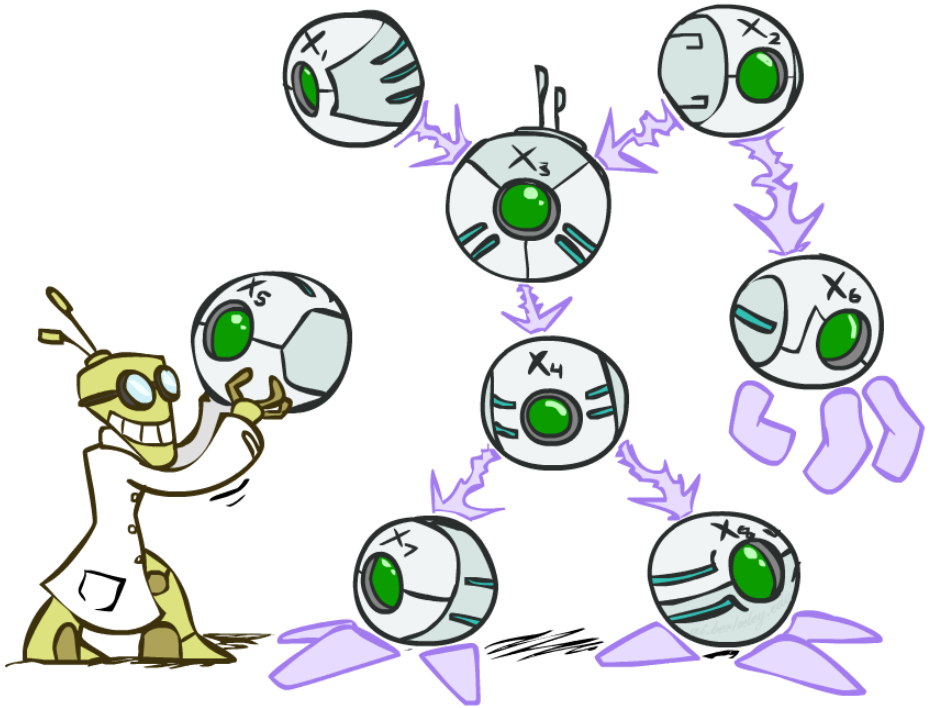
CS 4100 // artificial intelligence

instructor: byron wallace



Bayes Nets

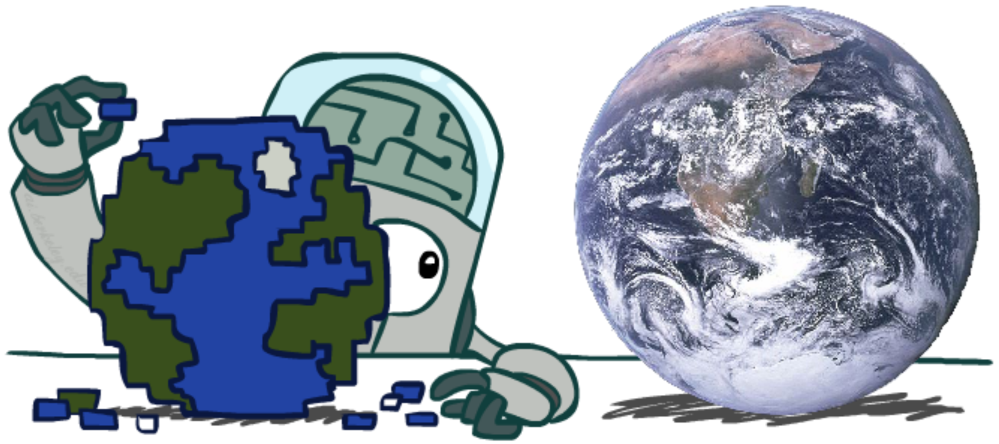
Attribution: many of these slides are modified versions of those distributed with the *UC Berkeley CS188* materials
Thanks to John DeNero and Dan Klein

Probabilistic models

Models describe how (a portion of) the world works

Models are always simplifications

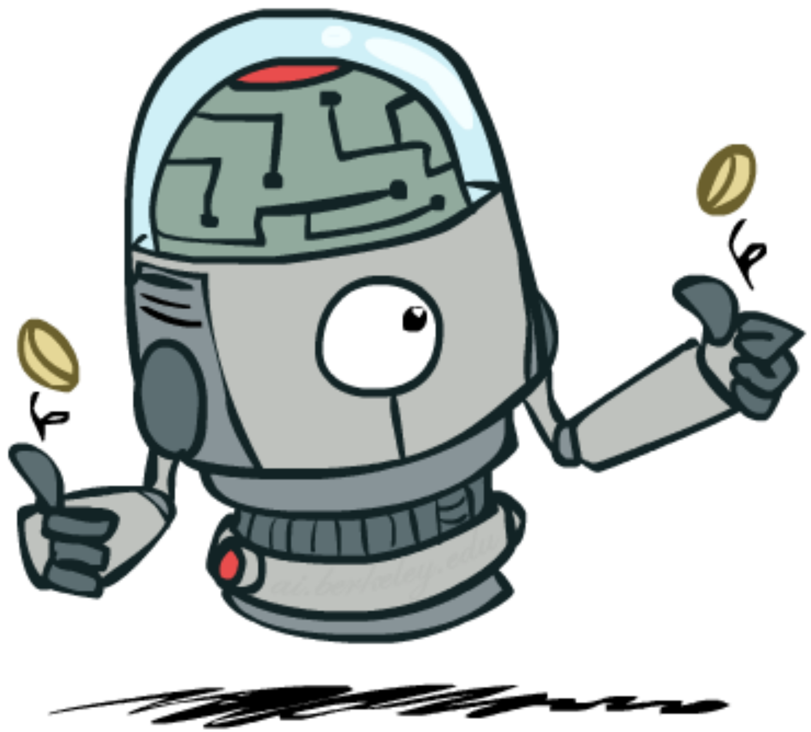
- May not account for every variable
- May not account for all interactions between variables
- "All models are wrong; but some are useful."
  – George E. P. Box



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

A brief review of independence



Independence

Two variables are independent if:

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

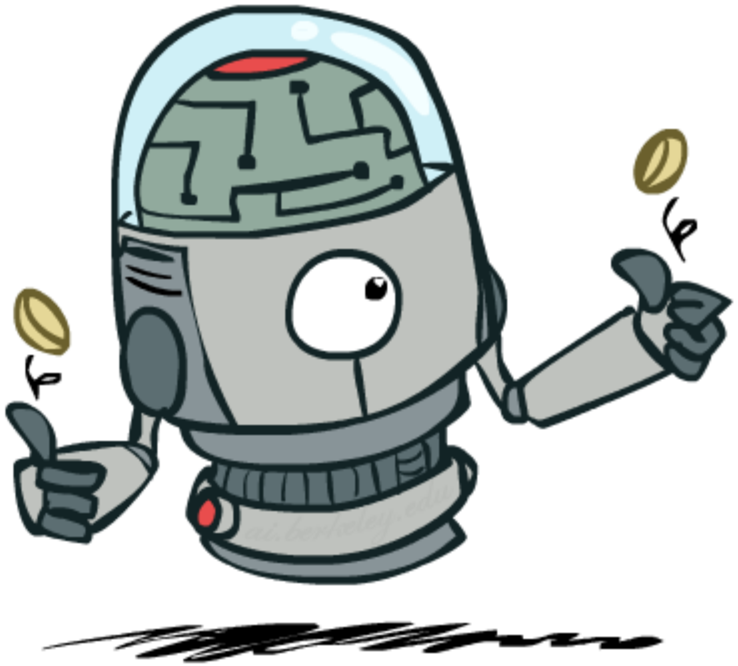
- This says that their joint distribution factors into a product two simpler distributions
- Another form:

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

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

Independence is a simplifying modeling assumption

- Empirical joint distributions: at best "close" to independent
- Recall: What could we assume for {Weather, Traffic, Cavity, Toothache}?



Verifying independence

T, W independent if all entries in P2 and P1 match!

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

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

joint distribution

$P(W)$

W	P
sun	0.6
rain	0.4

$P_2(T, W)$

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

= product of marginals
= P(T)P(W)

Conditional independence: the fundamental assumption for Bayes nets.

Conditional independence

P(Toothache, Cavity, Catch*) **catch* means probe finds a cavity

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

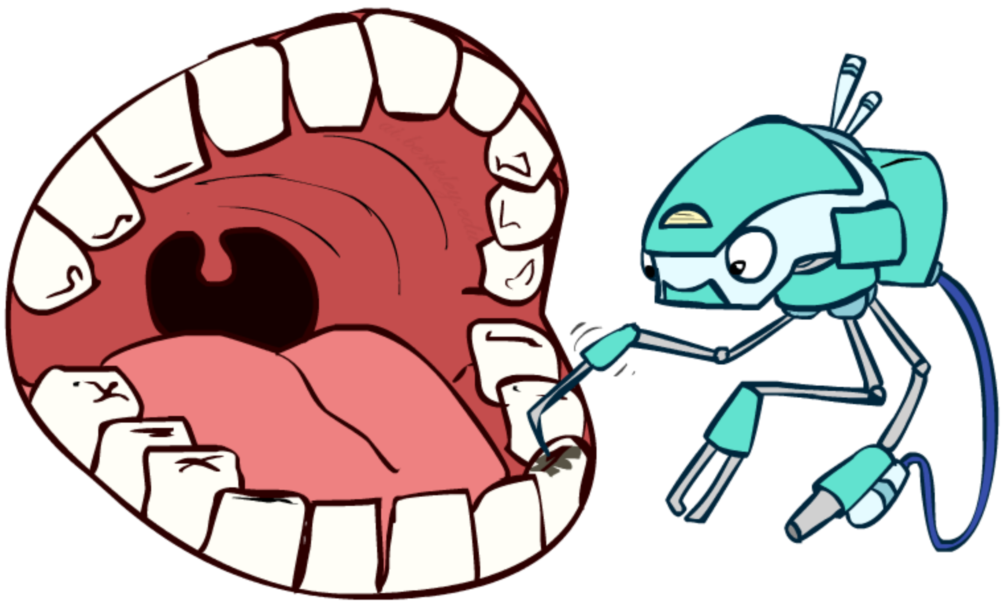
$$P(+catch \mid +toothache, +cavity) = P(+catch \mid +cavity)$$

