

Artificial Intelligence

CE-417, Group 1

Computer Eng. Department

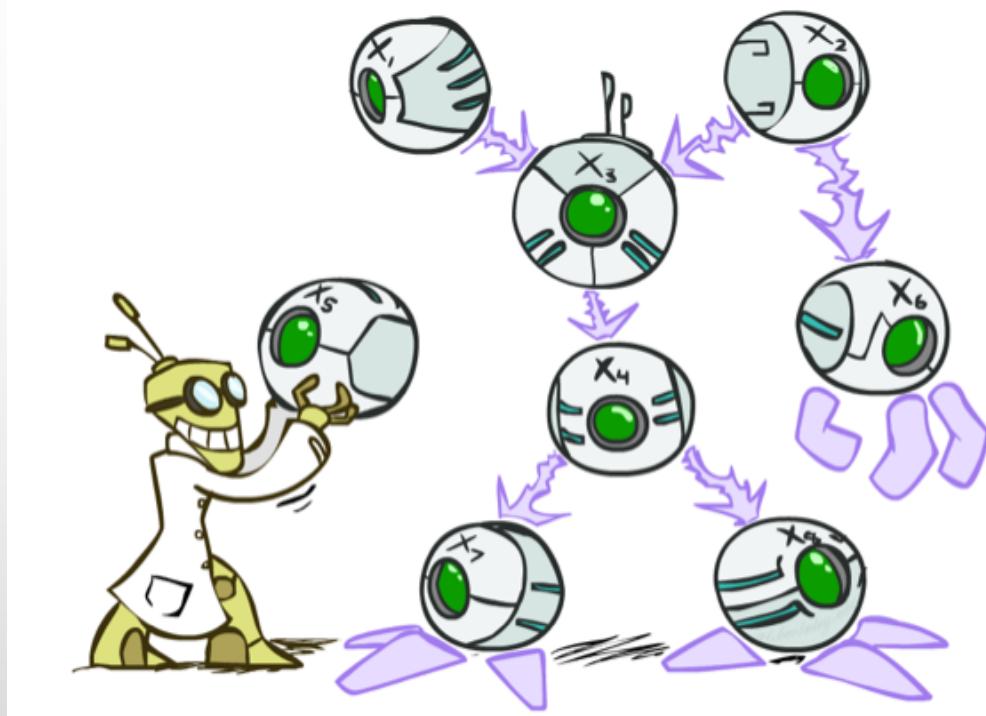
Sharif University of Technology

Spring 2020

By Mohammad Hossein Rohban, Ph.D.

Courtesy: Most slides are adopted from CSE-573 (Washington U.), original
slides for the textbook, and CS-188 (UC. Berkeley).

Introduction to Bayes' Networks



Probabilistic Models

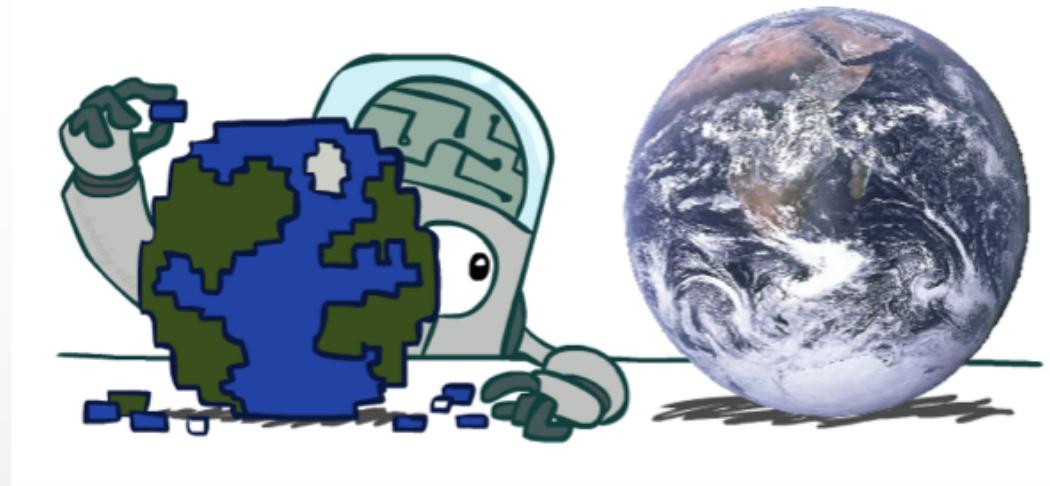
- Models describe how (a portion of) the world works

- **Models are always simplifications**

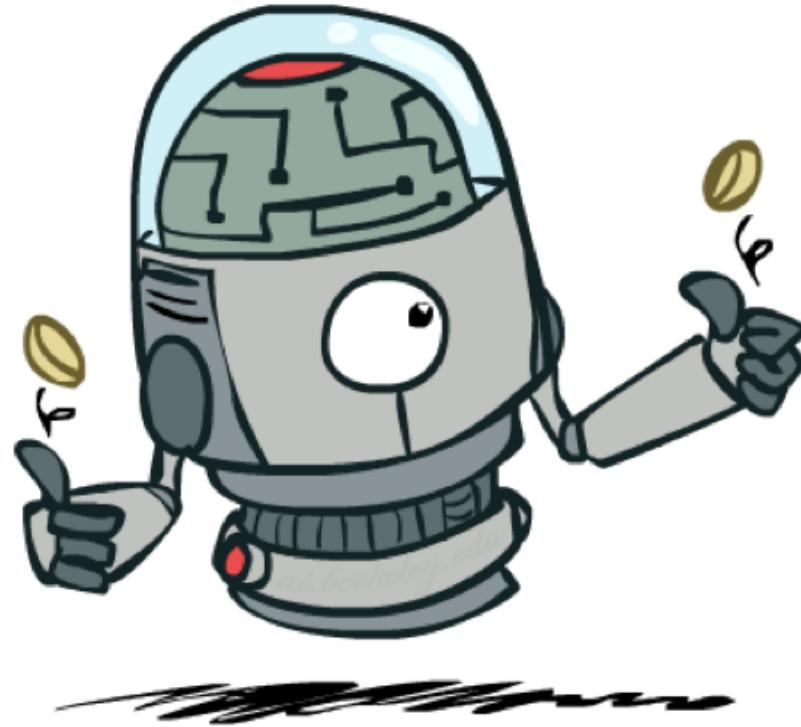
- May not account for every variable
- May not account for all interactions between variables
- “All models are wrong; but some are useful.”
– George E. P. Box

- What do we do with probabilistic models?

- We (or our agents) need to reason about unknown variables, given evidence
- Example: explanation (diagnostic reasoning)
- Example: prediction (causal reasoning)
- Example: value of information



Independence



Independence

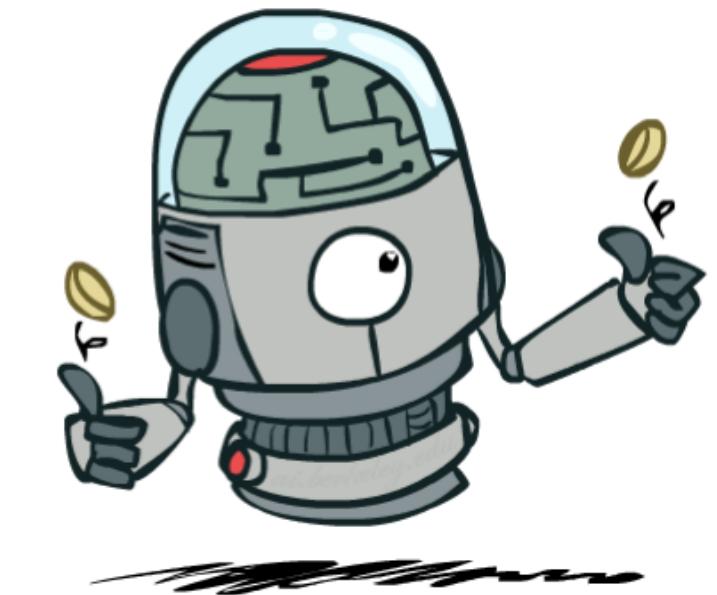
- Two variables are *independent* if:

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

- This says that their joint distribution *factors* into a product 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
 - What could we assume for {weather, traffic, cavity, toothache}?



Example: Independence?

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P(T)$

T	P
hot	0.5
cold	0.5

$P_2(T, W)$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

$P(W)$

W	P
sun	0.6
rain	0.4

Example: Independence

- N fair, independent coin flips:

 $P(X_1)$

H	0.5
T	0.5

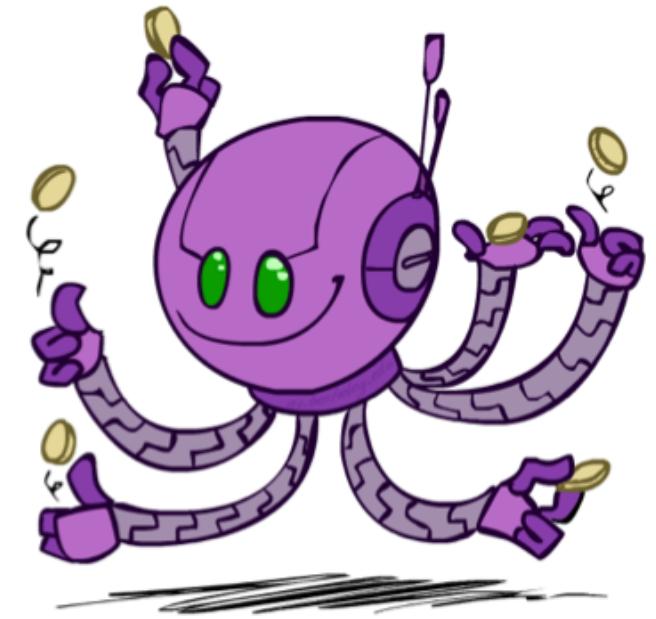
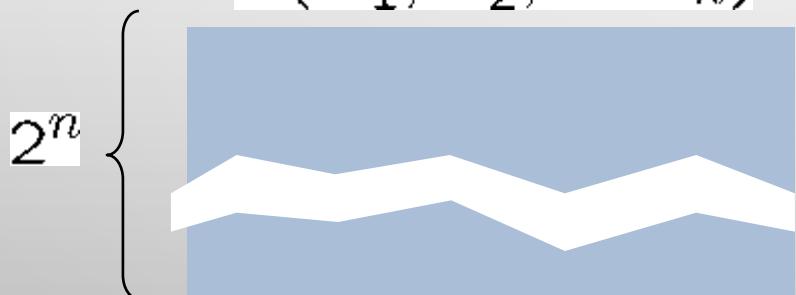
 $P(X_2)$

H	0.5
T	0.5

 $P(X_n)$

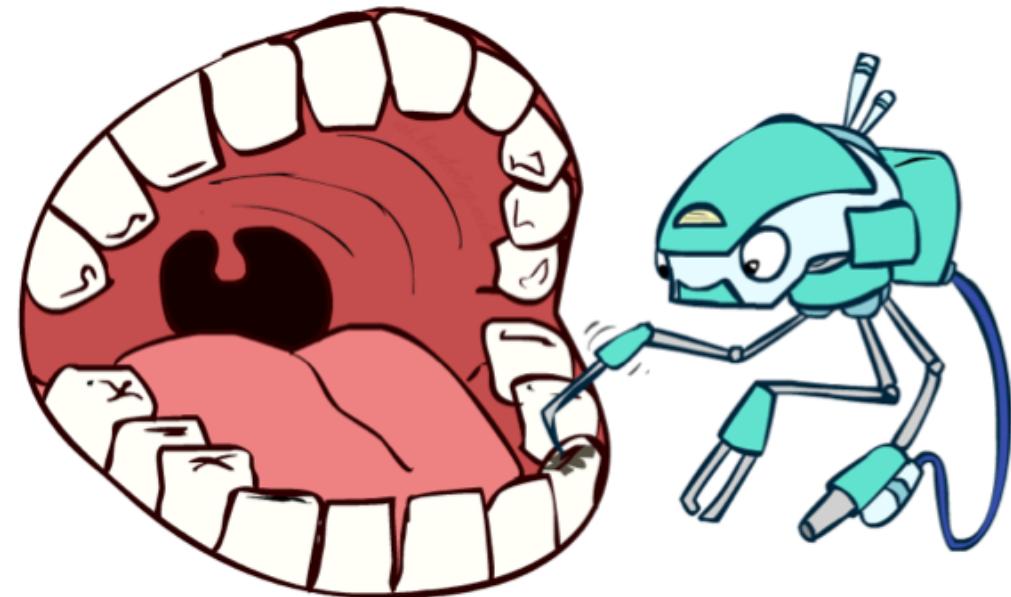
H	0.5
T	0.5

...

