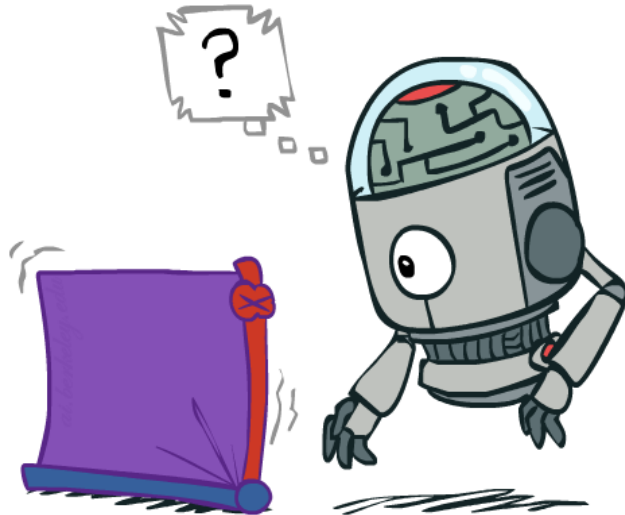
- How big is an N -node net if nodes have up to k parents?

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

- Both give you the power to calculate

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

- 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

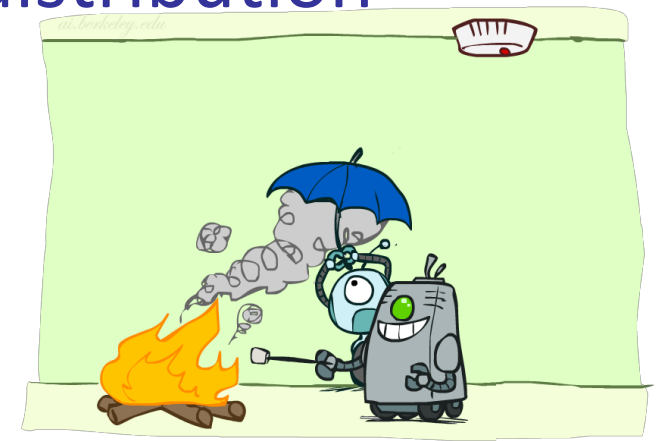
$$\forall x, y \quad P(x, y) = P(x)P(y) \quad \text{---} \rightarrow \quad X \perp Y$$

- X and Y are **conditionally independent** given Z

$$\forall x, y, z \quad P(x, y|z) = P(x|z)P(y|z) \quad \text{---} \rightarrow \quad X \perp Y|Z$$

- (Conditional) independence is a property of a distribution

- Example: $Alarm \perp Fire|Smoke$

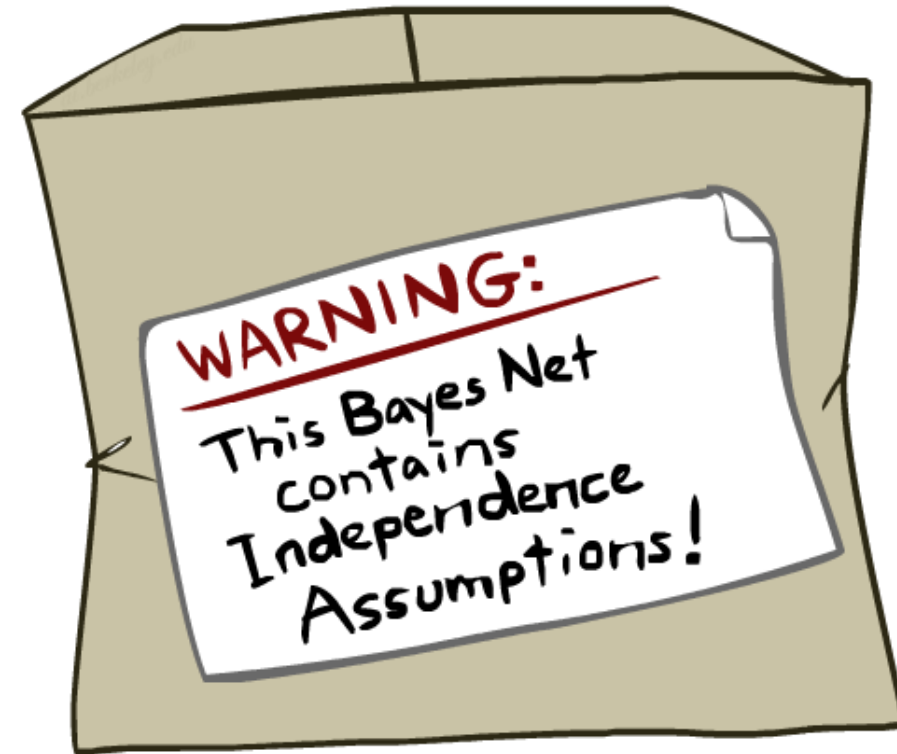


Bayes Nets: Assumptions

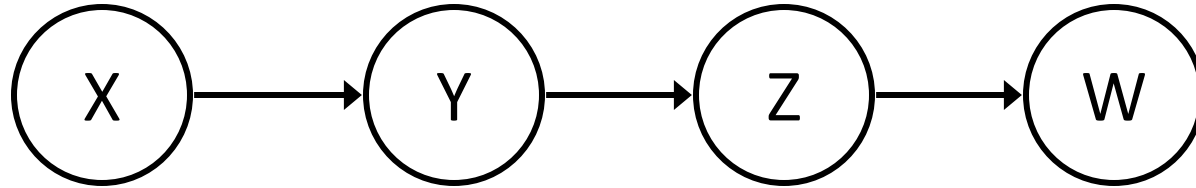
- Assumptions we are required to make to define the Bayes net when given the graph:

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

- Beyond above “chain rule \rightarrow Bayes net” conditional independence assumptions
 - Often additional conditional independences
 - They can be read off the graph
- Important for modeling: understand assumptions made when choosing a Bayes net graph



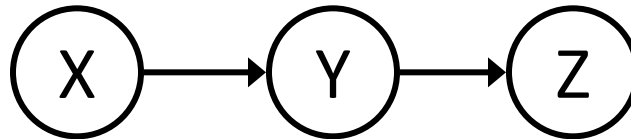
Example



- Conditional independence assumptions directly from simplifications in chain rule:
- Additional implied conditional independence assumptions?

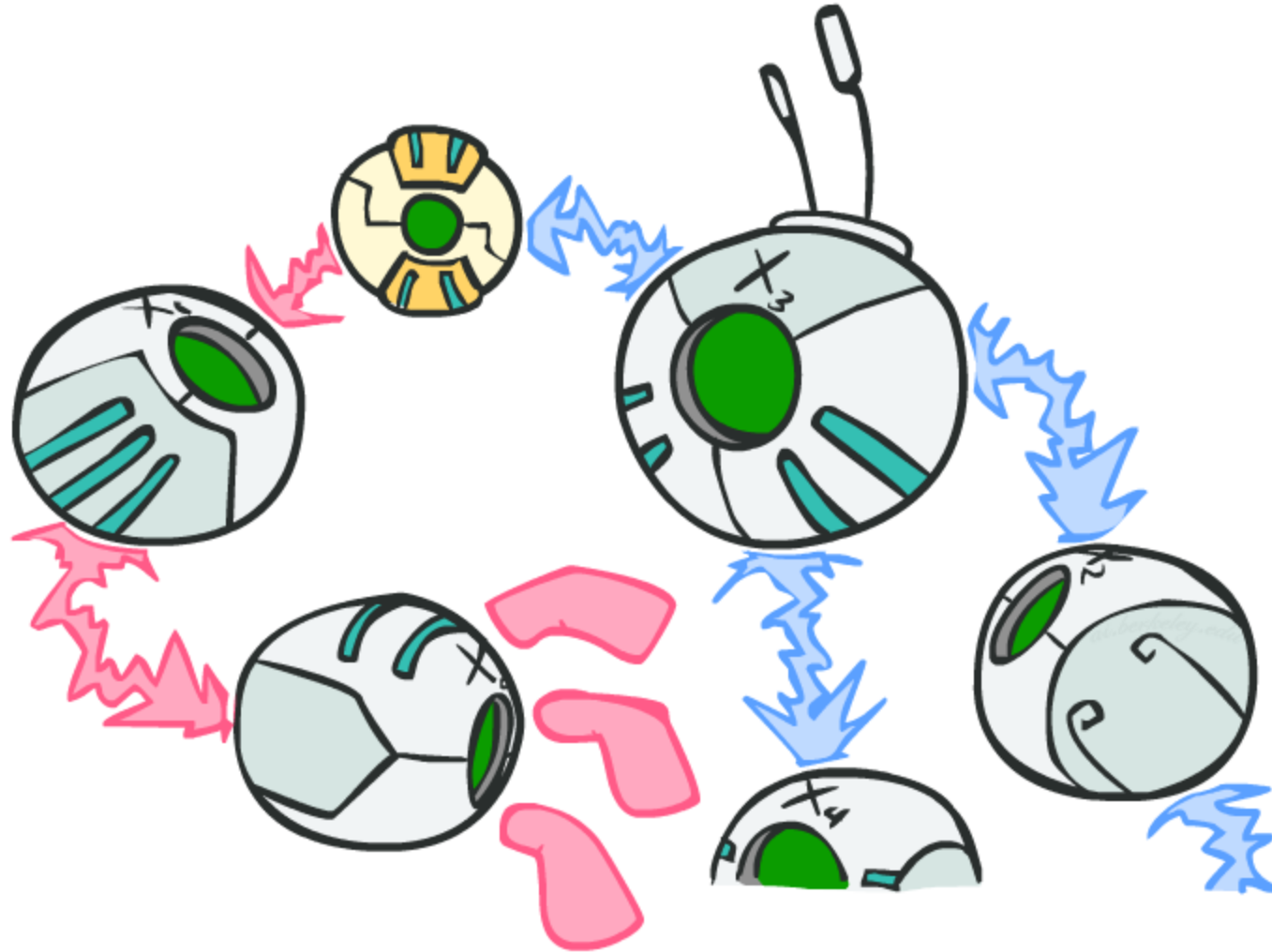
Independence in a BN

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



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

D-separation: Outline



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:

$$P(+y \mid +x) = 1, P(-y \mid -x) = 1, \\ P(+z \mid +y) = 1, P(-z \mid -y) = 1$$

Causal Chains

- This configuration is a “causal chain”

- Guaranteed X independent of Z given Y?