The same independence holds if I don’t have a cavity:

$$P(+catch \mid +toothache, -cavity) = P(+catch \mid -cavity)$$

Catch is *conditionally independent* of Toothache given Cavity:

$$P(Catch \mid Toothache, Cavity) = P(Catch \mid Cavity)$$

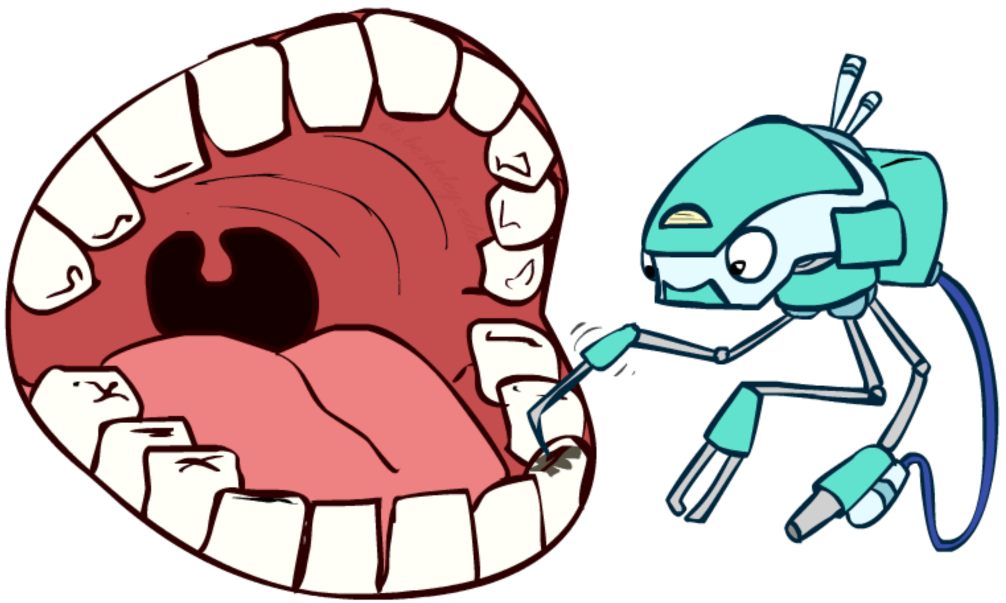


Conditional independence

P(Toothache, Cavity, Catch*) **catch* means probe finds a cavity

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

$P(+catch \mid +toothache, +cavity) = P(+catch \mid +cavity)$



The same independence holds if I don't have a cavity:

$P(+catch \mid +toothache, -cavity) = P(+catch \mid -cavity)$

Catch is *conditionally independent* of Toothache given Cavity:

$P(Catch \mid Toothache, Cavity) = P(Catch \mid Cavity)$

Equivalent statements:

1. $P(Toothache \mid Catch , Cavity) = P(Toothache \mid Cavity)$
2. $P(Toothache, Catch \mid Cavity) = P(Toothache \mid Cavity) P(Catch \mid Cavity)$
3. One can be derived from the other easily

Conditional independence (formal def)

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, written

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

if and only if:

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z) \qquad (1)$$

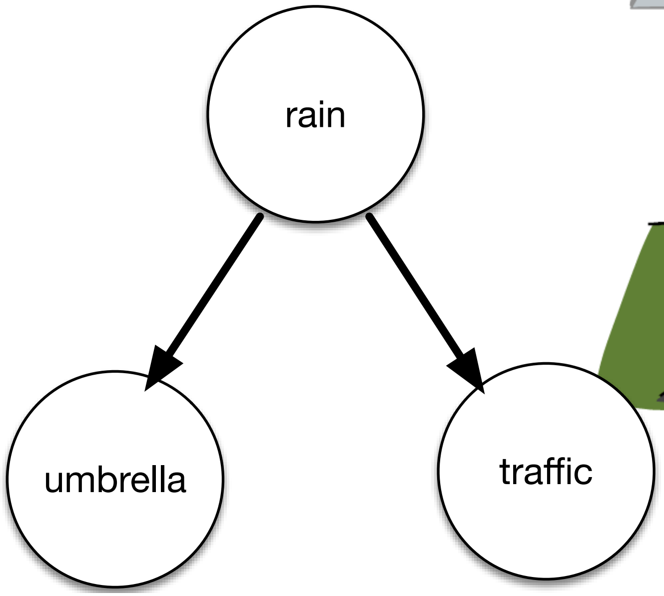
or, equivalently, if and only if

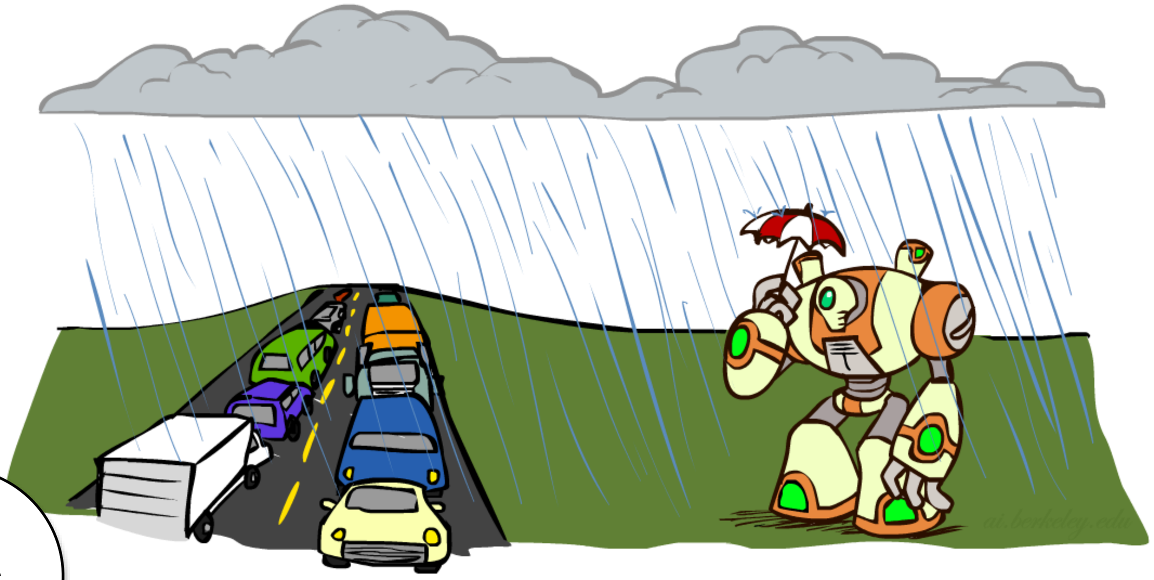
$$\forall x, y, z : P(x|z, y) = P(x|z) \qquad (2)$$

Conditional independence

And then we had this example:

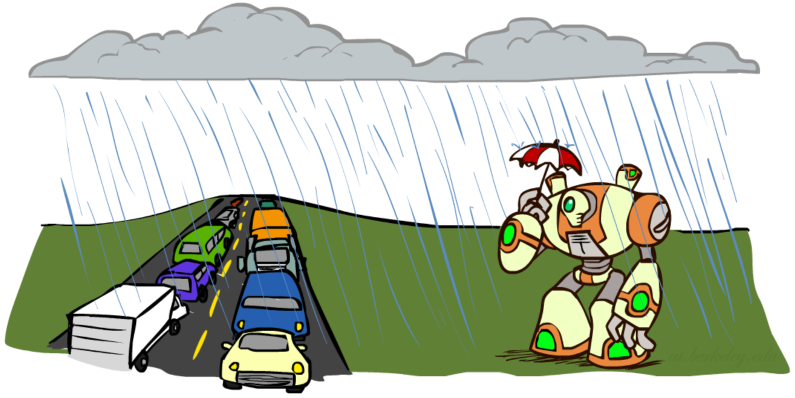
- Traffic
- Umbrella
- Raining





Conditional independence and the chain rule

Chain rule: $P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$

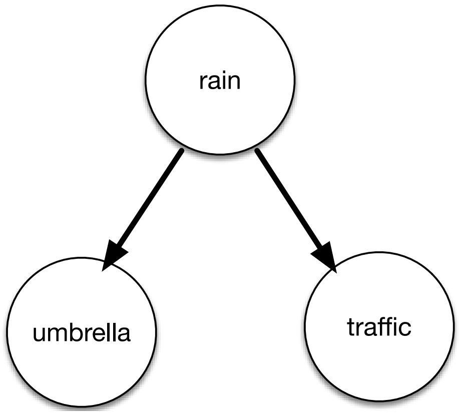


Trivial decomposition:

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

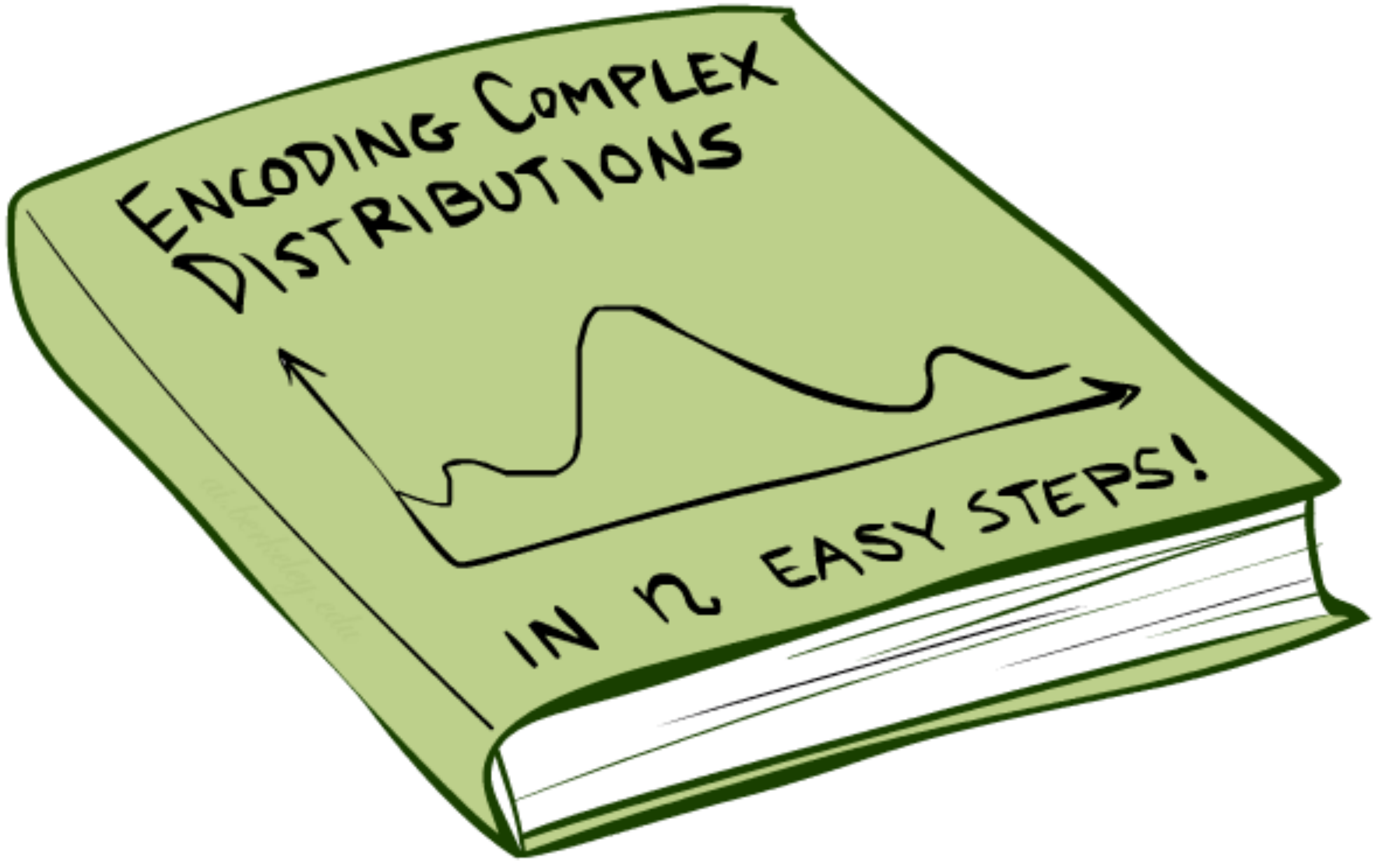
With assumption of conditional independence:

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



Bayes'nets / graphical models codify conditional independence assumptions

Bayes' Nets: the big picture



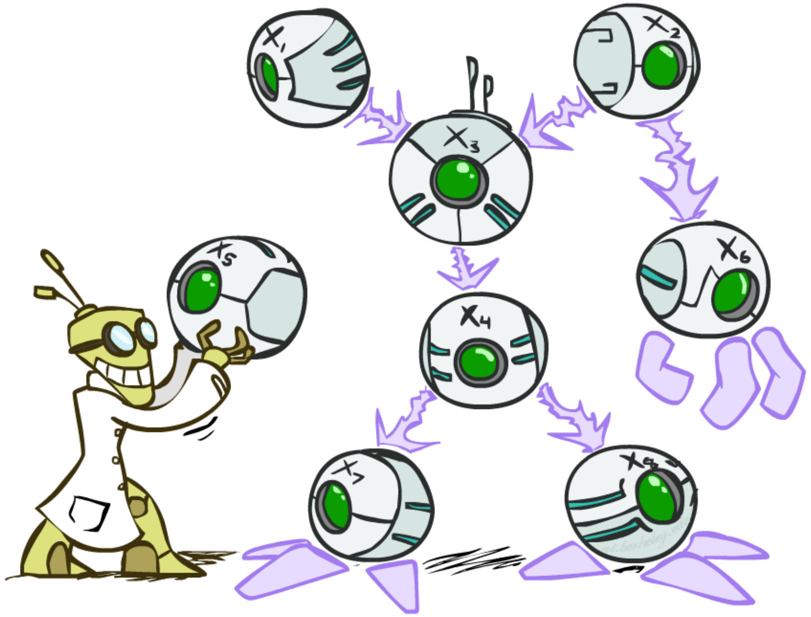
Bayes' Nets: the big picture

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



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?