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

Conditional Independence

- $P(\text{toothache}, \text{cavity}, \text{catch})$
- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
 - $P(+\text{catch} | +\text{toothache}, +\text{cavity}) = p(+\text{catch} | +\text{cavity})$
- The same independence holds if I don't have a cavity:
 - $P(+\text{catch} | +\text{toothache}, -\text{cavity}) = p(+\text{catch} | -\text{cavity})$
- Catch is *conditionally independent* of Toothache given cavity:
 - $P(\text{Catch} | \text{Toothache, Cavity}) = p(\text{Catch} | \text{Cavity})$
- **Equivalent statements:**
 - $P(\text{Toothache} | \text{Catch, Cavity}) = P(\text{Toothache} | \text{Cavity})$
 - $P(\text{Toothache, Catch} | \text{Cavity}) = P(\text{Toothache} | \text{Cavity}) P(\text{Catch} | \text{Cavity})$
 - One can be derived from the other easily



Conditional Independence

- Unconditional (absolute) independence very rare (why?)
- *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:

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

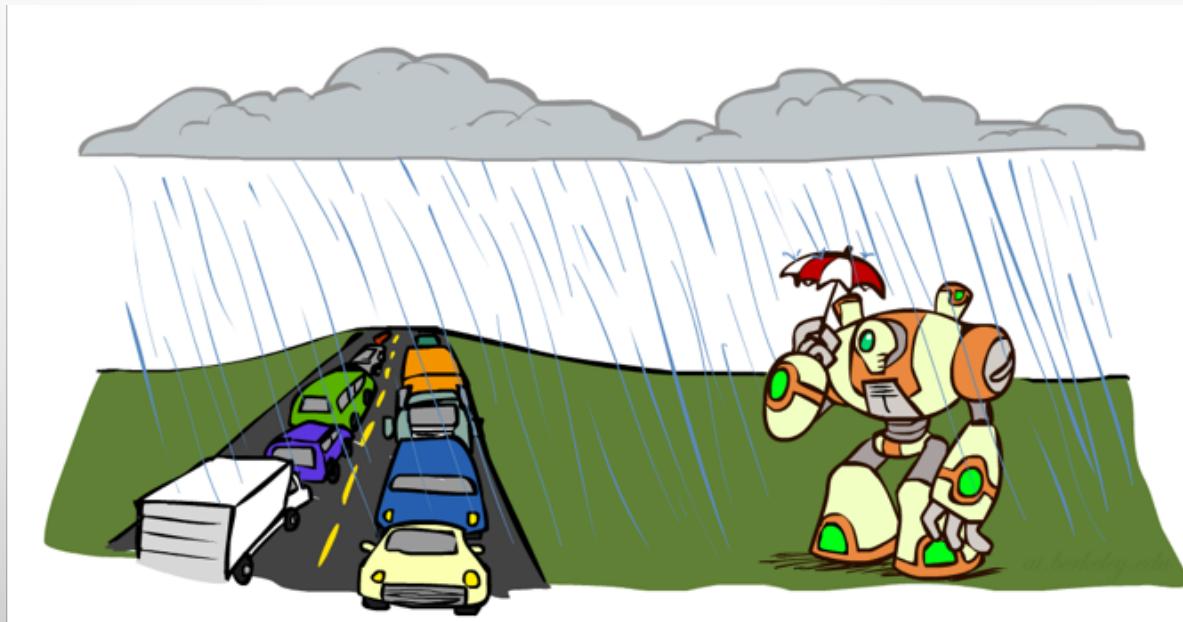
Or, equivalently, if and only if

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

Conditional Independence

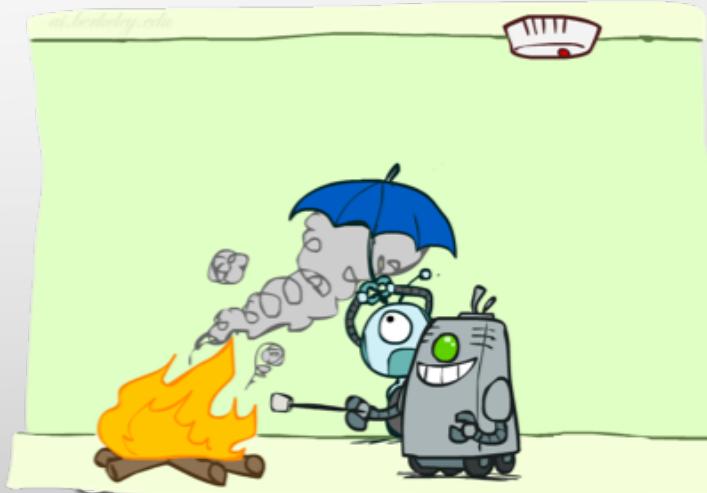
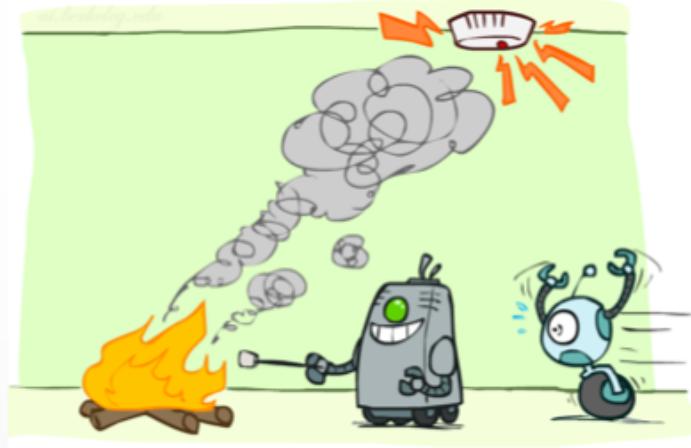
- What about this domain:

- Traffic
- Umbrella
- Raining



Conditional Independence

- What about this domain:
 - Fire
 - Smoke
 - Alarm



Conditional Independence and the Chain Rule

- Chain rule:

$$P(X_1, X_2, \dots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots$$

- Trivial decomposition:

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

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

- With assumption of conditional independence:

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

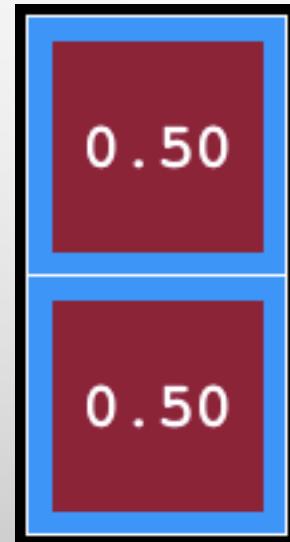
$$P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain})$$



- Bayes' nets / graphical models help us express conditional independence assumptions

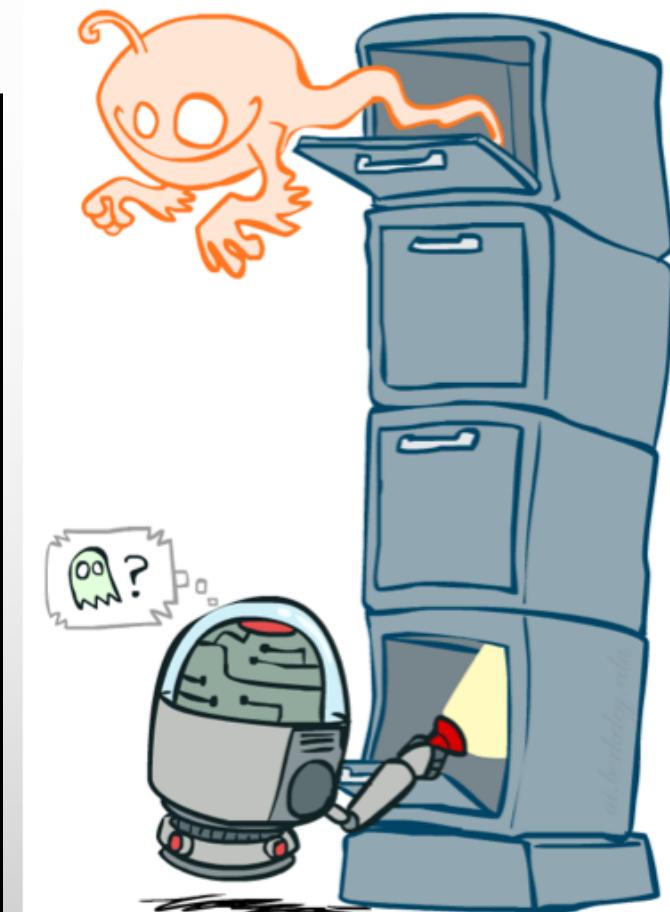
Ghostbusters Chain Rule

- Each sensor depends only on where the ghost is
- That means, the two sensors are conditionally independent, given the ghost position
- T: Top square is red
B: Bottom square is red
G: Ghost is in the top
- Givens:
 $P(+g) = 0.5$
 $P(+t | +g) = 0.8$
 $P(+t | -g) = 0.4$
 $P(+b | +g) = 0.4$
 $P(+b | -g) = 0.8$

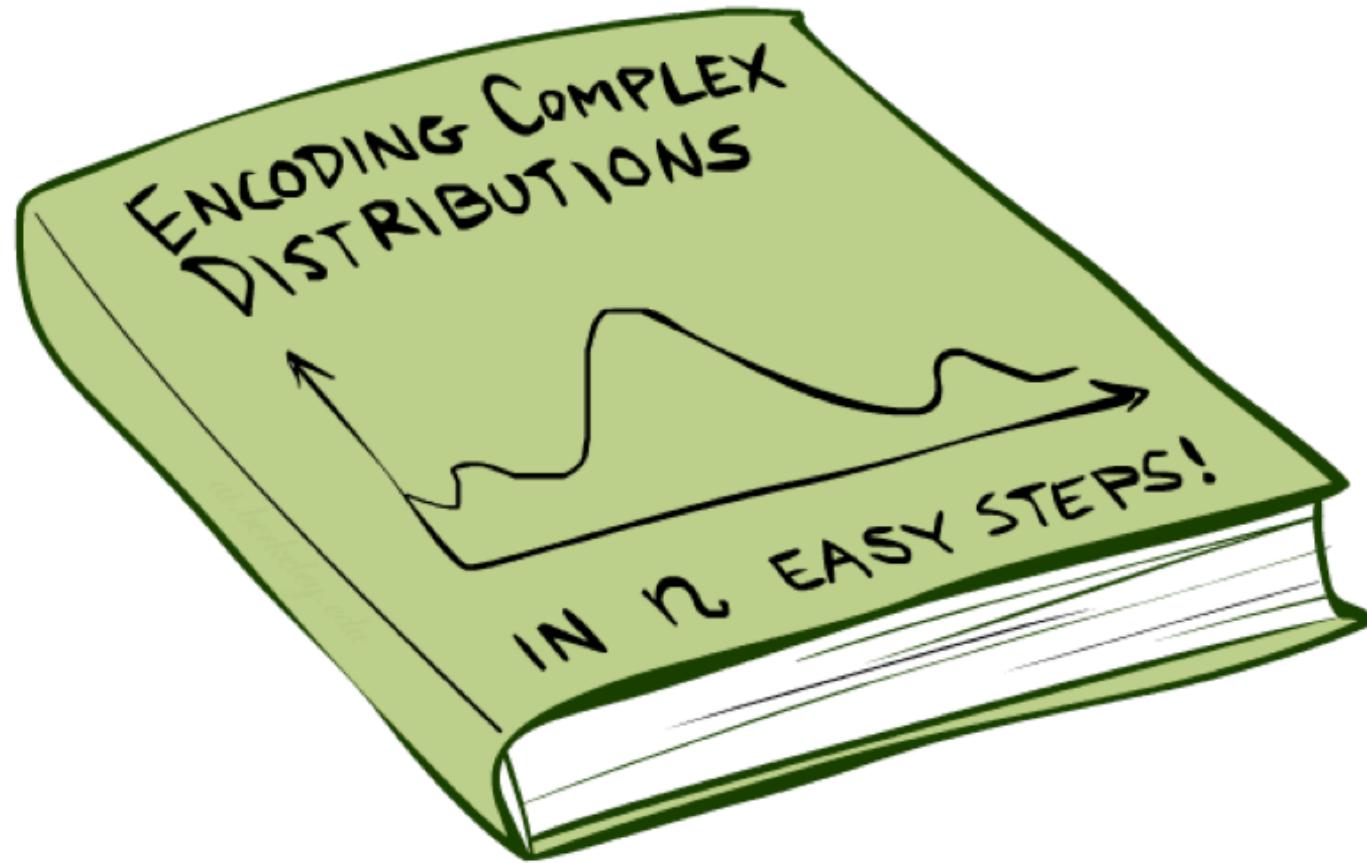


$$P(T, B, G) = P(G) P(T|G) P(B|G)$$

T	B	G	$P(T, B, G)$
+t	+b	+g	0.16
+t	+b	-g	0.16
+t	-b	+g	0.24
+t	-b	-g	0.04
-t	+b	+g	0.04
-t	+b	-g	0.24
-t	-b	+g	0.06
-t	-b	-g	0.06