X: Low pressure

Y: Rain

Z: Traffic

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

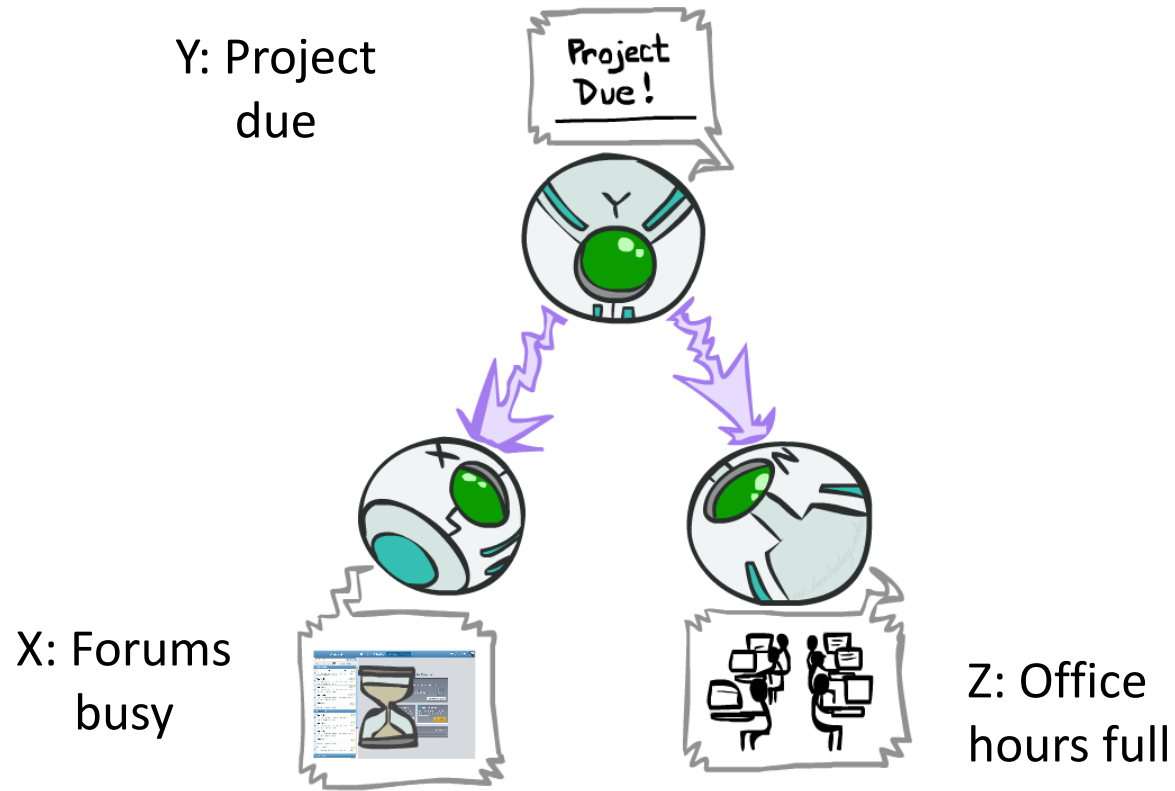
$$\begin{aligned} P(z|x, y) &= \frac{P(x, y, z)}{P(x, y)} \\ &= \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)} \\ &= P(z|y) \end{aligned}$$

Yes!

- Evidence along the chain “blocks” the influence

Common Cause

- This configuration is a “common cause”



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

- 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:

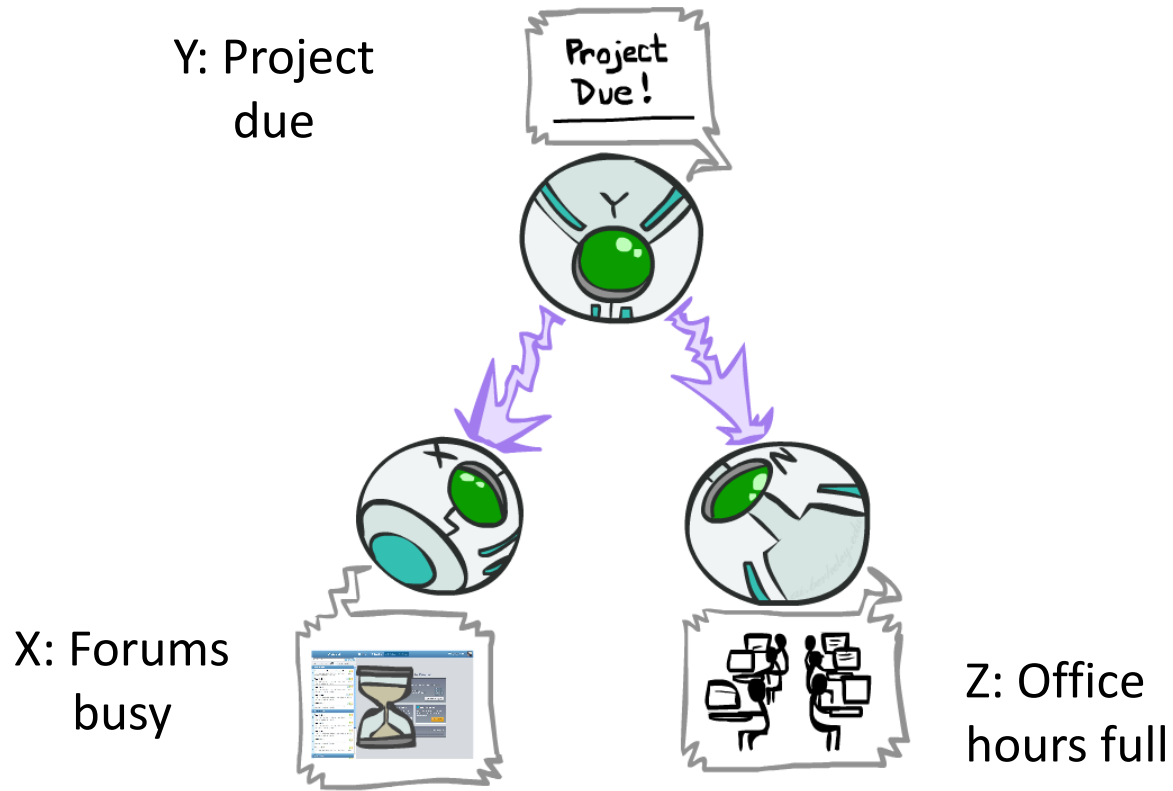
- Project due causes both forums busy and office hours to be full