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

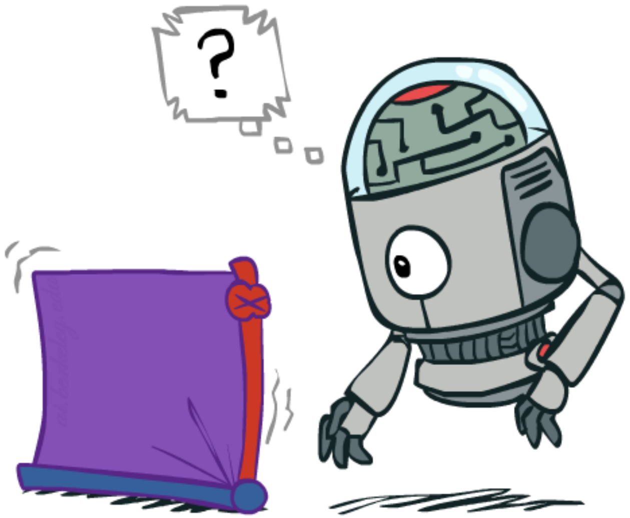
Both give you the power to calculate

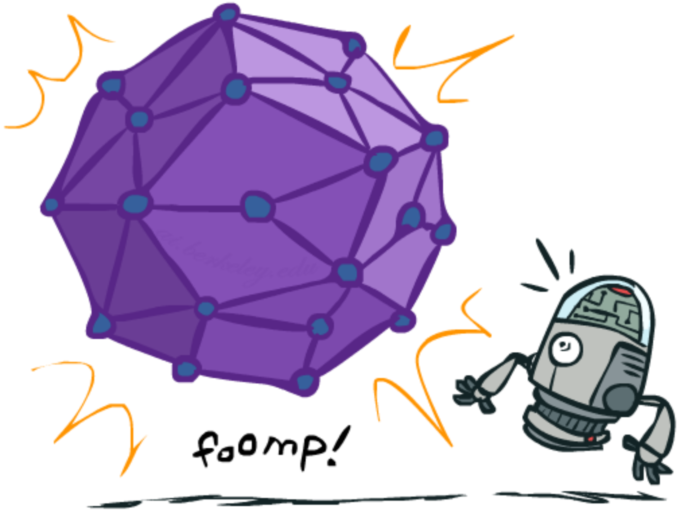
$$P(X_1, X_2, \ldots X_n)$$

BNs: Huge space savings!

Also easier to elicit local CPTs

Also faster to answer queries (coming)





Extreme case: total independence

N fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

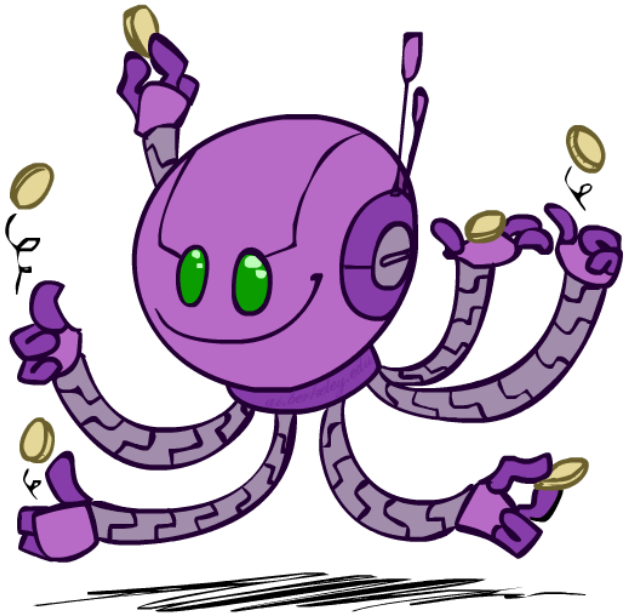
$P(X_2)$

H	0.5
T	0.5

$\cdots$

$P(X_n)$

H	0.5
T	0.5



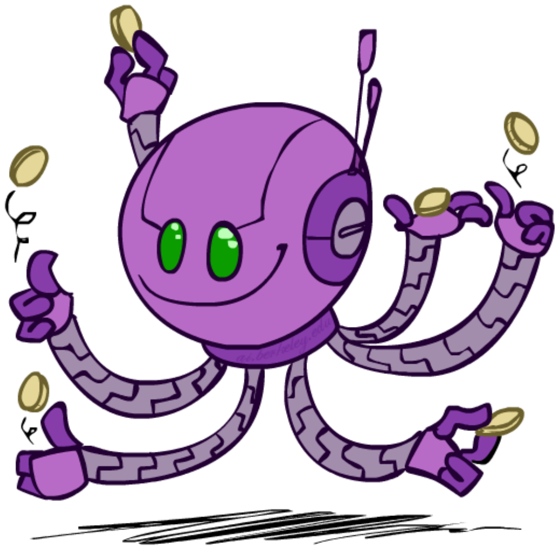
$P(X_1, X_2, \ldots X_n)$

2^n

Example: coin flips

N independent coin flips

X_1 X_2 $\cdots$ X_n



No interactions between variables: **absolute independence**

Example: coin flips

X_1 X_2 ... X_n

$P(X_1)$

h	0.5
t	0.5

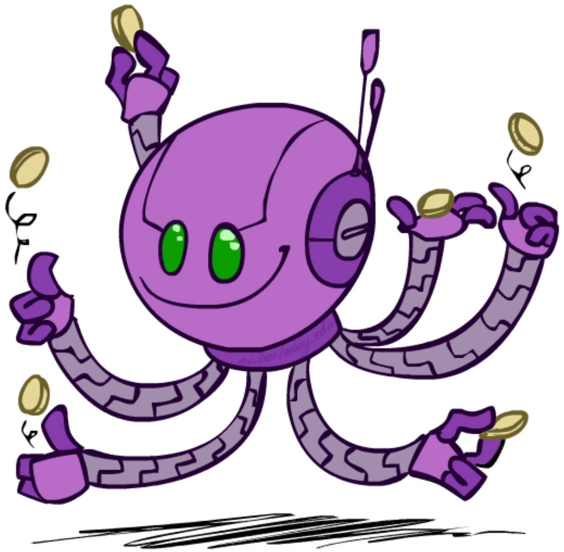
$P(X_2)$

h	0.5
t	0.5

...