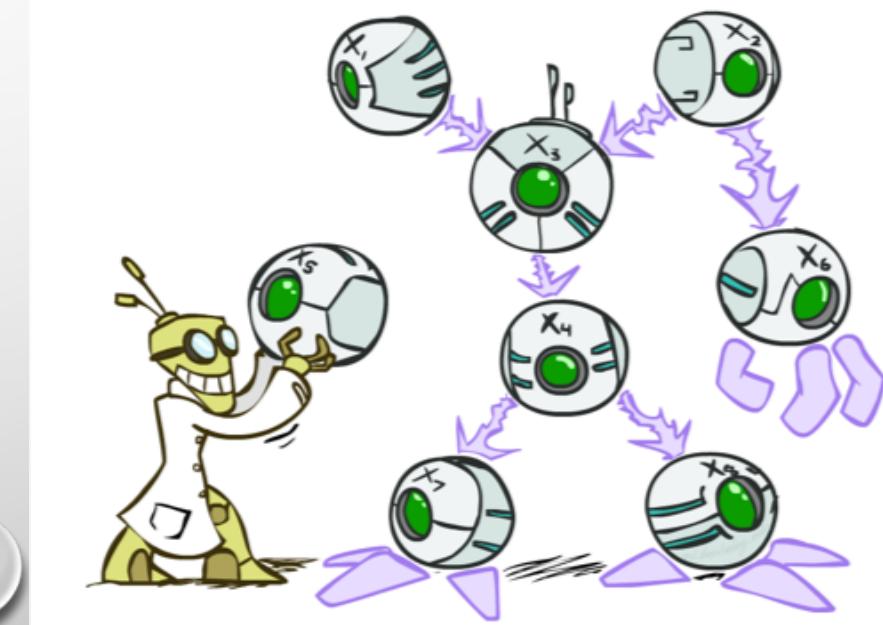


Bayes' Nets: Big Picture

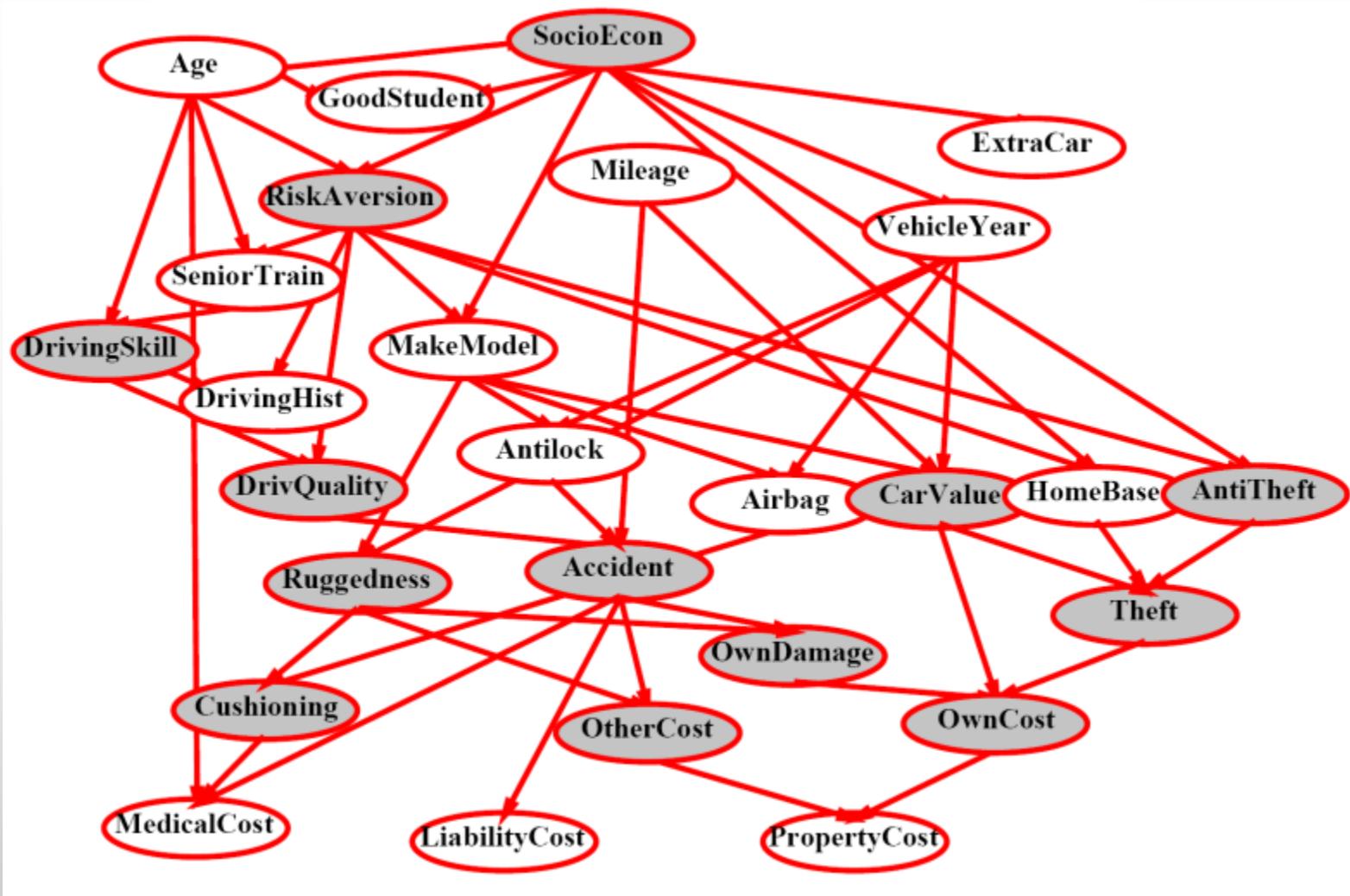


Bayes' Nets: Big Picture

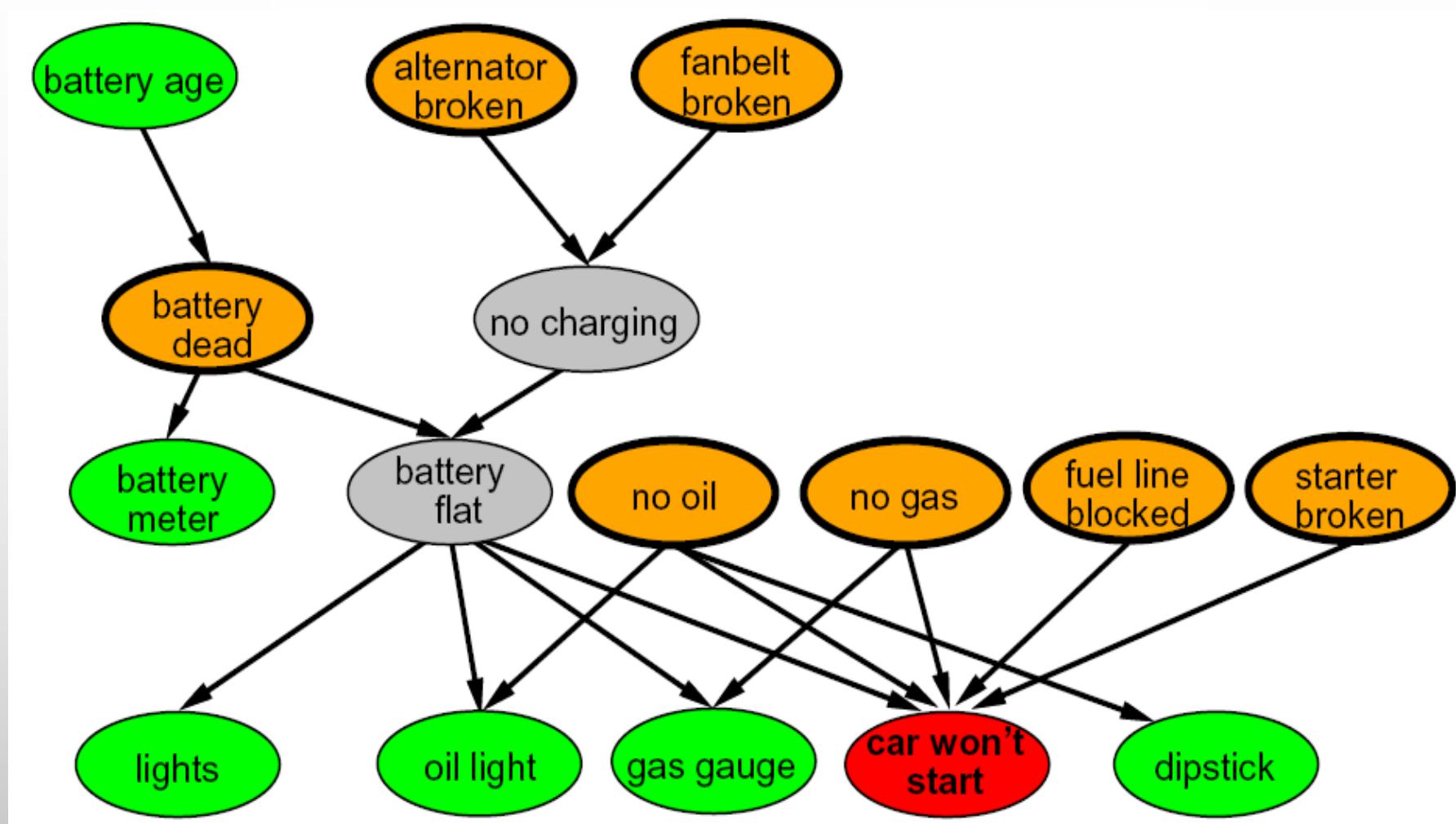
- Two problems with using full joint distribution tables as our probabilistic models:
 - Unless there are only a few variables, the joint is WAY too big to represent explicitly
 - Hard to learn (estimate) anything empirically about more than a few variables at a time
- **Bayes' nets:** a technique for 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
 - For about 10 min, we'll be vague about how these interactions are specified



Example Bayes' Net: Insurance



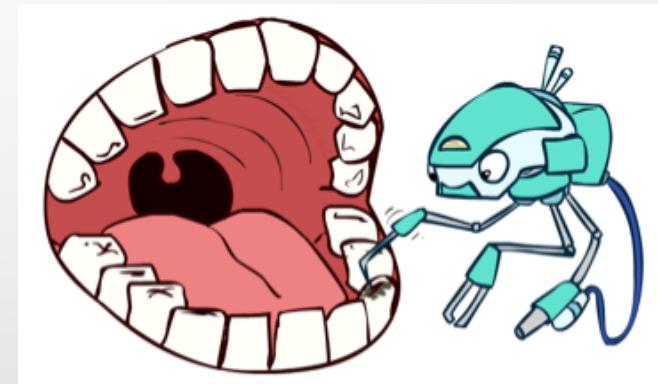
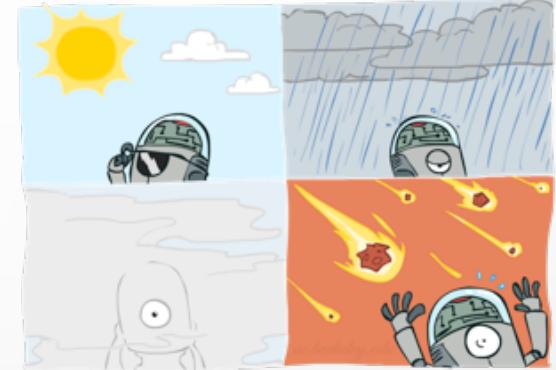
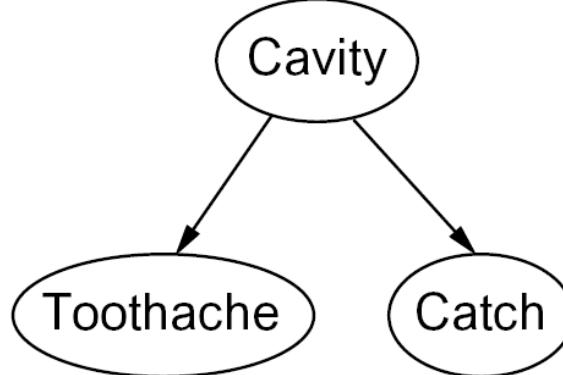
Example Bayes' Net: Car



Graphical Model Notation

- Nodes: variables (with domains)
 - Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions
 - Similar to CSP constraints
 - Indicate “direct influence” between variables
 - Formally: encode conditional independence (more later)
- For now: imagine that arrows mean direct causation (in general, they don’t!)