- In numbers:

$$P(+x \mid +y) = 1, P(-x \mid -y) = 1, \\ P(+z \mid +y) = 1, P(-z \mid -y) = 1$$

Common Cause

- This configuration is a “common cause”
- Guaranteed X and Z independent given Y?



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

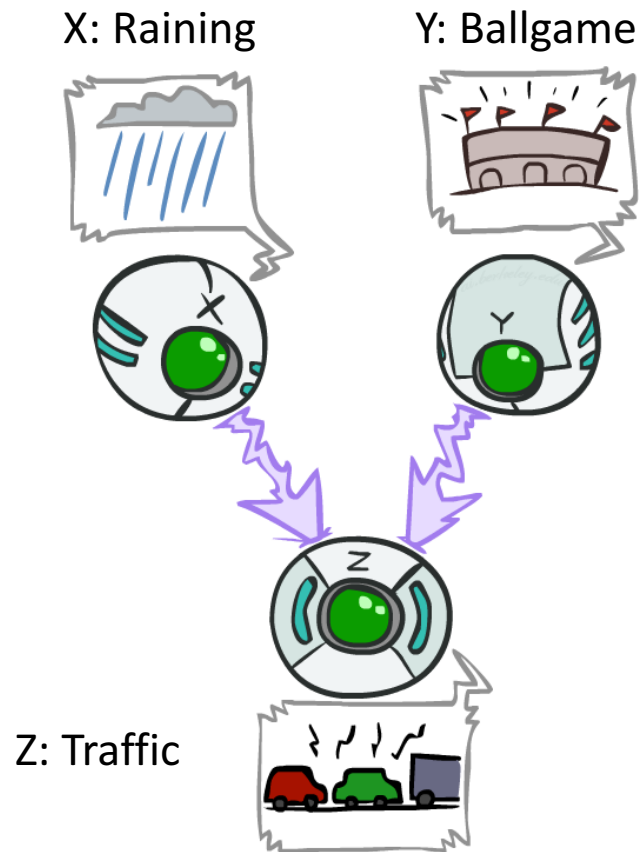
$$\begin{aligned} P(z|x, y) &= \frac{P(x, y, z)}{P(x, y)} \\ &= \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)} \\ &= P(z|y) \end{aligned}$$

Yes!

- Observing the cause blocks influence between effects.

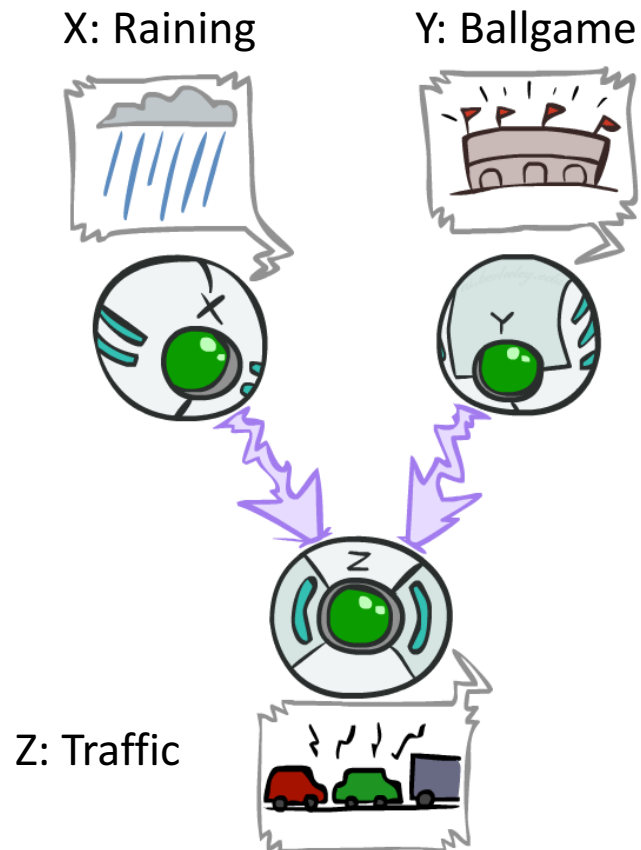
Common Effect

- Last configuration: two causes of one effect (v-structures)
- Are X and Y independent?