$P(X_n)$

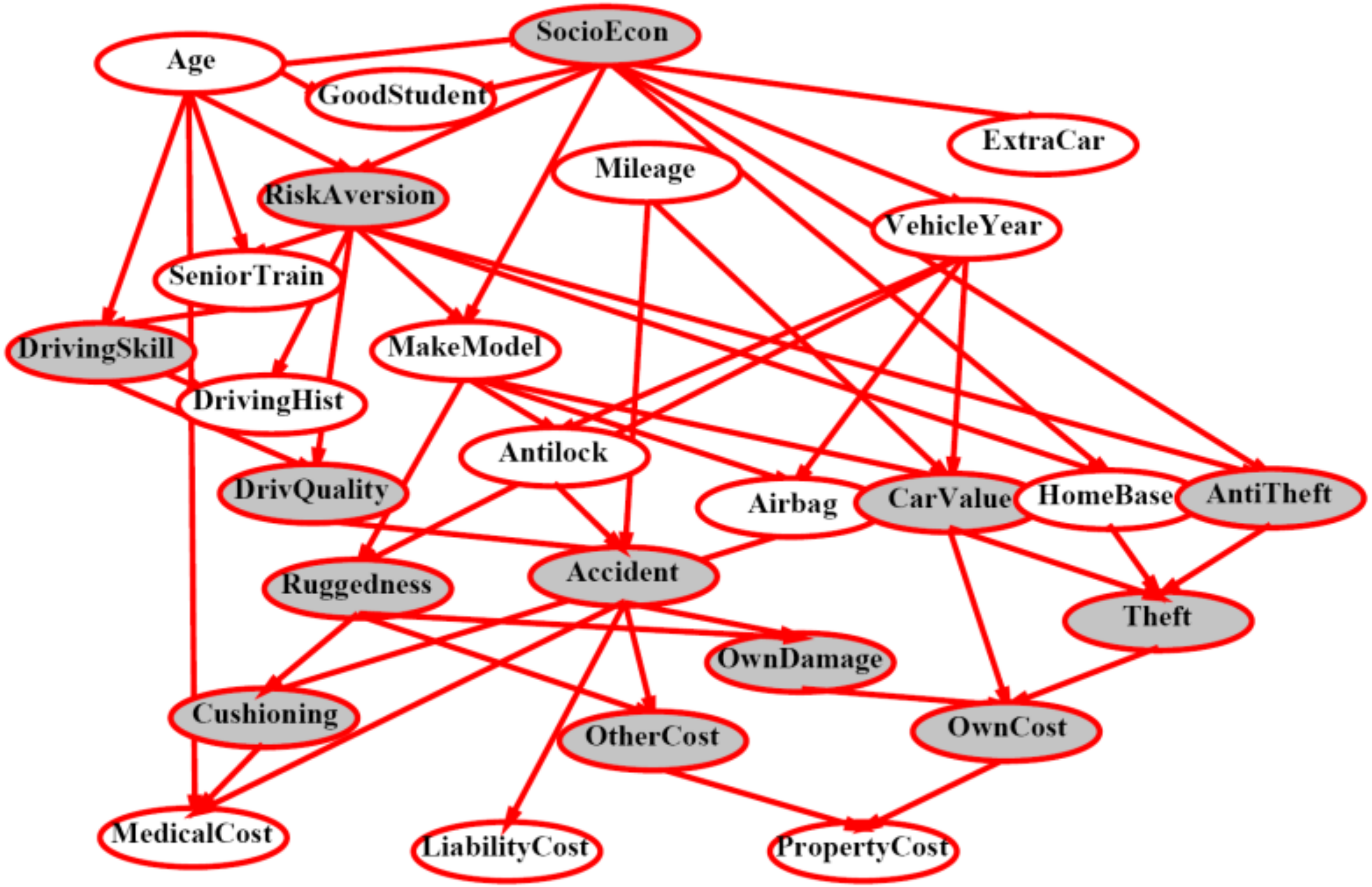
h	0.5
t	0.5



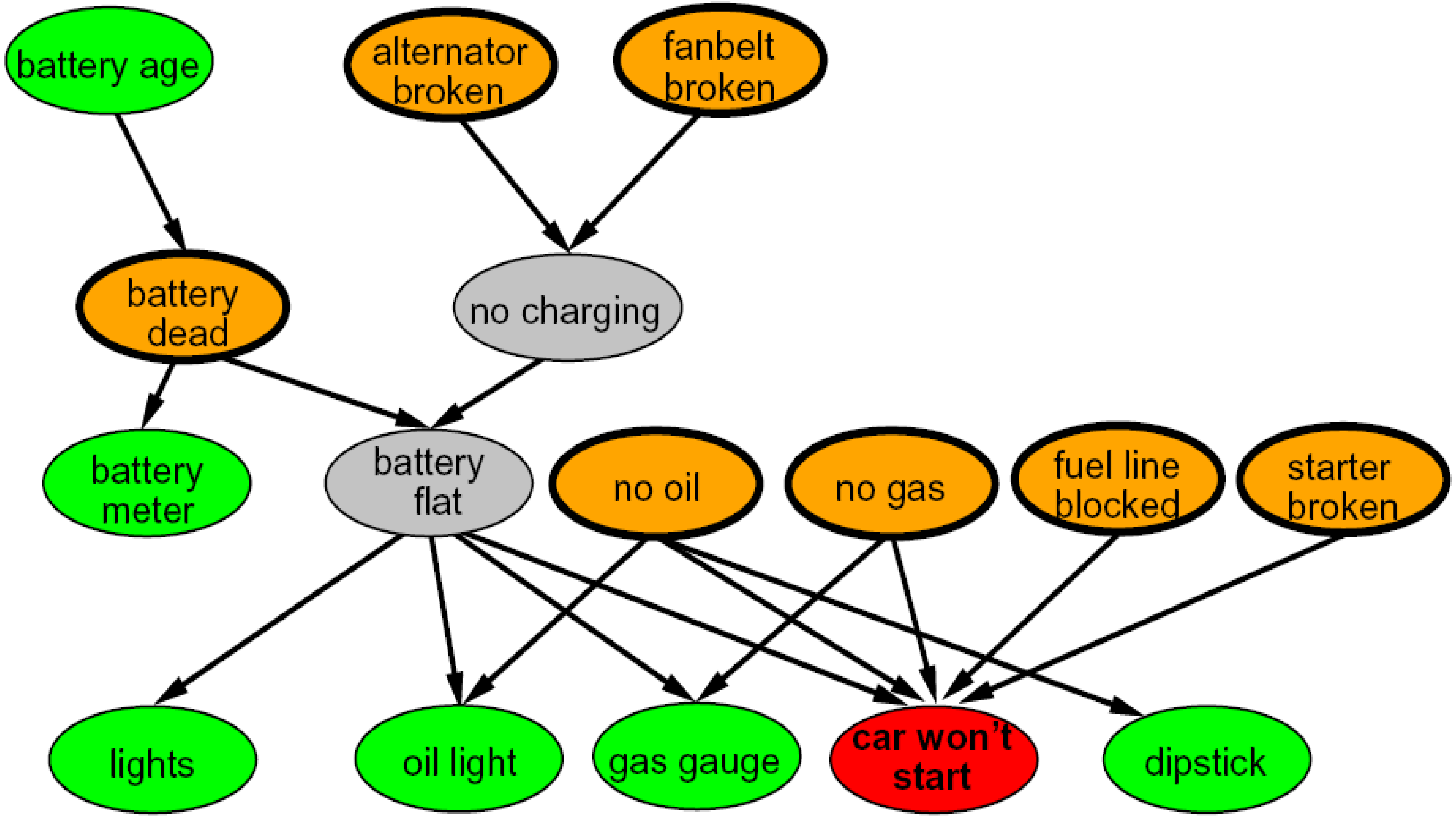
$$P(h, h, t, h) =$$

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

Example Bayes' Net: insurance



Example Bayes' Net: troubleshooting car

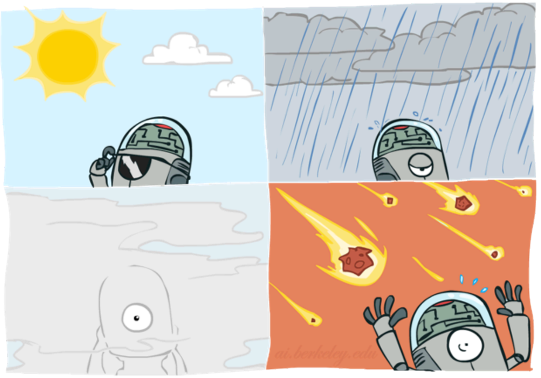


Bayes' Net semantics

Graphical model notation

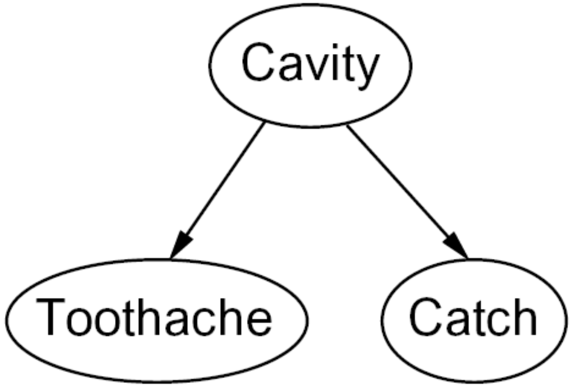
Nodes: variables (with domains)

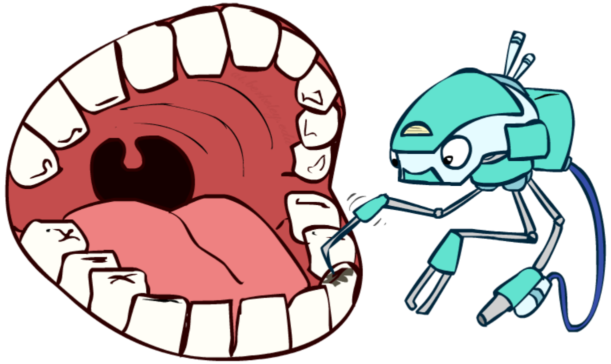
- Can be assigned (observed) or unassigned (unobserved)



Arcs: direct interactions

- Similar to CSP constraints
- Indicate “direct influence” between variables
- Formally: encode conditional independence (more later)



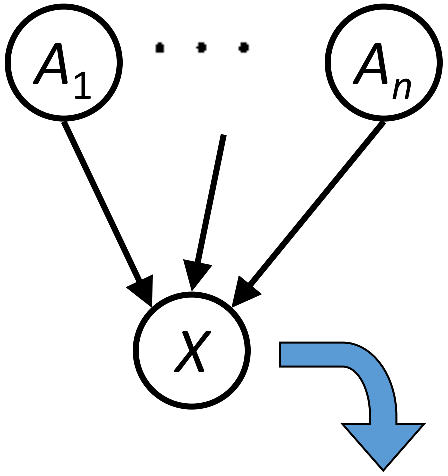


For now: imagine that arrows mean direct causation (in general, they don’t! but in practice, they often do!)

Bayes' Net semantics

A set of nodes, one per variable X

A **directed, acyclic graph (DAG)**



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

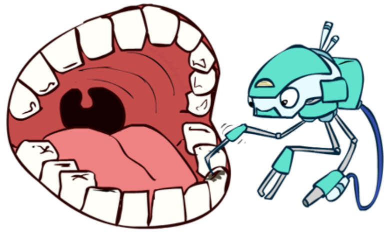
Probabilities in BNs

Bayes' nets implicitly encode joint distributions

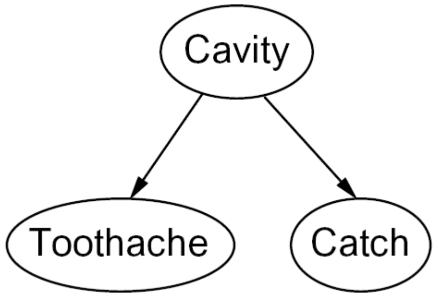
- As a product of local conditional distributions
- To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

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

- Example:



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



Probabilities in BNs

Why are we guaranteed that setting

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

results in a proper joint distribution?

Chain rule (valid for all distributions): $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$

<u>Assume</u> conditional independences: $P(x_i | x_1, \ldots x_{i-1}) = P(x_i | parents(X_i))$

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

Bayes Nets: assumptions

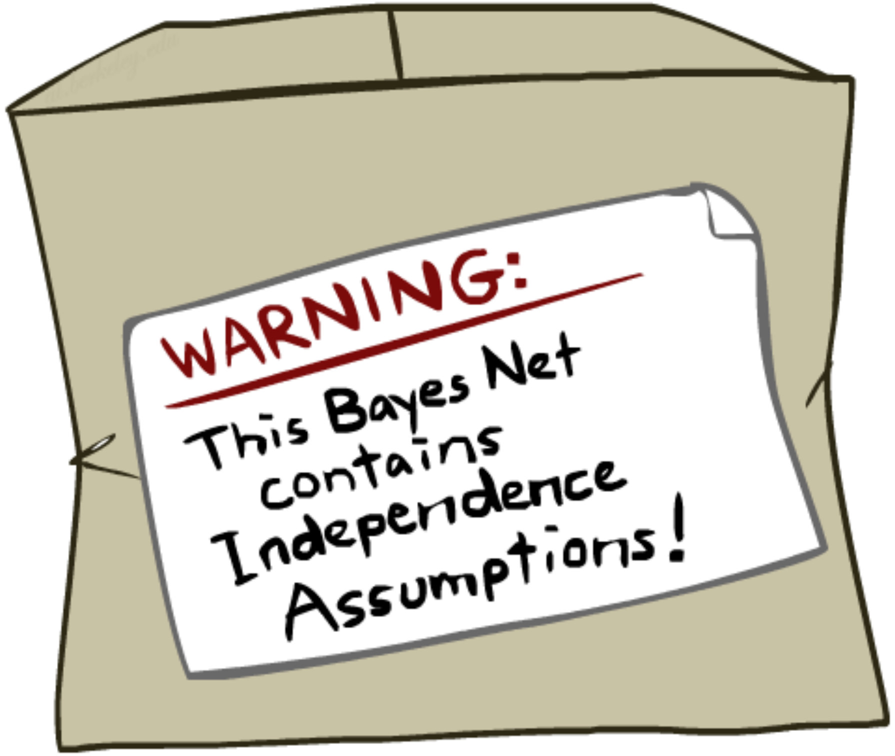
Assumptions we are required to make to define the Bayes net when given the graph:

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

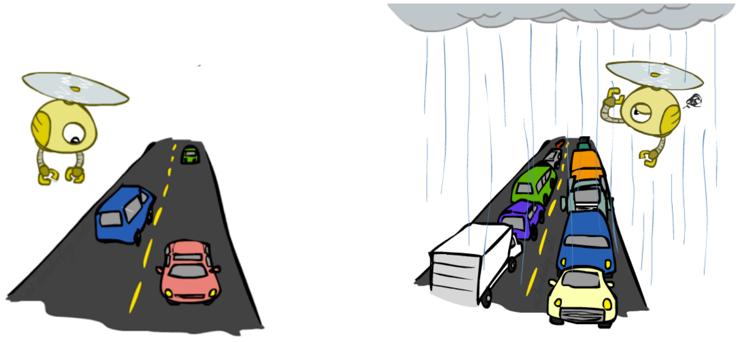
Beyond above "chain rule → Bayes net" conditional independence assumptions

- Often additional conditional independences
- They can be read off the graph

Important for modeling: understand assumptions made when choosing a Bayes net graph



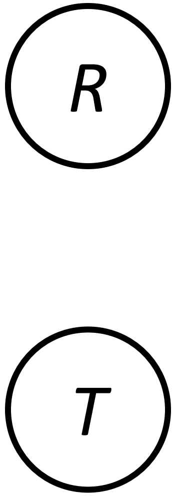
Example: traffic



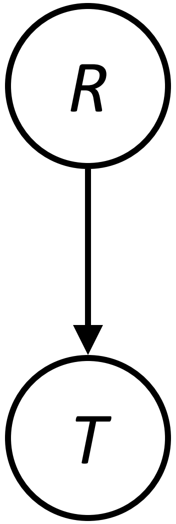
Variables:

- R: It rains
- T: There is traffic

Model 1: independence

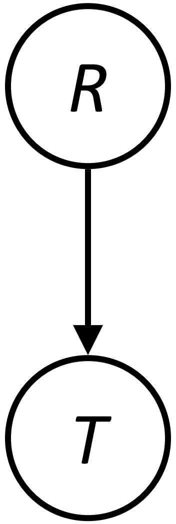


Model 2: rain causes traffic



Why is an agent using model 2 better?

Example: Traffic



$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

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

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

$P(T, R)$

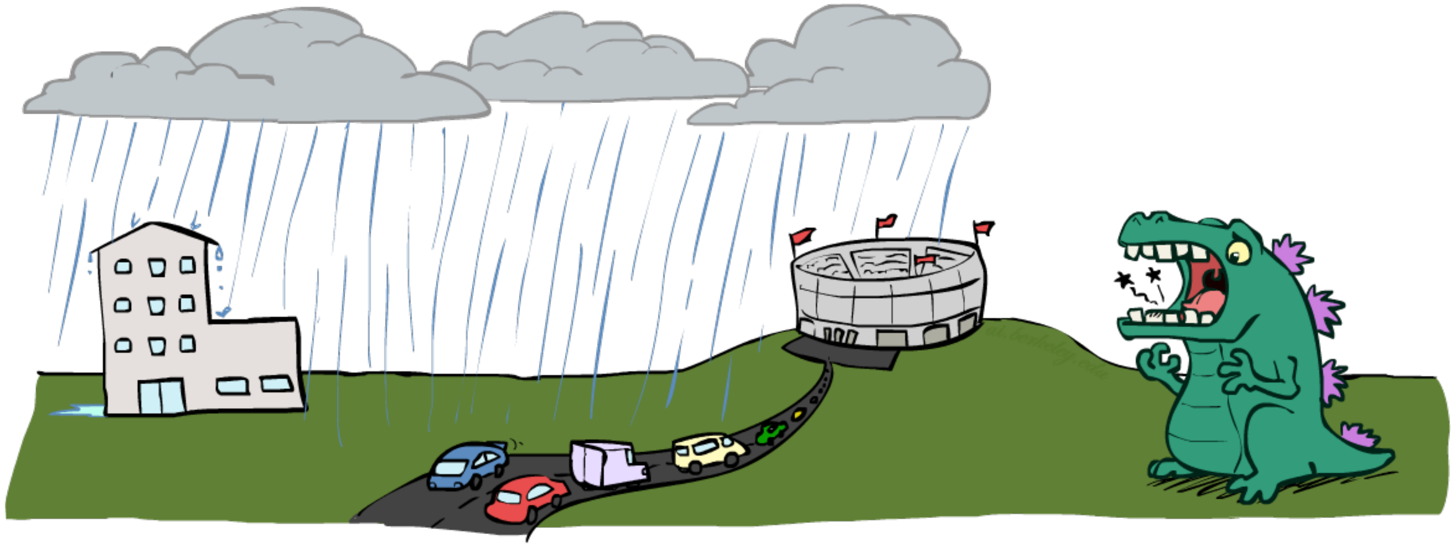
+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

Example: traffic

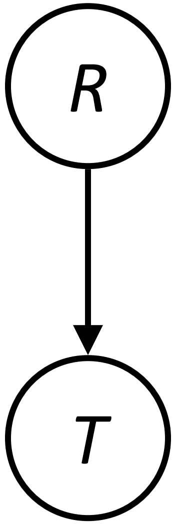
Let's build a causal graphical model!

Variables

- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
- C: Cavity



On causality: returning to our traffic example



$P(R)$

+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

Probabilities in BNs

Why are we guaranteed that setting

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

results in a proper joint distribution?

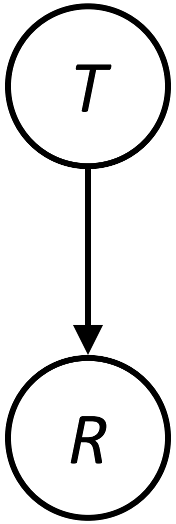
Chain rule (valid for all distributions): $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$

<u>Assume</u> conditional independences: $P(x_i | x_1, \ldots x_{i-1}) = P(x_i | \mathit{parents}(X_i))$

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

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

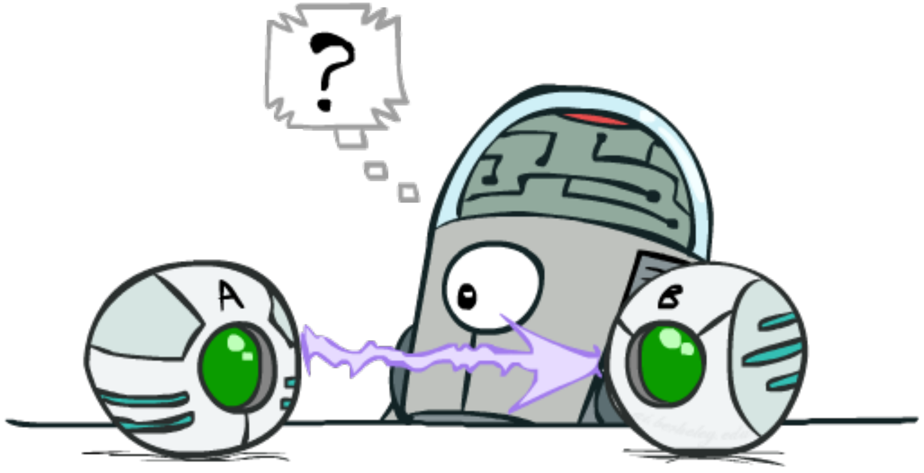
$P(T, R)$

+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

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

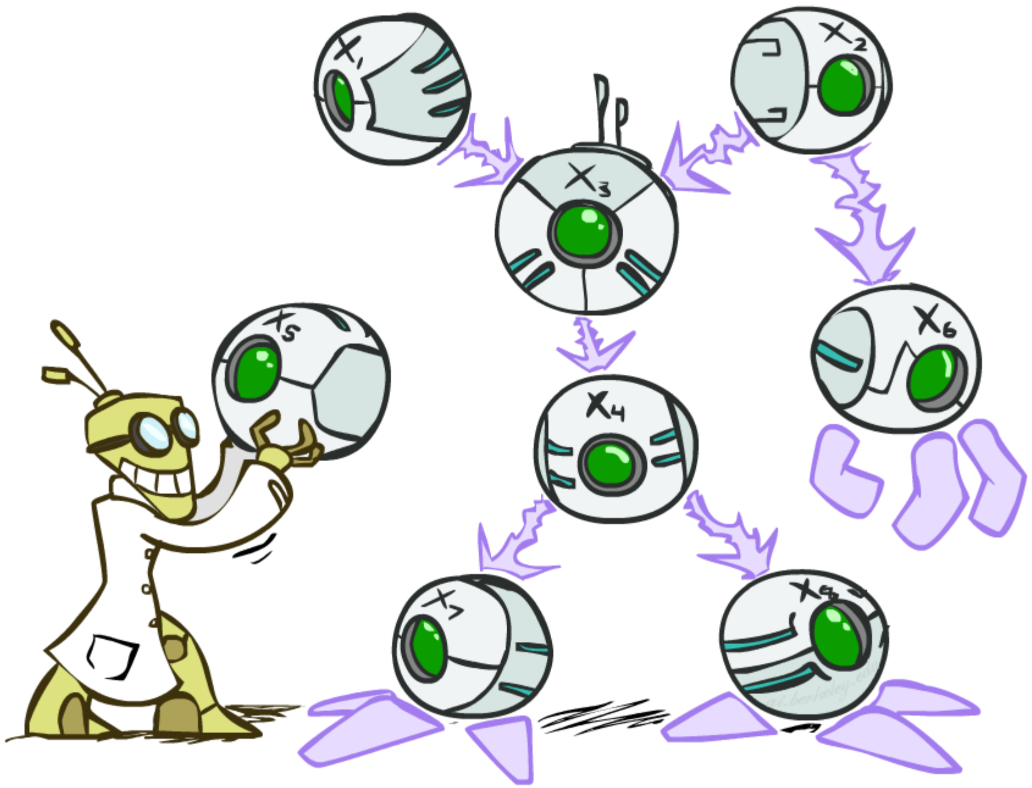
Bayes' Nets

So far: how a Bayes' net encodes a joint distribution

Next: how to answer queries about that distribution

- *Main goal: answer queries about conditional independence and influence*

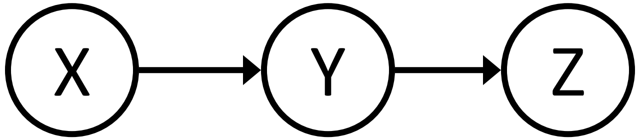
After that: how to answer numerical queries (inference)



Independence in a BN

Important question about a BN:

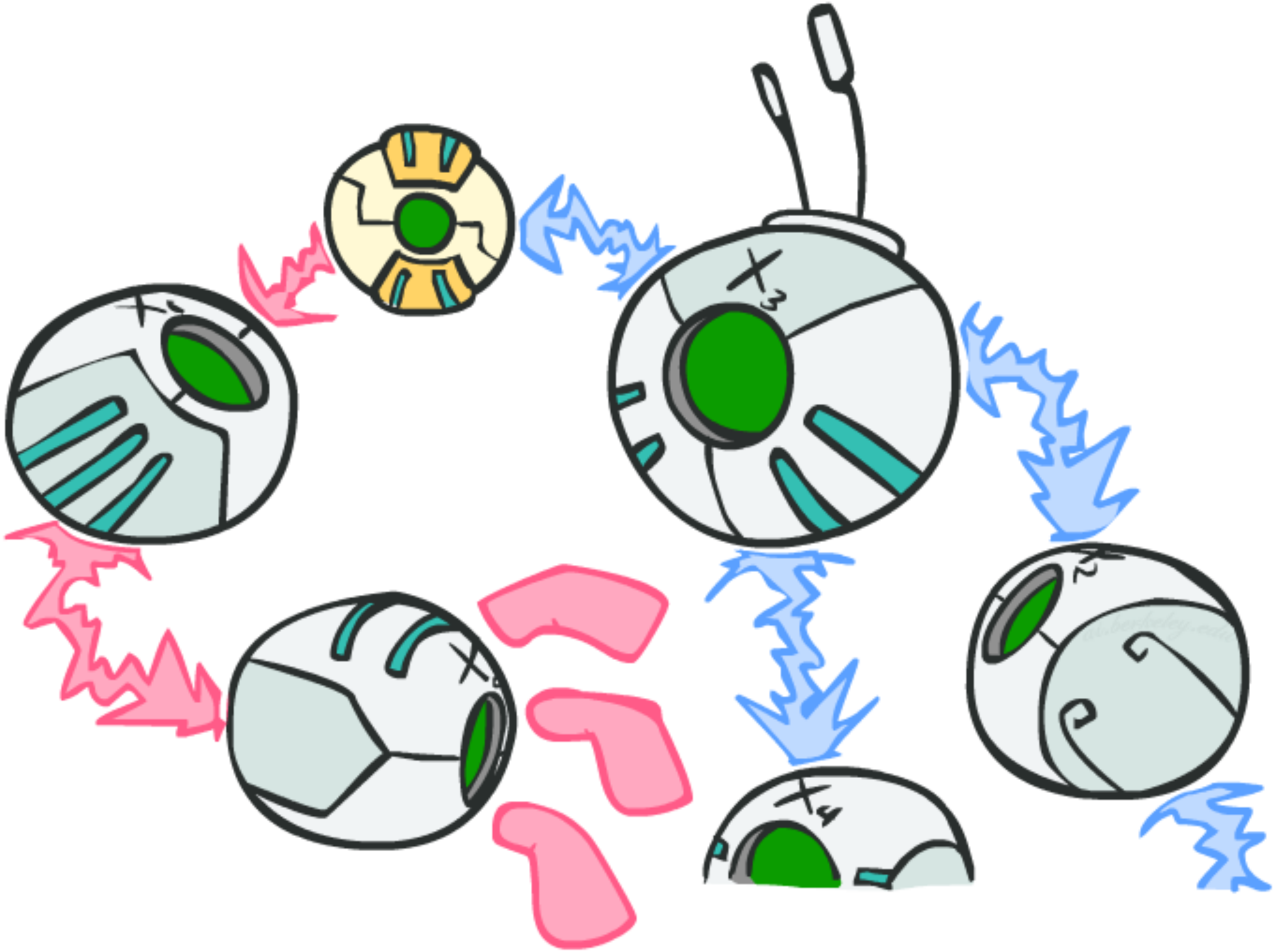
- Are two nodes independent given certain evidence?
- If yes, can prove using algebra (tedious in general)
- If no, can prove with a counter example
- Example:



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

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**?

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

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

$$= P(z|y)$$

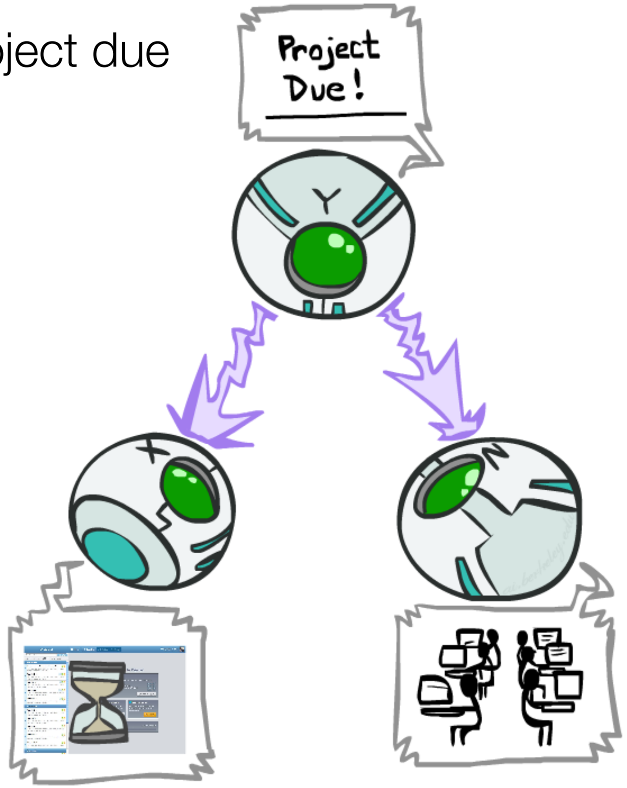
Yes!

Evidence along the chain "blocks" the influence

Common 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.

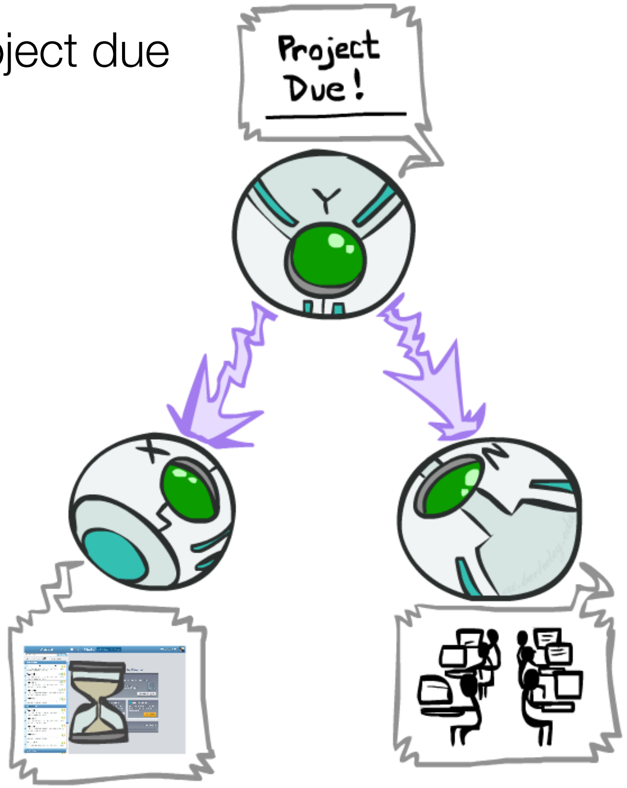
Example:

- Project due causes both forums busy and lab full

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?

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

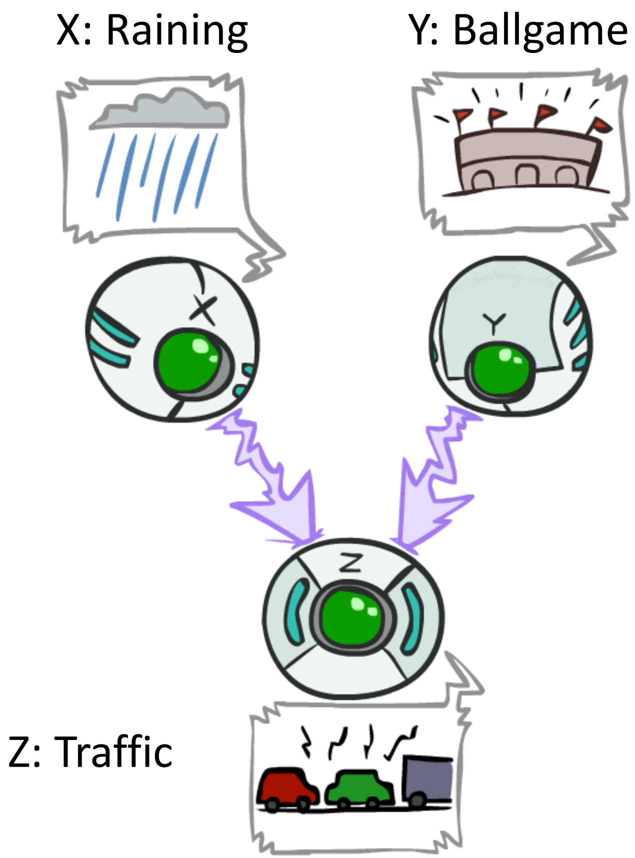
$$= \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}$$