Weather



Example: Coin Flips

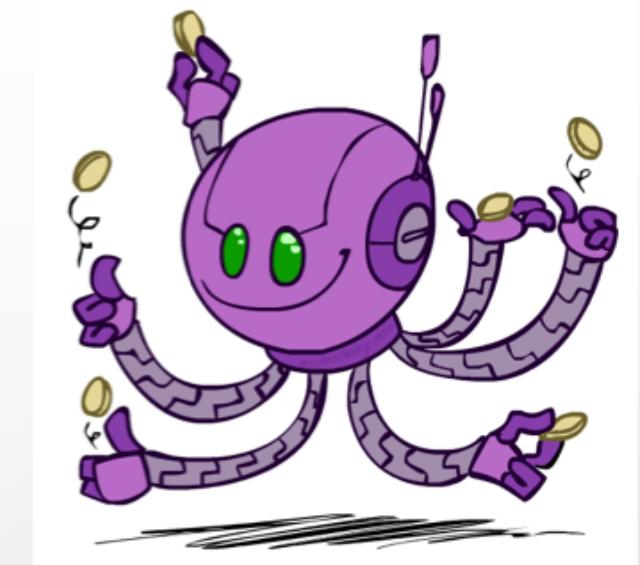
- N independent coin flips

X_1

X_2

...

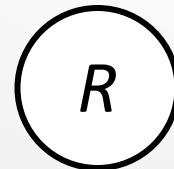
X_n



- No interactions between variables: **absolute independence**

Example: Traffic

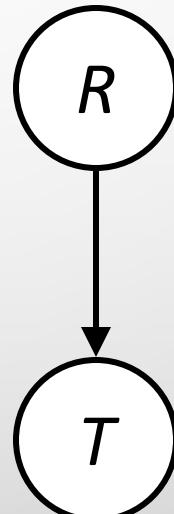
- Variables:
 - R: it rains
 - T: there is traffic



- Model 1: independence
- Why is an agent using model 2 better?



- Model 2: rain causes traffic



Example: Traffic II

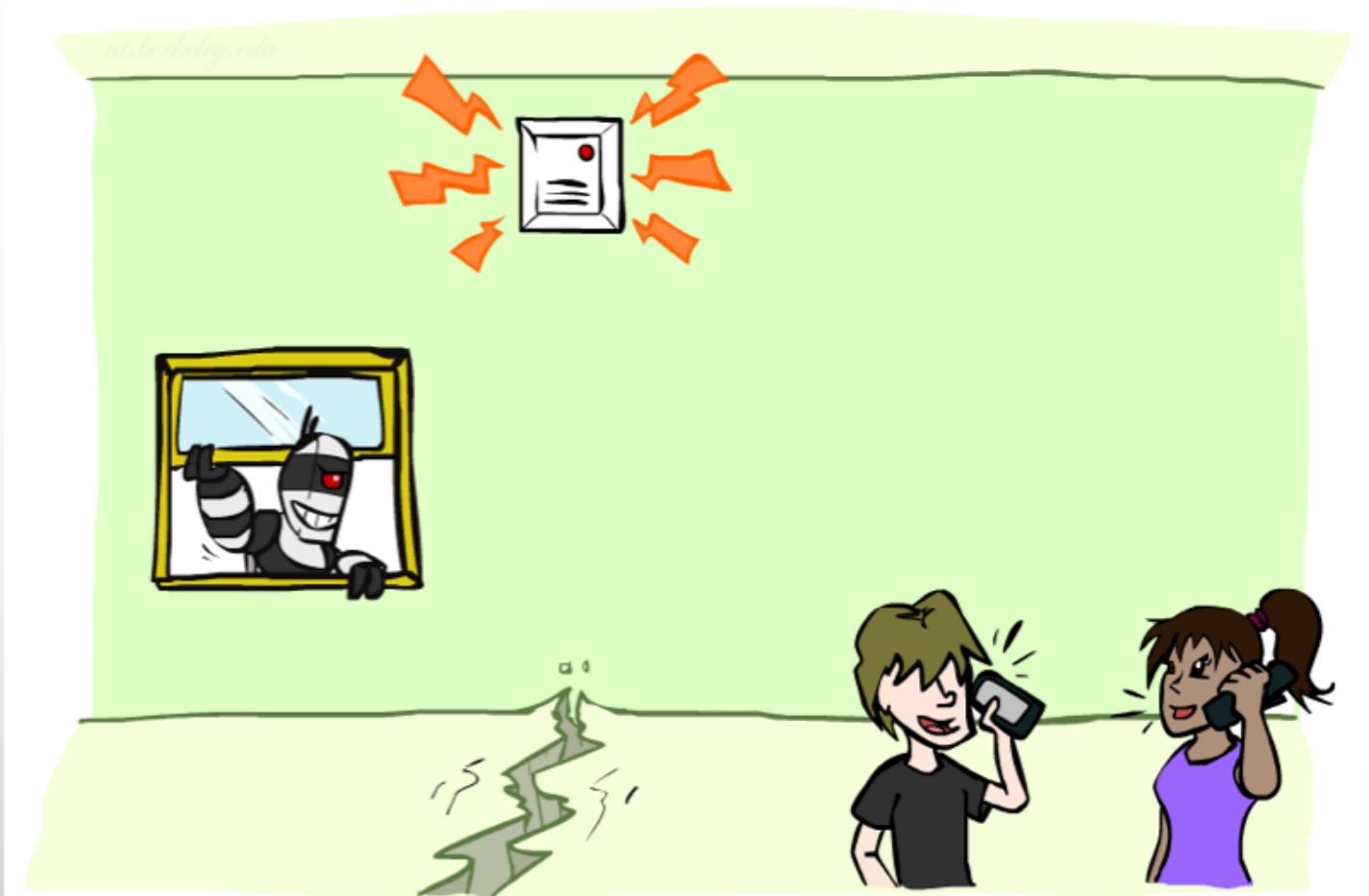
- Let's build a causal graphical model!
- Variables
 - T: traffic
 - R: it rains
 - L: low pressure
 - D: roof drips
 - B: ballgame
 - C: cavity



Example: Alarm Network

- Variables

- B: burglary
- A: alarm goes off
- M: Mary calls
- J: John calls
- E: earthquake!



Bayes' Net Semantics

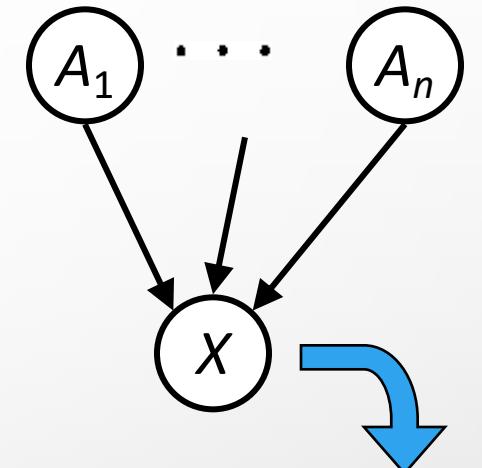


Bayes' Net Semantics



- 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

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



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

- CPT: conditional probability table
- Description of a noisy “causal” process

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

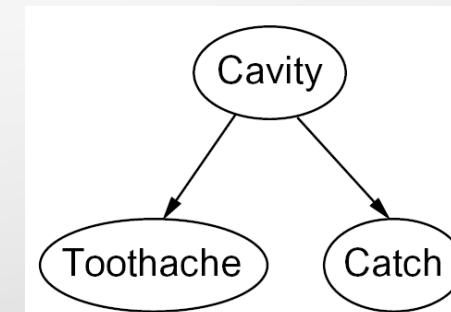
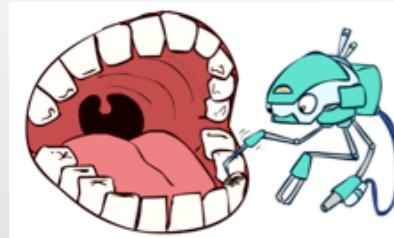
Probabilities in BNs



- 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:

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

- Example:



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

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):
- Assume conditional independences:

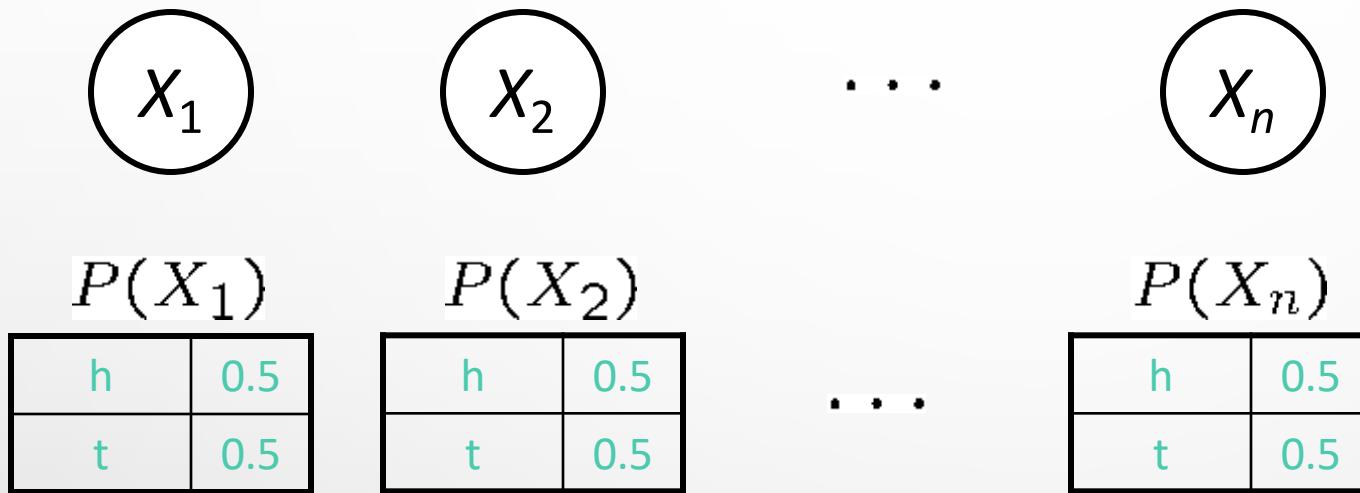
$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | x_1 \dots x_{i-1})$$
$$P(x_i | x_1, \dots, x_{i-1}) = P(x_i | \text{parents}(X_i))$$

→ Consequence:

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

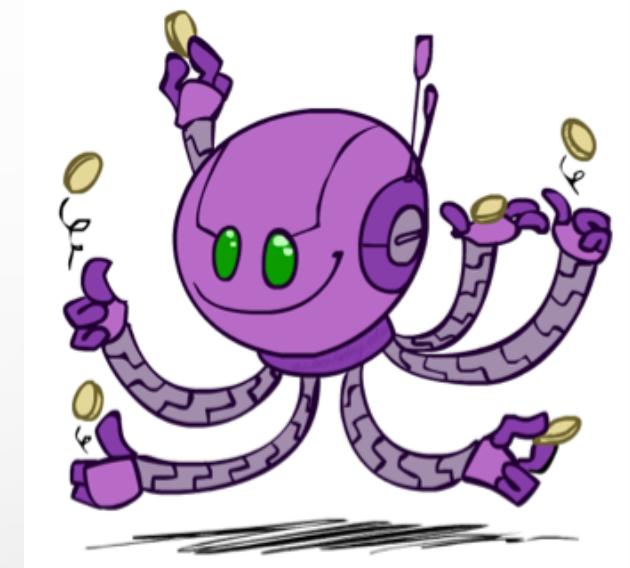
- Not every BN can represent every joint distribution
 - The topology enforces certain conditional independencies

Example: Coin Flips

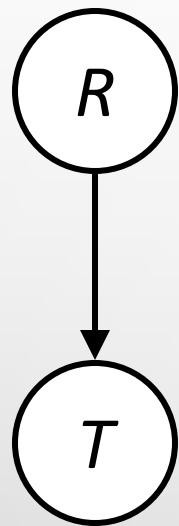


$$P(h, h, t, h) =$$

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



Example: Traffic



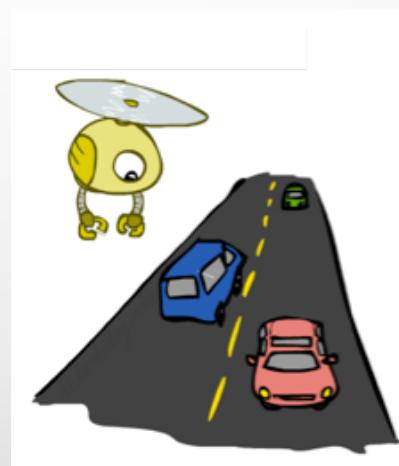
$P(R)$

$+r$	$1/4$
$-r$	$3/4$

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

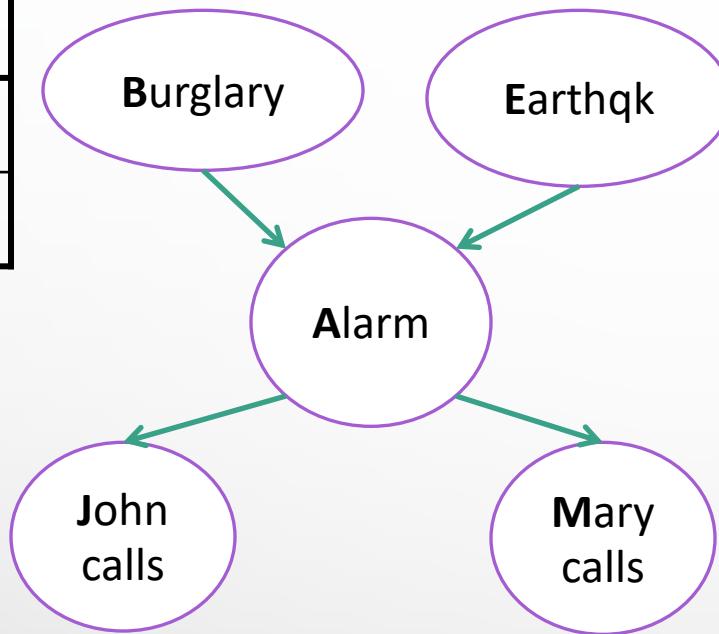
$P(T|R)$

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



Example: Alarm Network

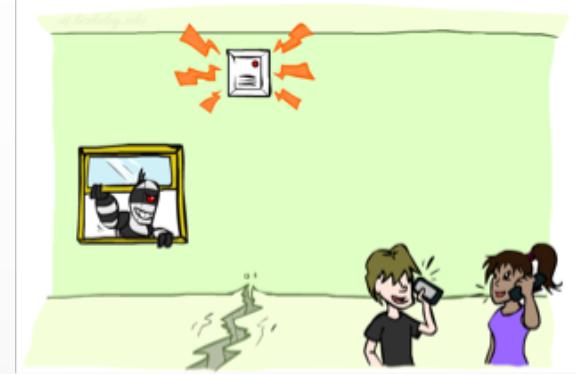
B	P(B)
+b	0.001
-b	0.999



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

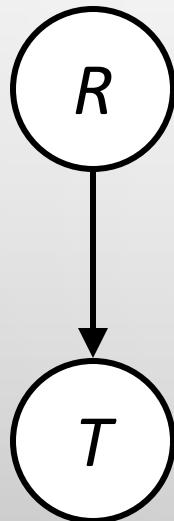
E	P(E)
+e	0.002
-e	0.998



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

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

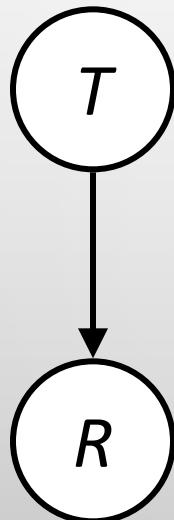
 $P(T, R)$

+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

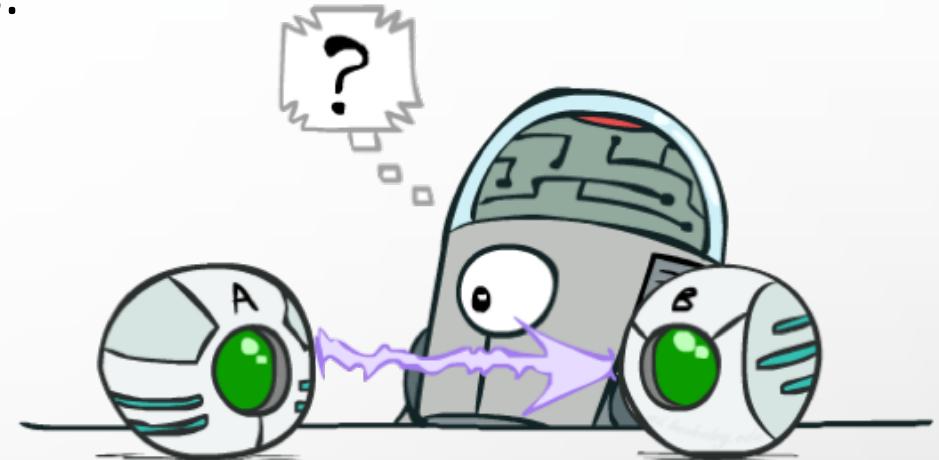
 $P(T, R)$

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

Causality?

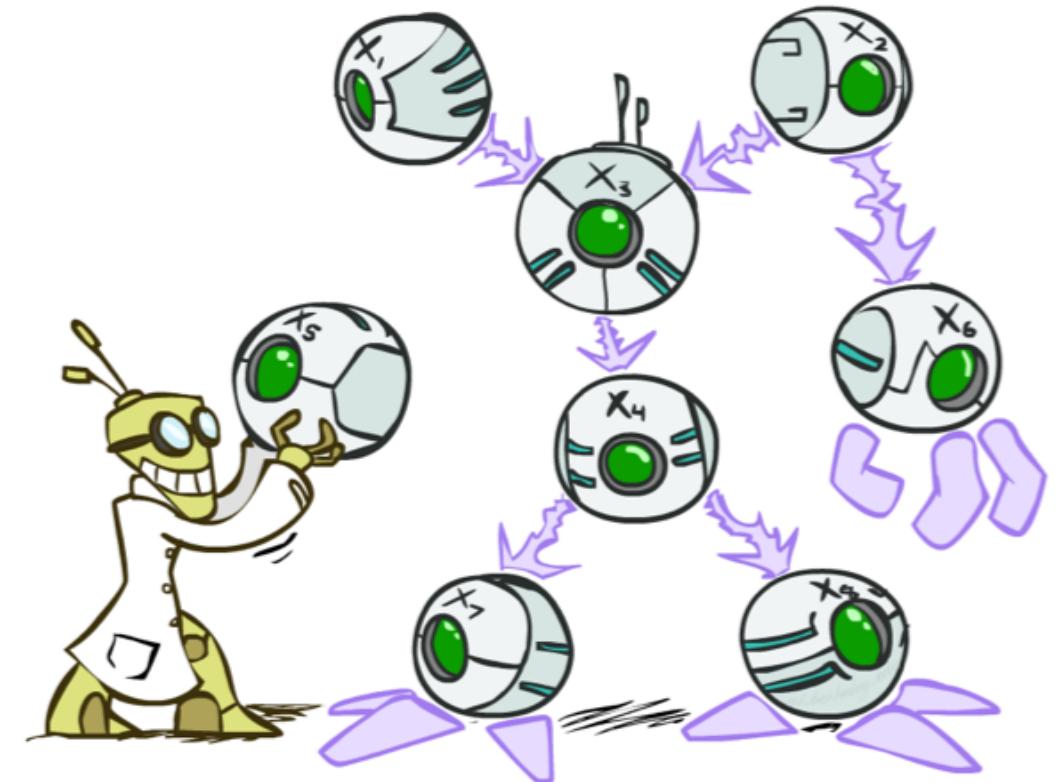
- When Bayes' nets reflect the true causal patterns:
 - Often simpler (nodes have fewer parents)
 - Often easier to think about
 - Often easier to elicit from experts
- BNs need not actually be causal
 - Sometimes no causal net exists over the domain (especially if variables are missing)
 - e.g. Consider 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**

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



Bayes' Nets

- So far: how a Bayes' net encodes a joint distribution
- Next: how to answer queries about that distribution
 - Today:
 - First assembled BNs using an intuitive notion of conditional independence as causality
 - Then saw that key property is conditional independence
 - Main goal: answer queries about conditional independence and influence
- After that: how to answer numerical queries (inference)



Size of a Bayes' Net

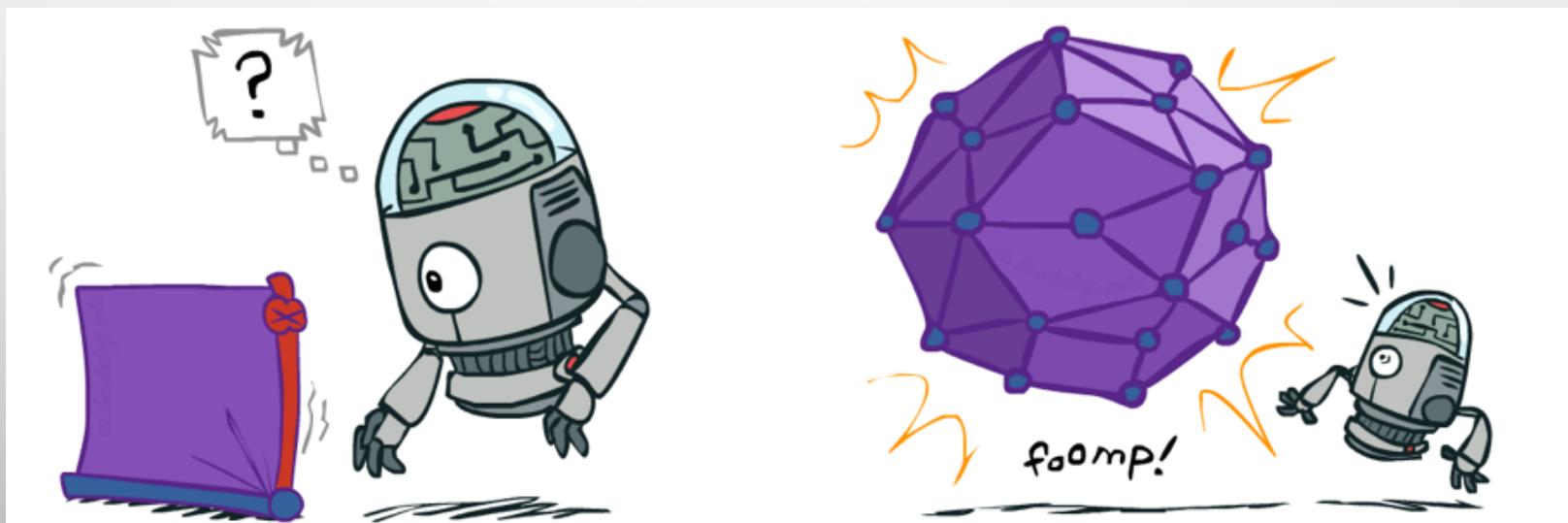
- How big is a joint distribution over N Boolean variables?