Common Effect

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



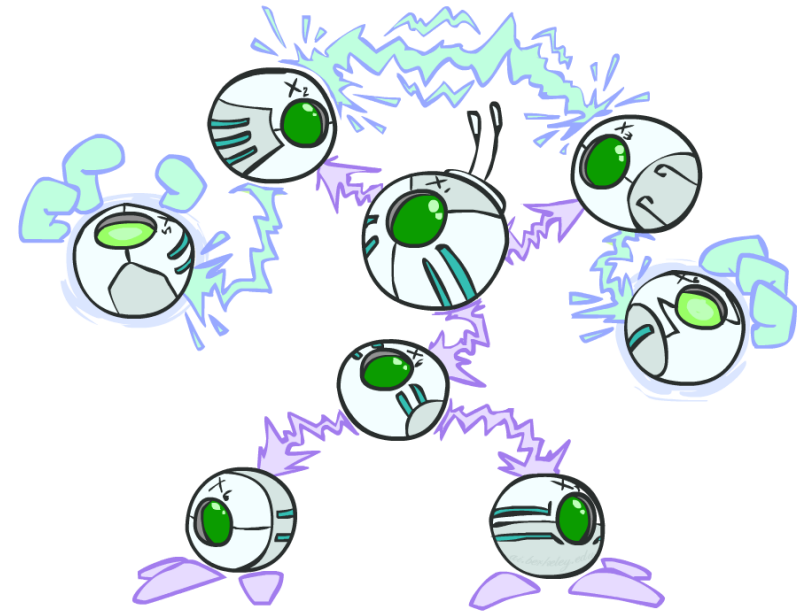
- Are X and Y independent?
 - **Yes**: the ballgame and the rain cause traffic, but they are not correlated
 - Still need to prove they must be (try it!)
- Are X and Y independent given Z?
 - **No**: seeing traffic puts the rain and the ballgame in competition as explanation.
- **This is backwards from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