$$= P(z|y)$$ ***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

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

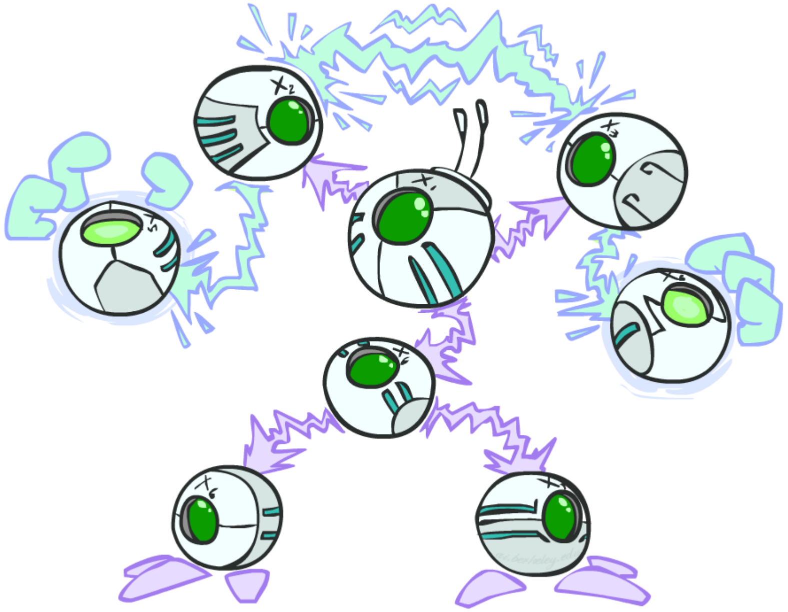
The general case

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



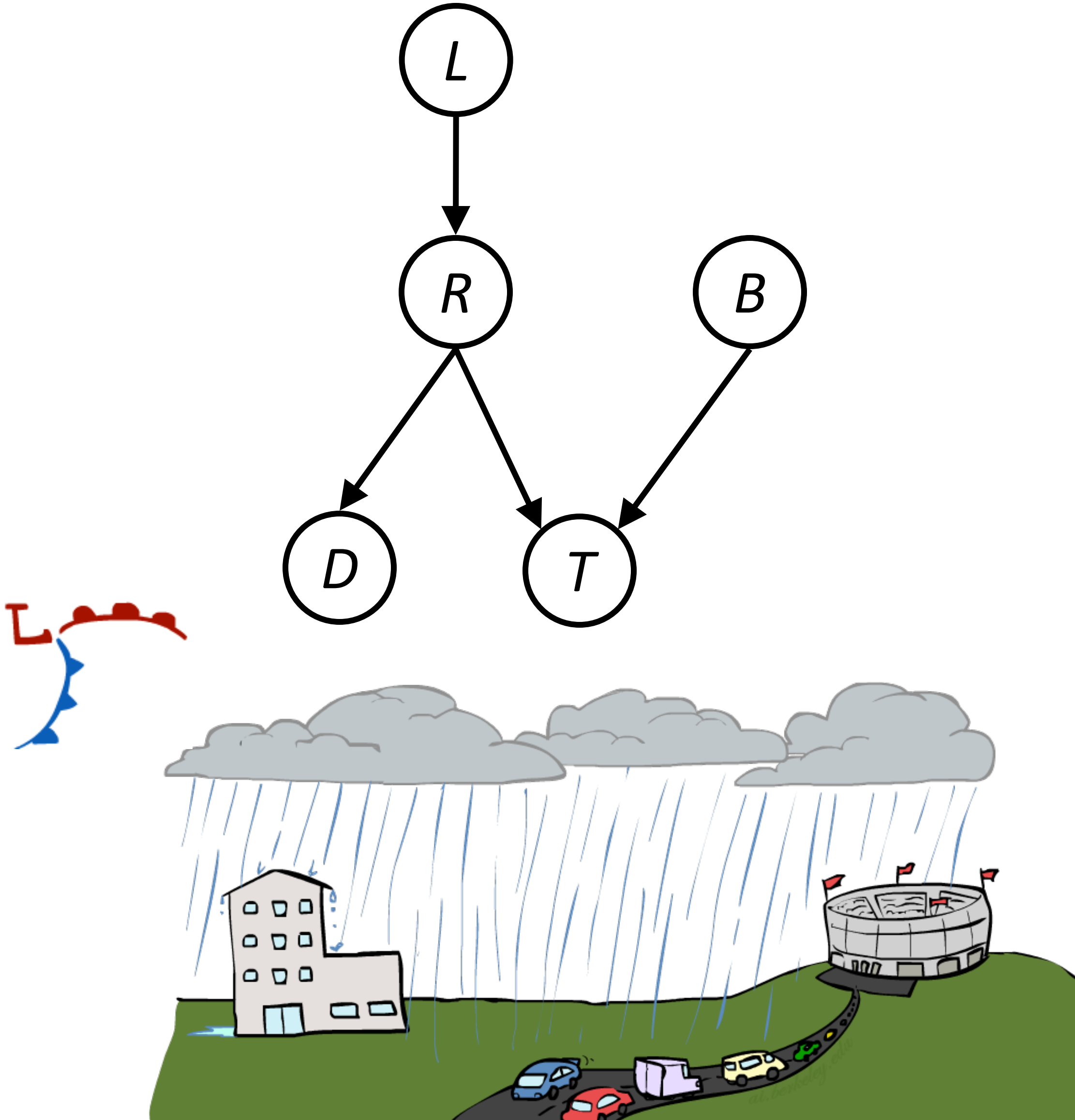
Reachability

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

Attempt 1: check if paths between two given nodes are blocked or not

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

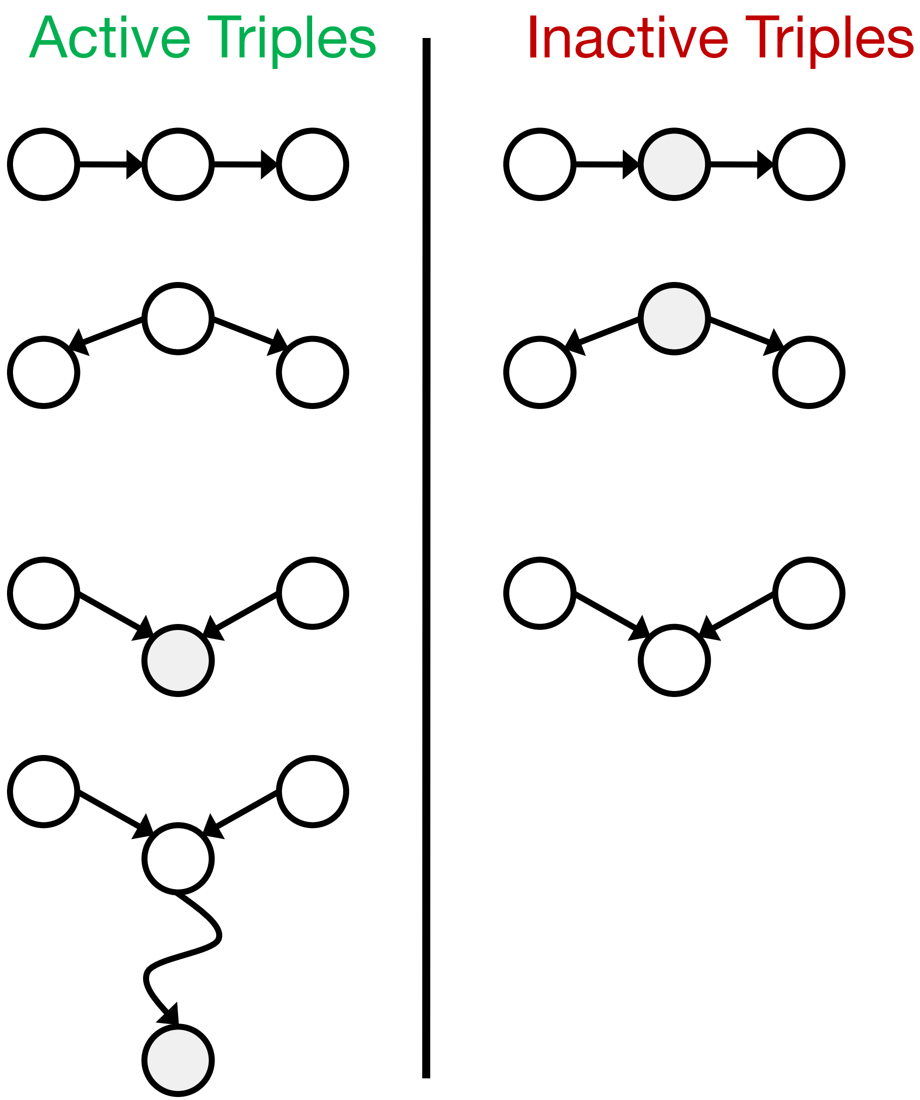
Question: Are X and Y conditionally independent given evidence variables {Z}?

- 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 descendents is observed

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



D-Separation

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

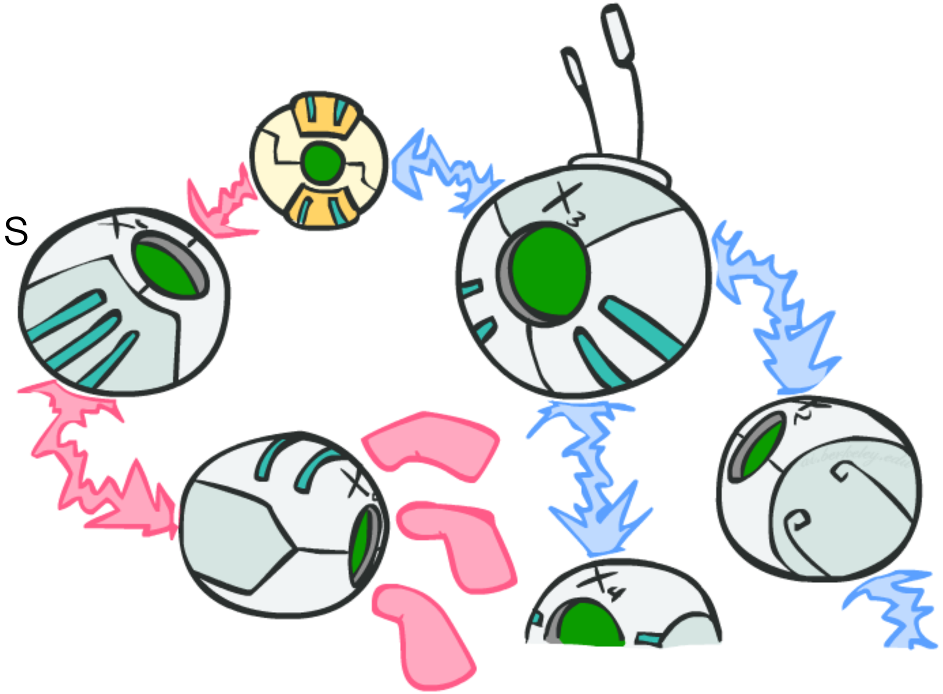
Check all (undirected!) paths between X_i and X_j

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

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

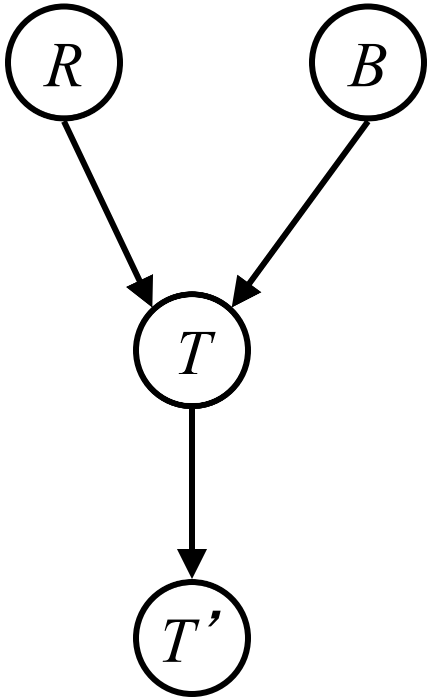


Example

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

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

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



Example

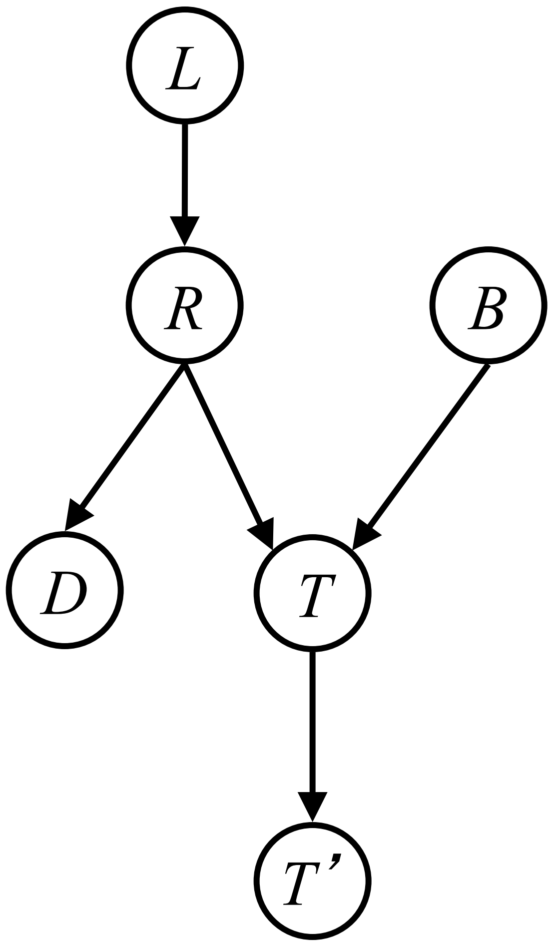
$L \perp\!\!\!\perp T' | T$ Yes

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

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

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

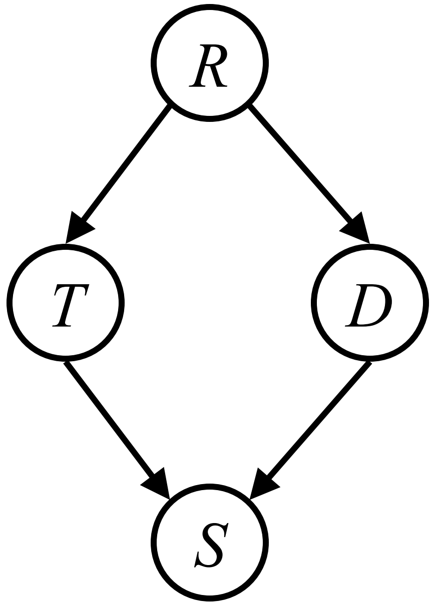
$L \perp\!\!\!\perp B | T, R$ Yes



Example

Variables:

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



Questions:

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

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

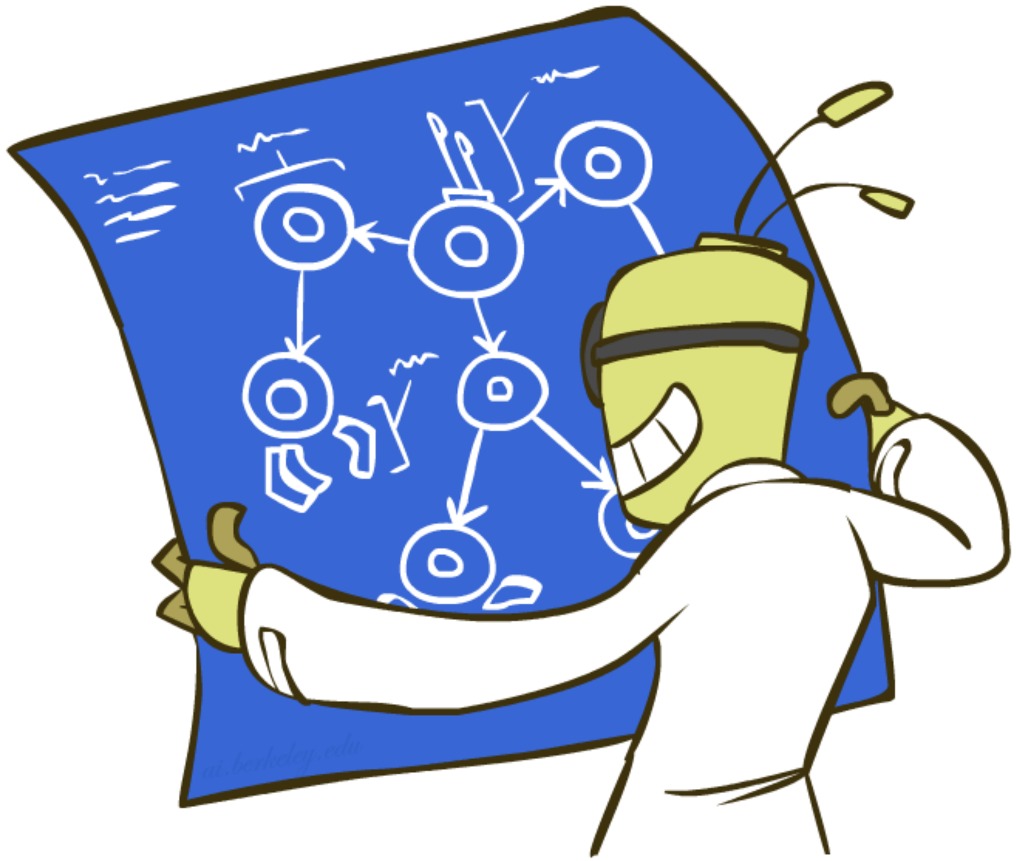
$T \perp\!\!\!\perp D | R, S$

Structure implications

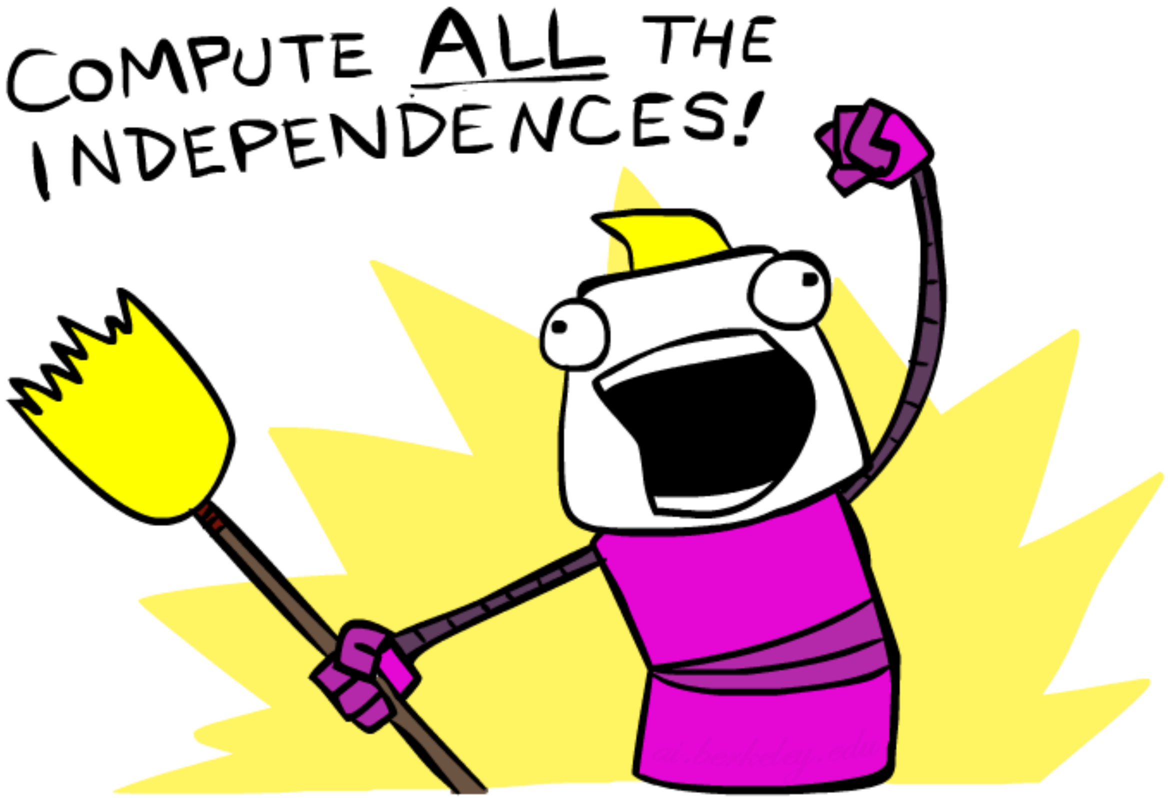
Given a Bayes net structure, can run *d*-separation algorithm to build a complete list of conditional independences that are necessarily true of the form

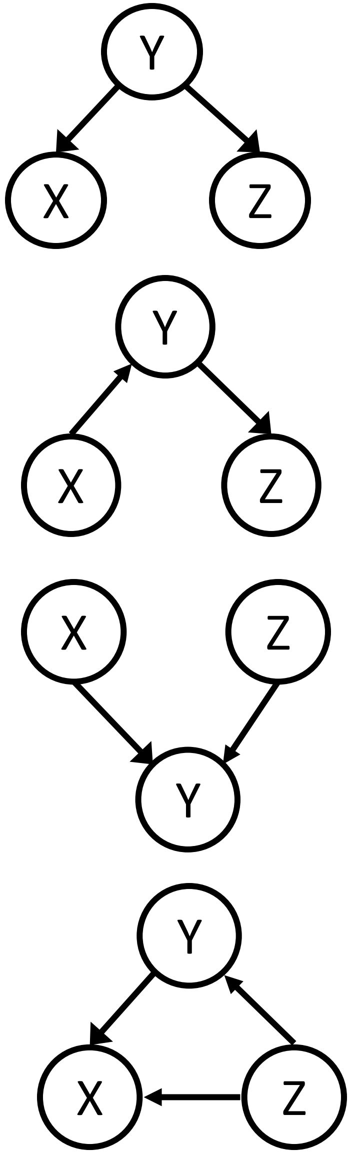
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, \ldots, X_{k_n}\}$$

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



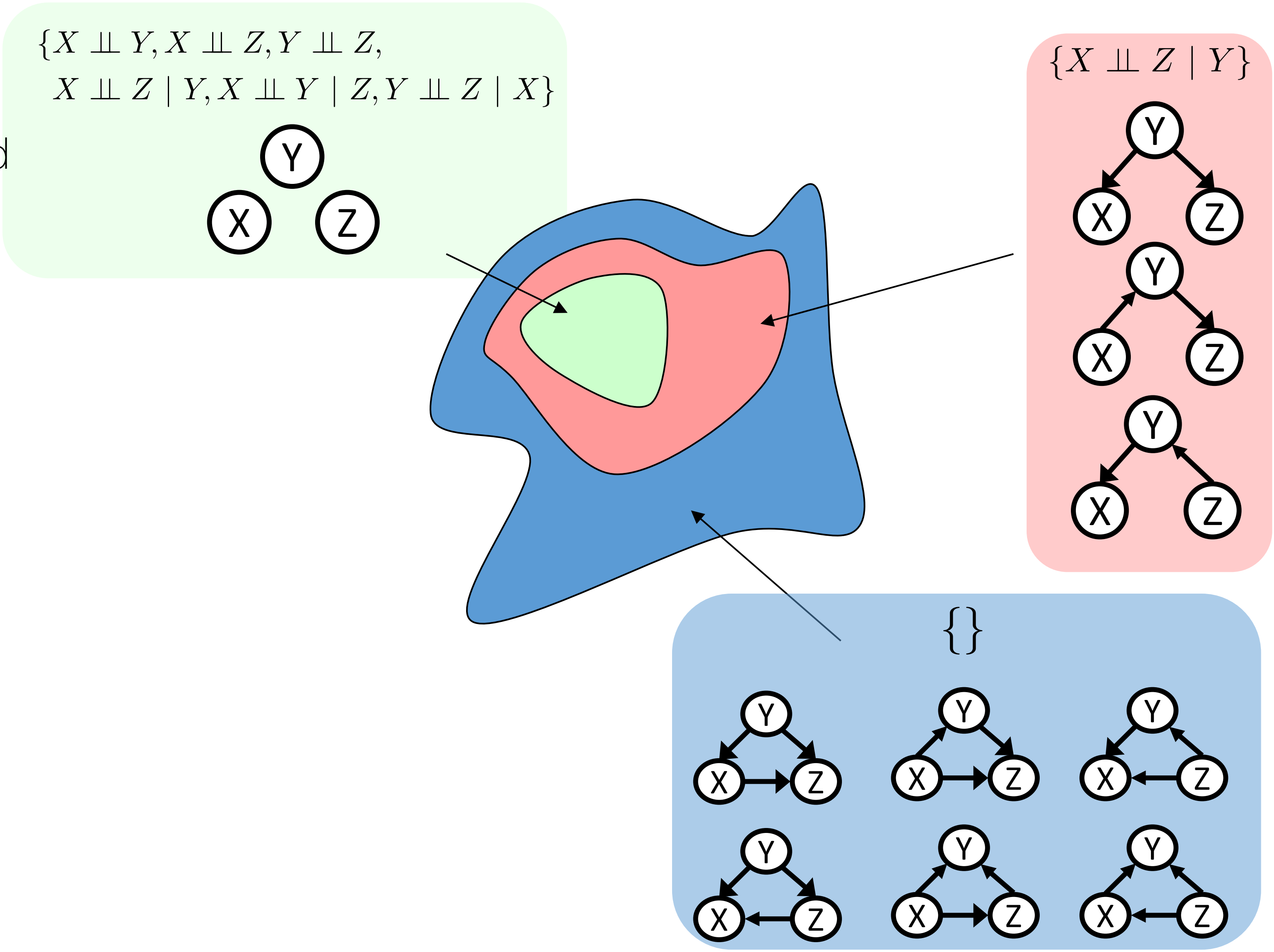
Computing all independences





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



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

That’s it for today!

- Next time: more on Bayes nets! (Estimation and inference!)
- ***START ON HOMEWORK 4***!!!