$$2^N$$

- 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

Bayes Nets: Assumptions

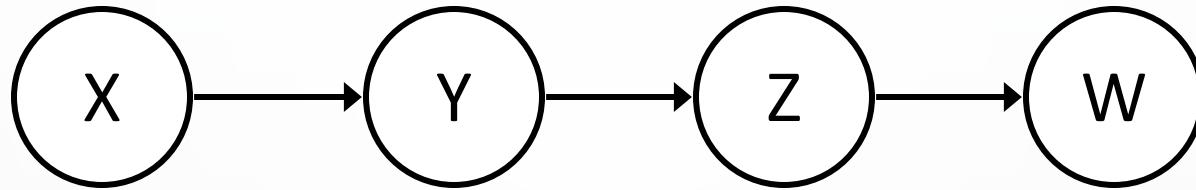
- Assumptions we are required to make to define the Bayes net when given the graph:

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

- 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



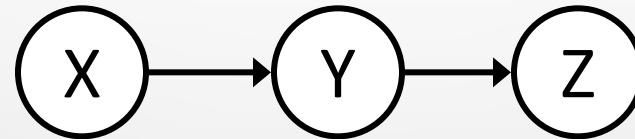
Example



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

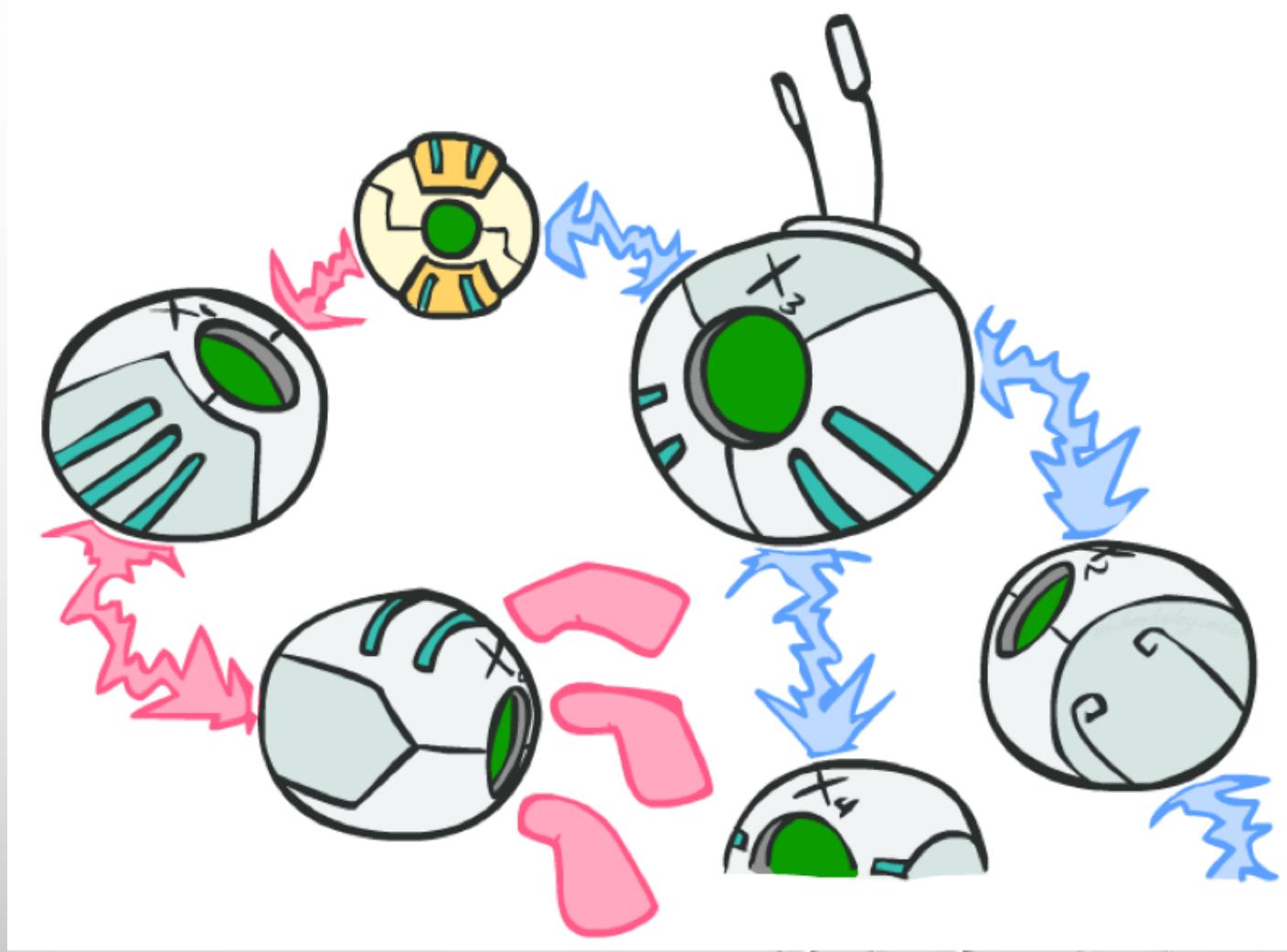
Independence in a BN

- Important question about a BN:
 - Are two nodes independent given certain evidence?
 - If yes, can prove using algebra (tedious in general)
 - 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



D-separation: Outline

- Study independence properties for triples
- Analyze complex cases in terms of member triples
- D-separation: a condition / algorithm for answering such queries

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!*

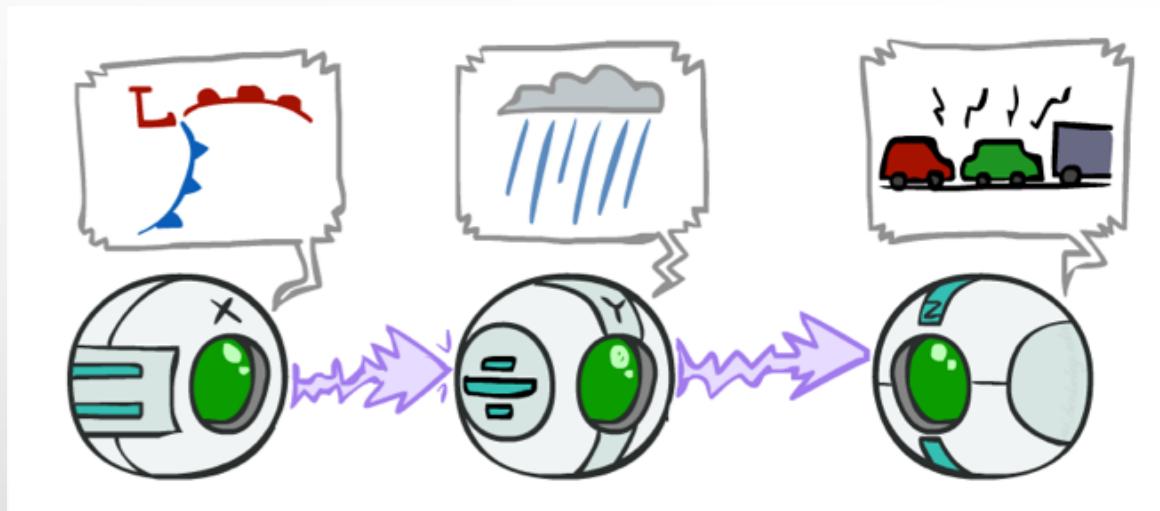
- One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
- Example:
 - Low pressure causes rain causes traffic, high pressure causes no rain causes no traffic
- In numbers:

$$\begin{aligned}P(+y | +x) &= 1, P(-y | -x) = 1, \\P(+z | +y) &= 1, P(-z | -y) = 1 \\P(+x) &= P(-x) = 0.5\end{aligned}$$

Causal Chains

- This configuration is a “causal chain”

- Guaranteed X independent of Z given Y?



X: Low pressure

Y: Rain

Z: Traffic

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

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

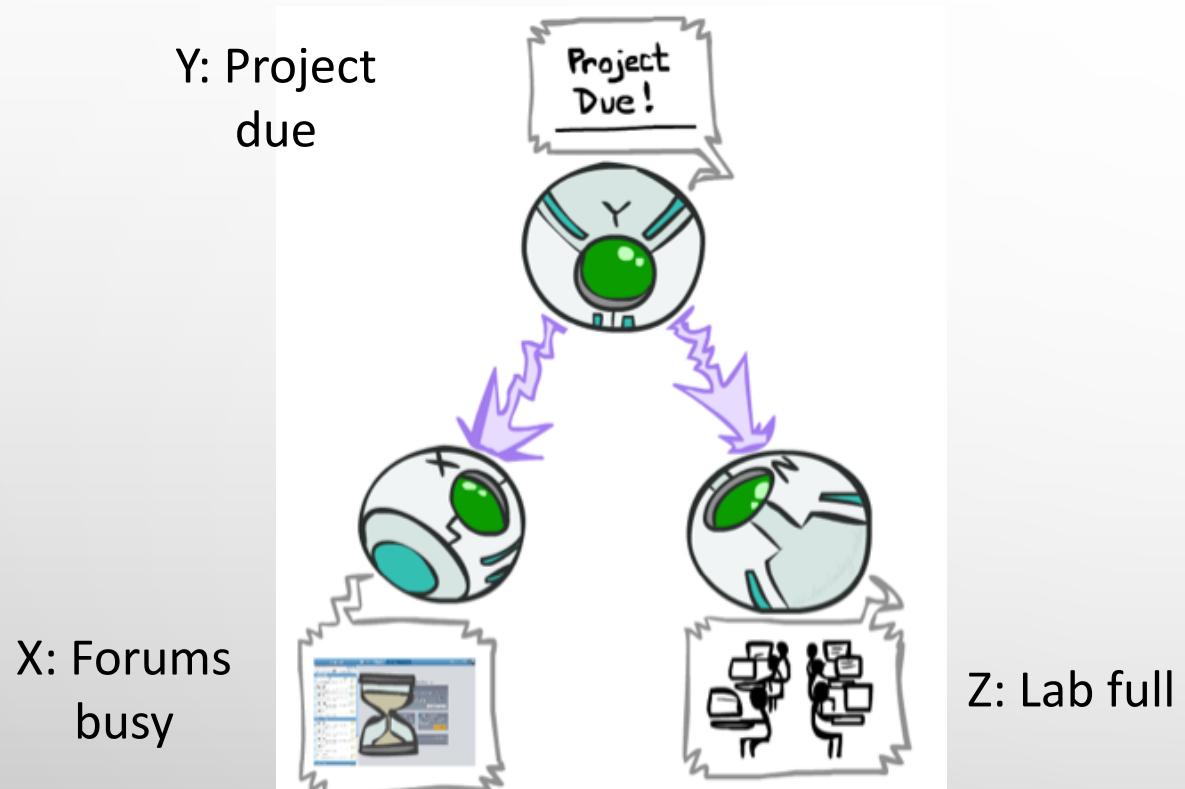
$$= P(z|y)$$

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:

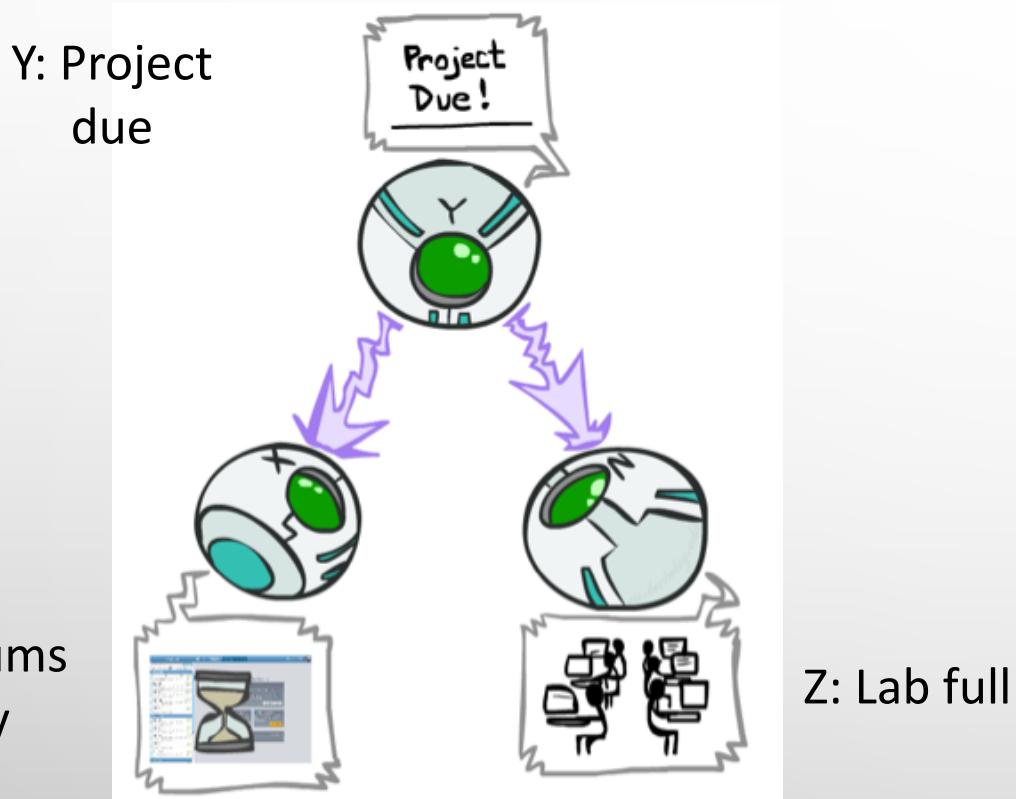
- Project due causes both forums busy and lab full

- In numbers:

$$\begin{aligned} P(+x | +y) &= 1, P(-x | -y) = 1, \\ P(+z | +y) &= 1, P(-z | -y) = 1 \\ P(+y) &= p(-y) = 0.5 \end{aligned}$$

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?