The General Case



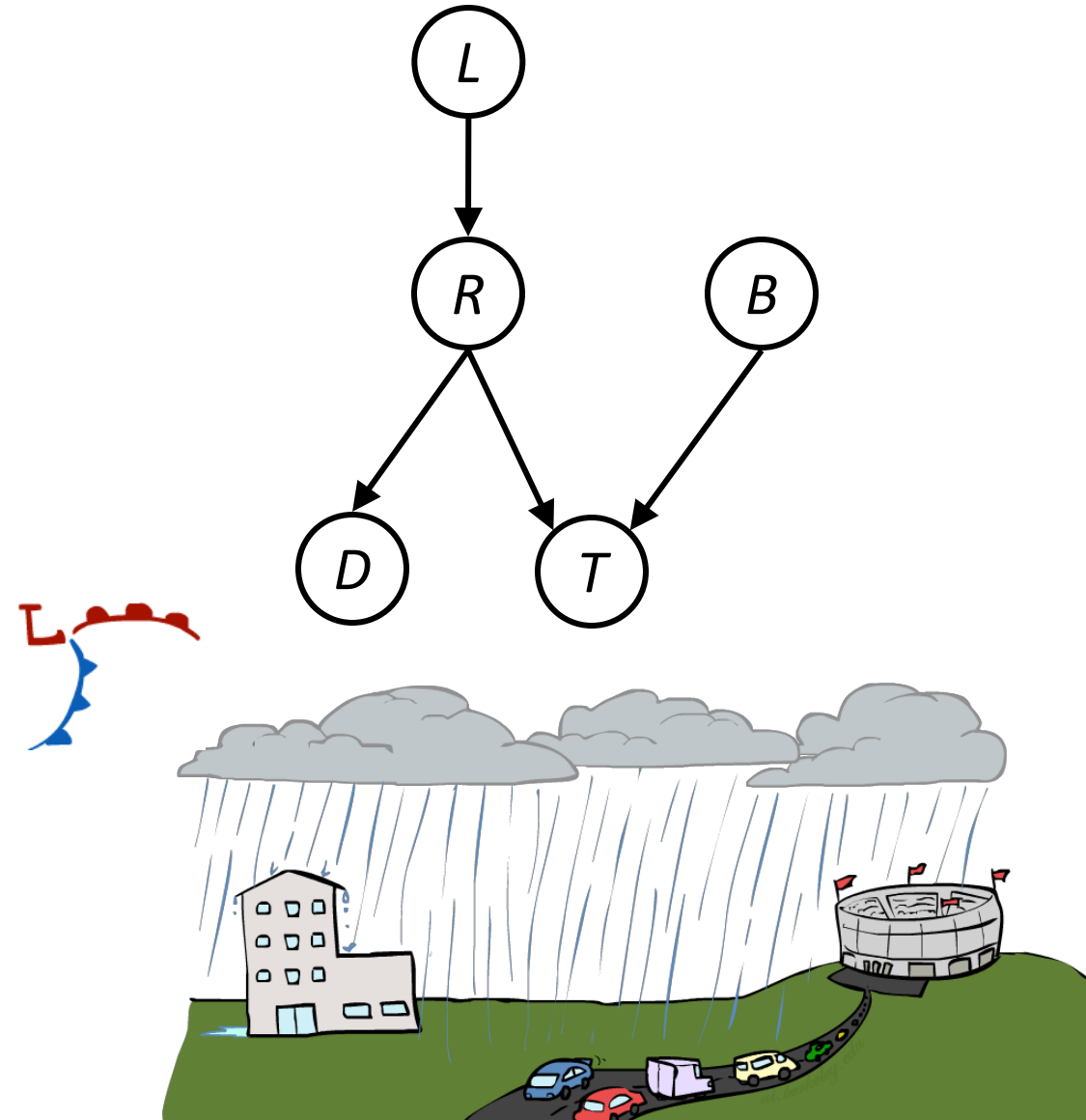
The General Case

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



Reachability

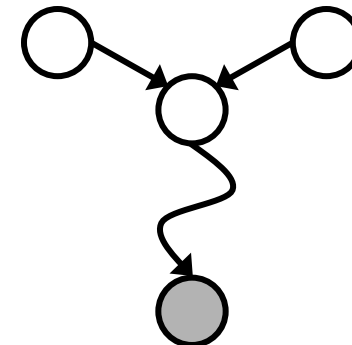
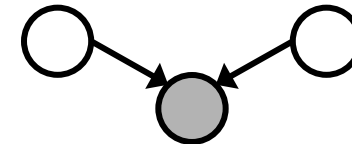
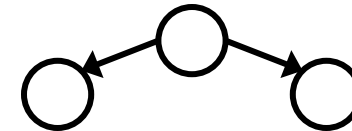
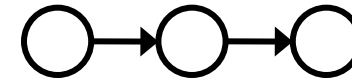
- Recipe: shade evidence nodes, look for paths in the resulting graph
- Attempt 1: if two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent
- Almost works, but not quite
 - Where does it break?
 - Answer: the v-structure at T doesn't count as a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



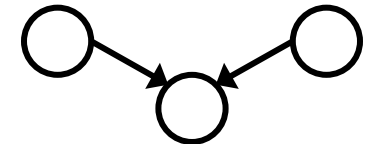
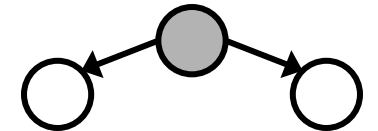
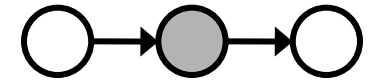
Active / Inactive Paths

- 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 $A \rightarrow B \rightarrow C$ 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 descendants is observed
- All it takes to block a path is a single inactive segment