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

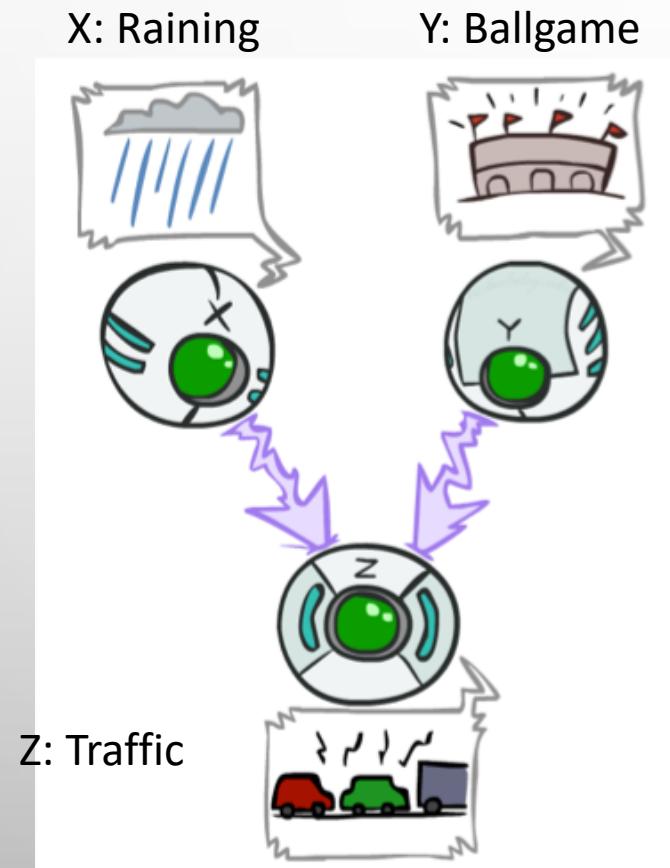
$$= P(z|y)$$

Yes!

- Observing the cause blocks influence between effects.

Common Effect

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



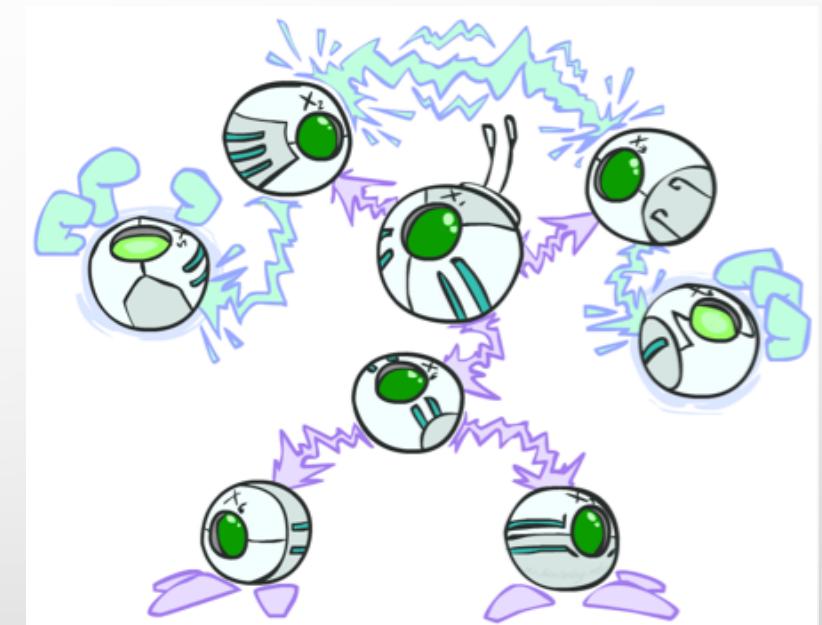
- Are X and Y independent?
 - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
 - Still need to prove they must be (try it!)
- Are X and Y independent given Z?
 - *No*: seeing traffic puts the rain and the ballgame in competition as explanation.
- This is backwards from the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

The General Case



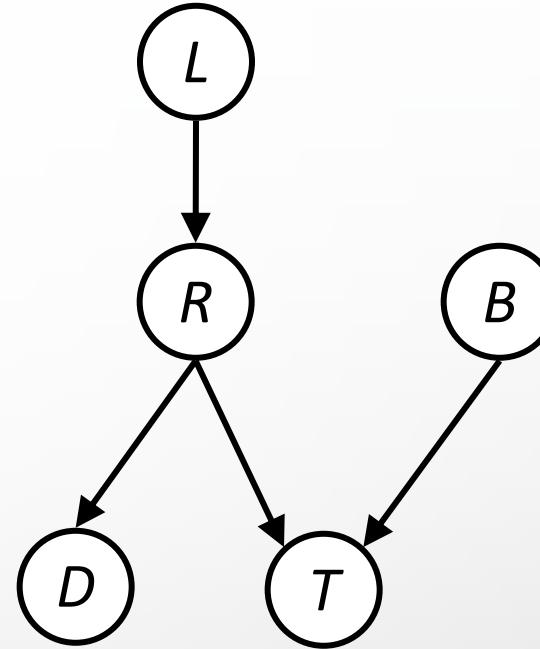
The General Case

- General question: in a given BN, are two variables independent (given 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!

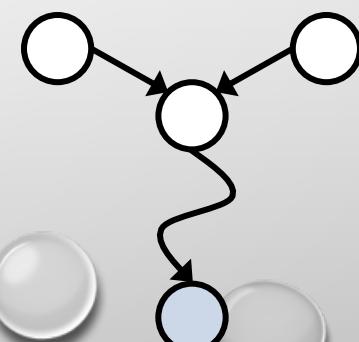
- A path is active if each triple is active:

- 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)

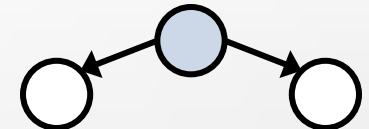
$A \rightarrow B \leftarrow C$ where B or one of its descendants is observed

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

Active Triples



Inactive Triples



D-Separation

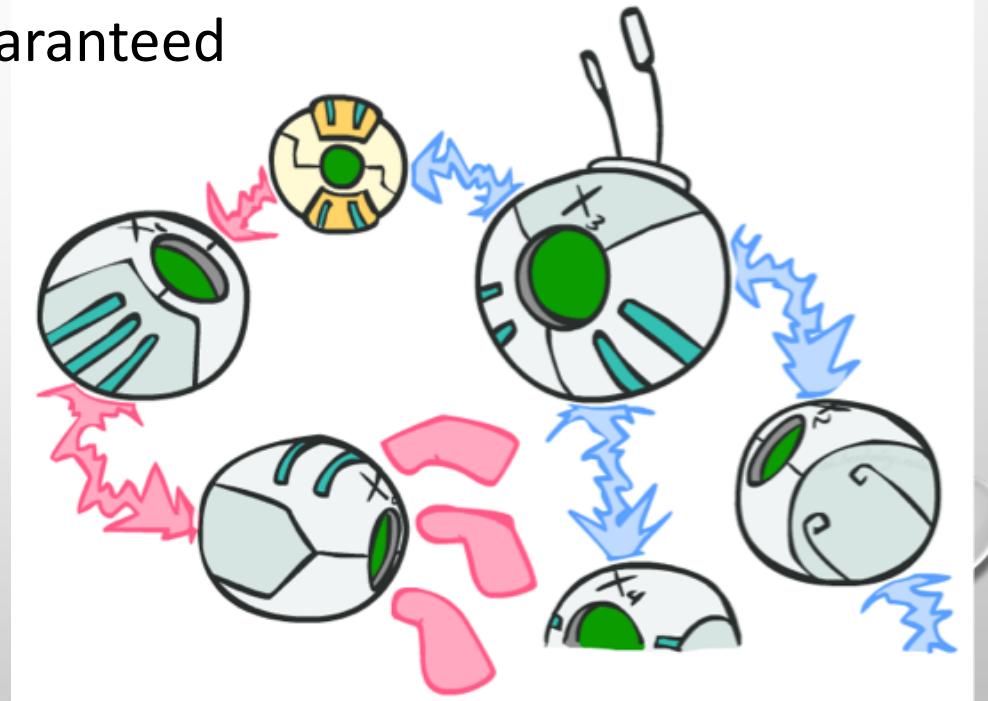
- Query: $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, \dots, X_{k_n}\}$?
- Check all (undirected!) Paths between X_i and X_j

- If one or more active, then independence not guaranteed

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

- Otherwise (i.e. If all paths are inactive),
Then independence is guaranteed

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



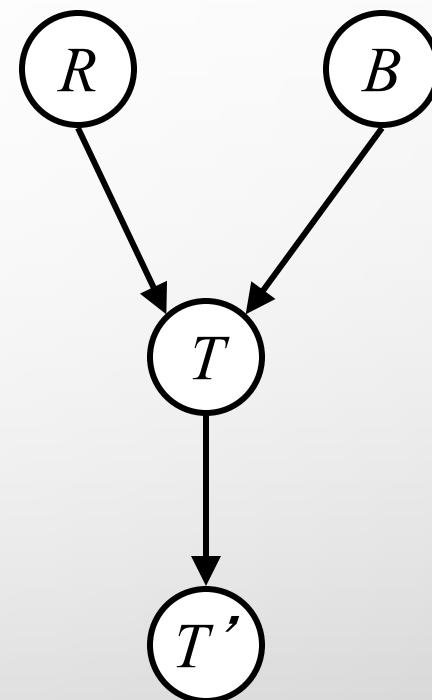
Example

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

Yes

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

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



Example

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

Yes

$$L \perp\!\!\!\perp B$$

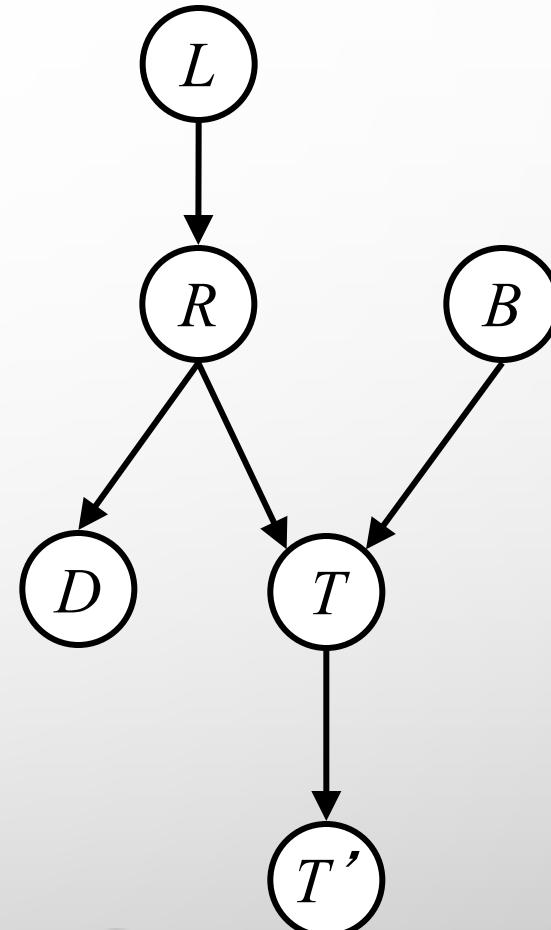
Yes

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

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

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

Yes



Example

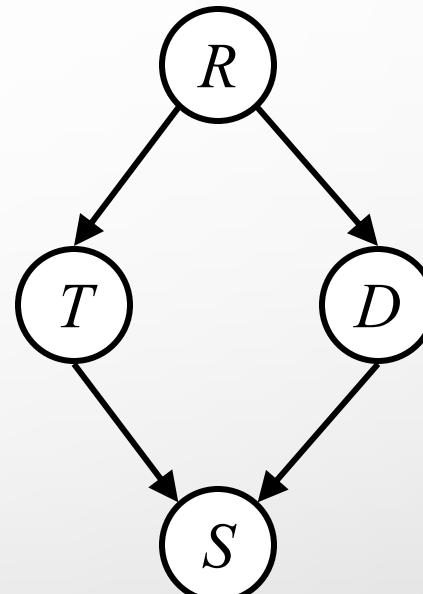
- Variables:
 - R: raining
 - T: traffic
 - D: roof drips
 - S: I'm sad

- Questions:

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

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

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

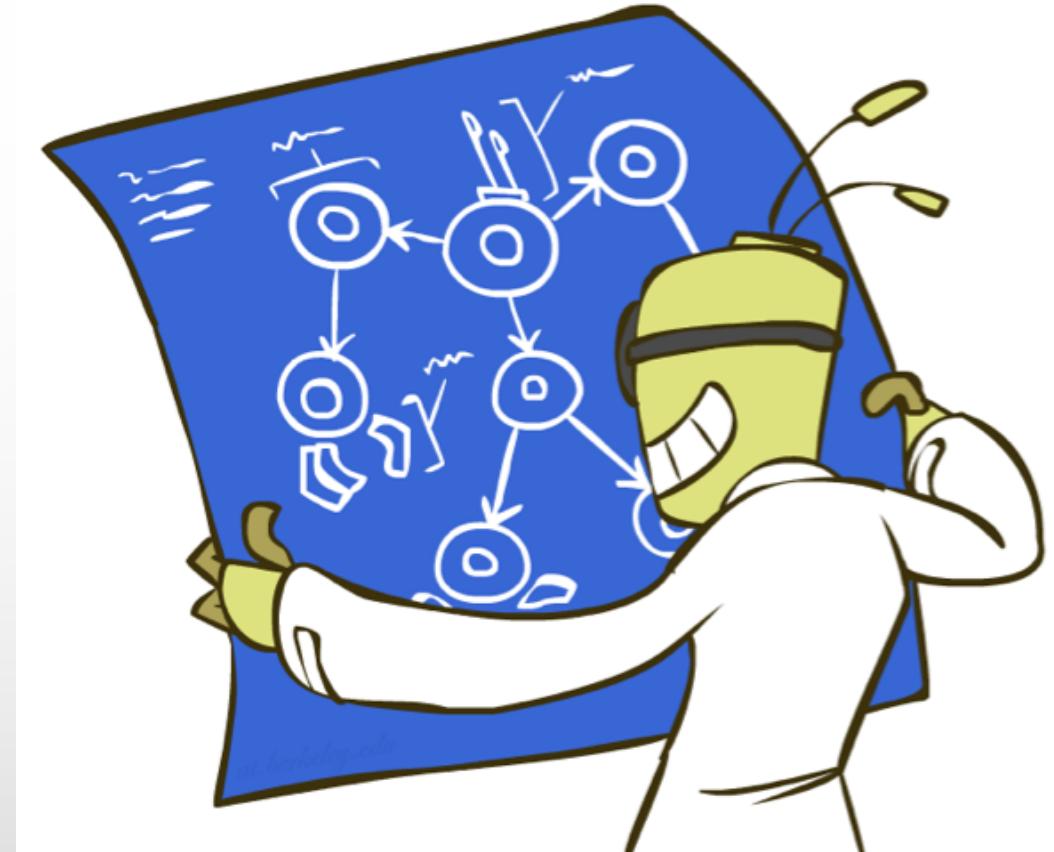


Structure Implications

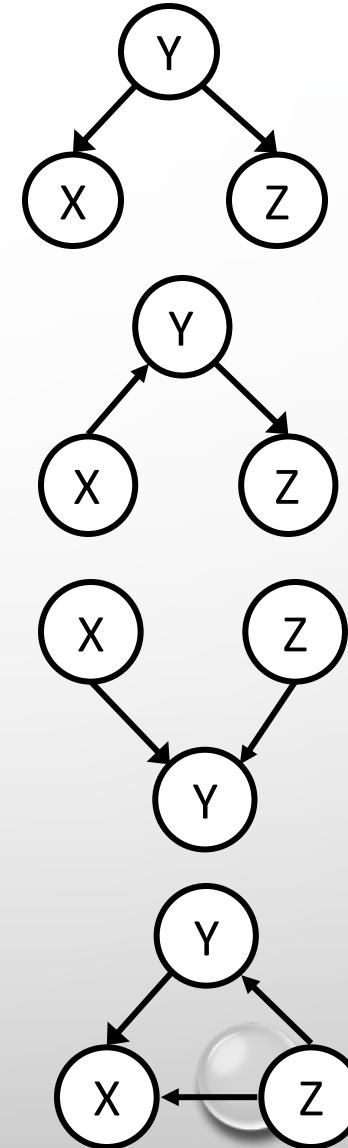
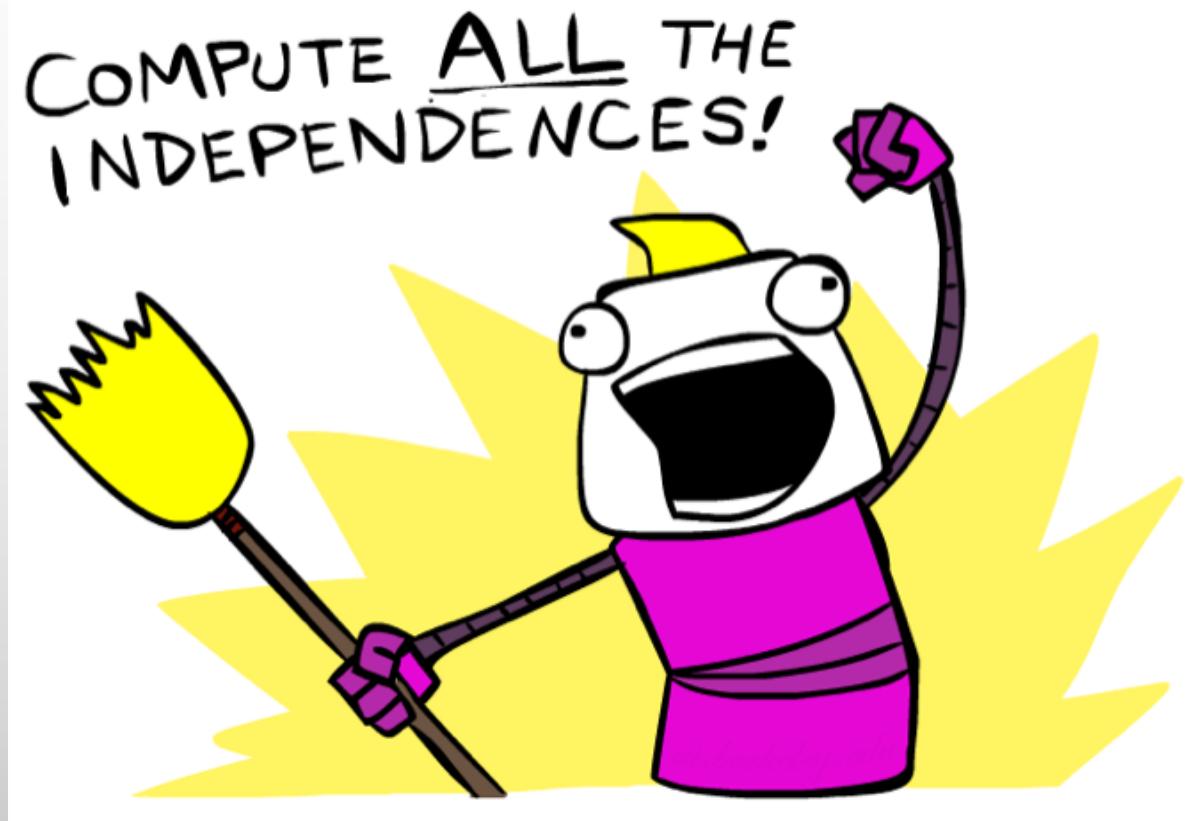
- Given a Bayes net structure, can run d-separation algorithm to build a complete list of conditional independences that are necessarily true of the form

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

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



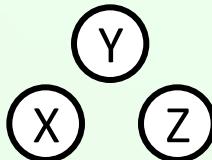
Computing All Independences



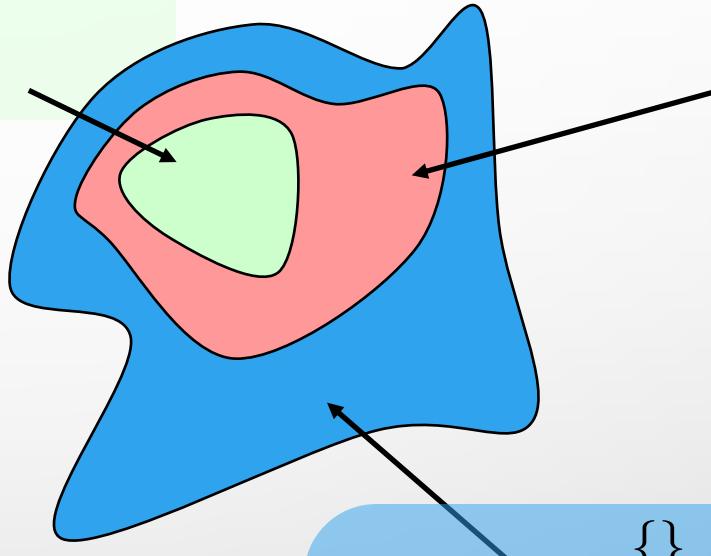
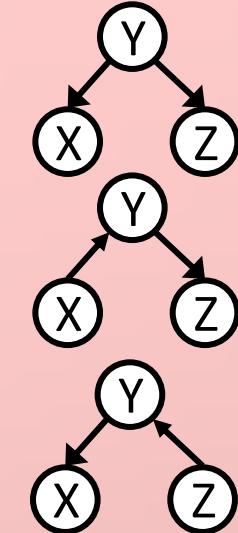
Topology Limits Distributions

- 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

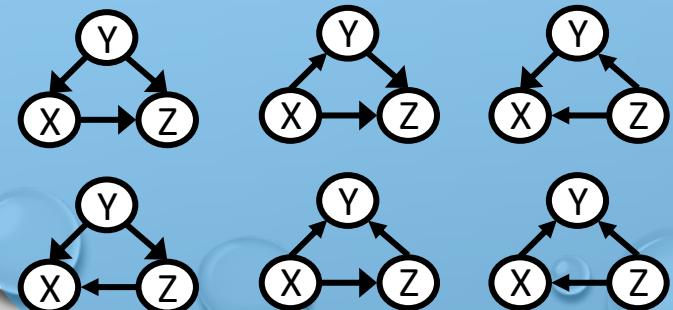
$$\{X \perp\!\!\!\perp Y, X \perp\!\!\!\perp Z, Y \perp\!\!\!\perp Z, X \perp\!\!\!\perp Z \mid Y, X \perp\!\!\!\perp Y \mid Z, Y \perp\!\!\!\perp Z \mid X\}$$



$$\{X \perp\!\!\!\perp Z \mid Y\}$$



{}



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