Active Triples



Inactive Triples



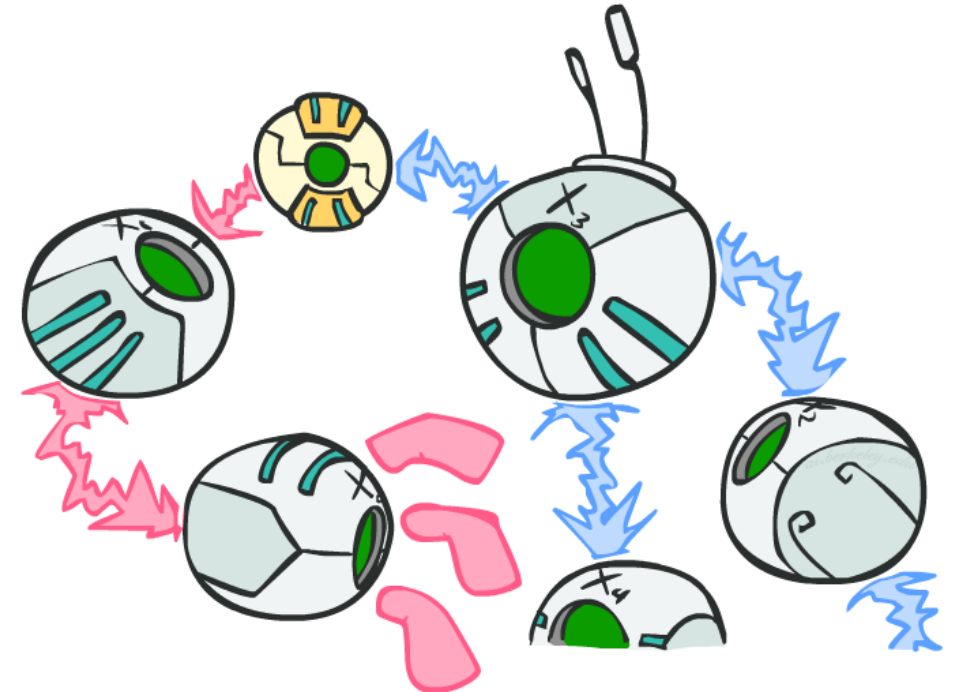
D-Separation

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

$$X_i \not\perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$$

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

$$X_i \perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$$



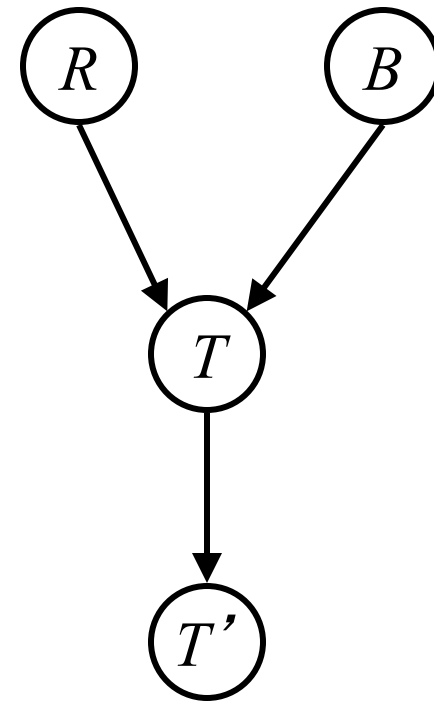
Example

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

Yes

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

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



Example

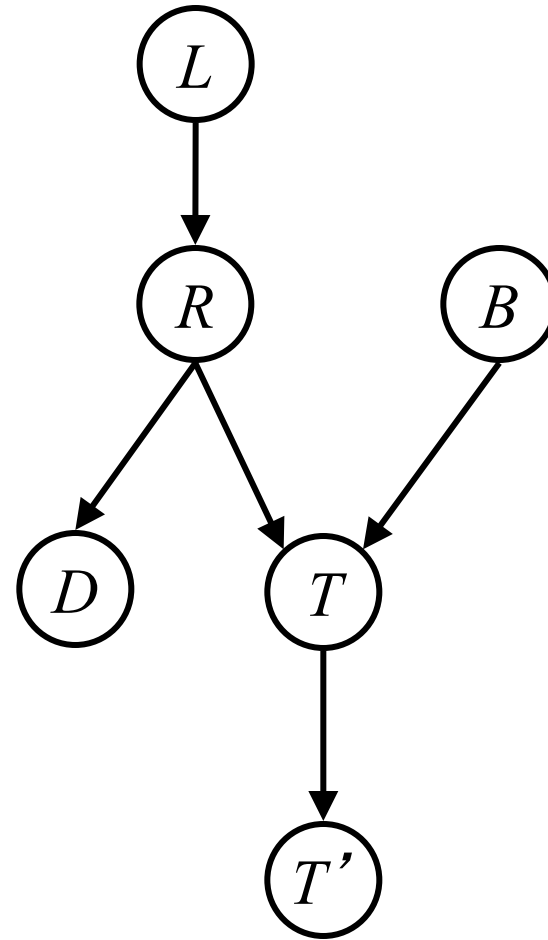
$L \perp\!\!\!\perp T' | T$ *Yes*

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

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

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

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



Example

- Variables:

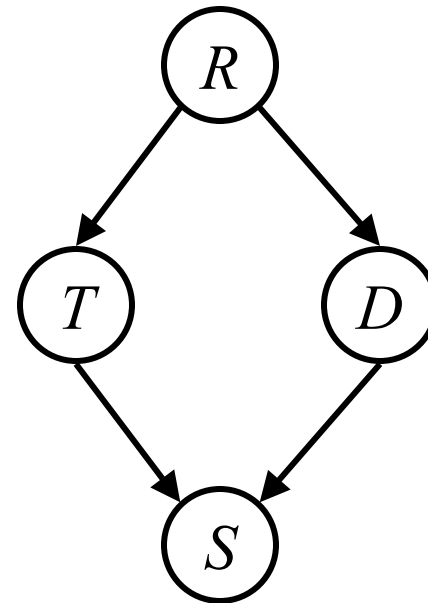
- R: Raining
- T: Traffic
- D: Roof drips
- S: I'm sad

- Questions:

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

$$T \perp\!\!\!\perp D | R \quad \text{Yes}$$

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

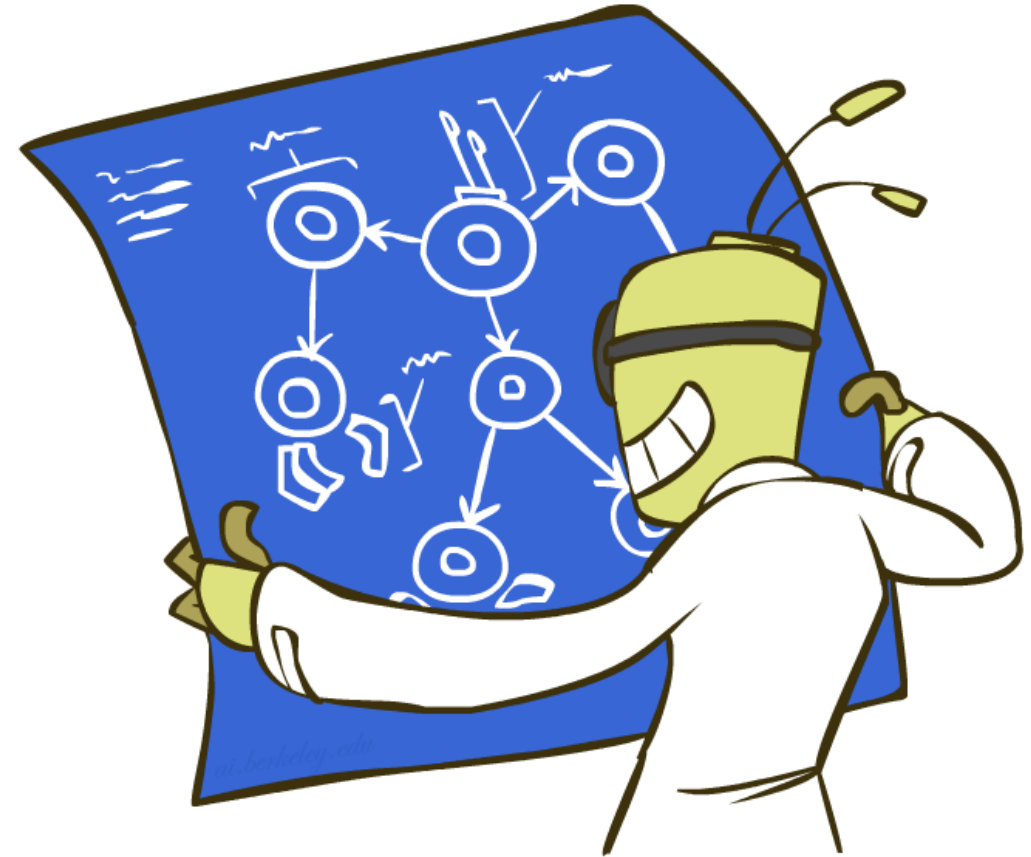


Structure Implications

- 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

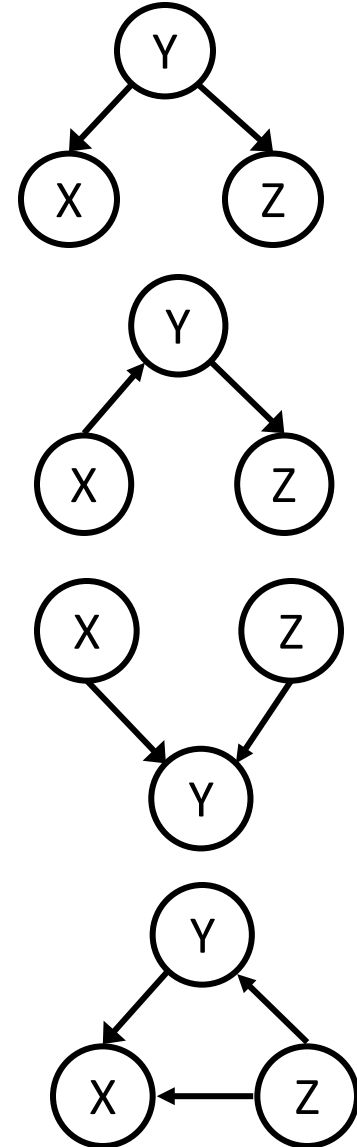
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, \dots, X_{k_n}\}$$

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



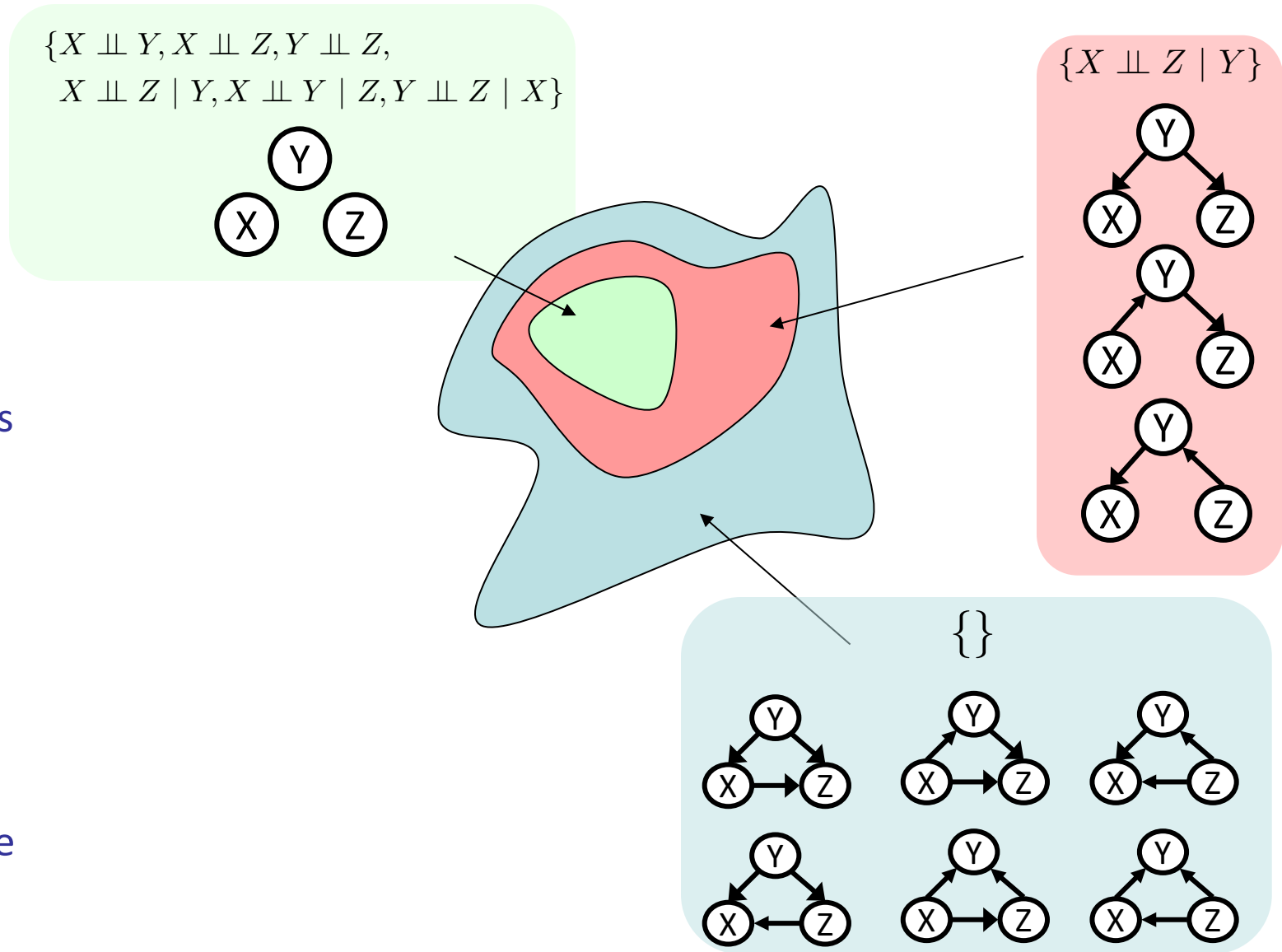
Computing All Independences

COMPUTE ALL THE
INDEPENDENCES!



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



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