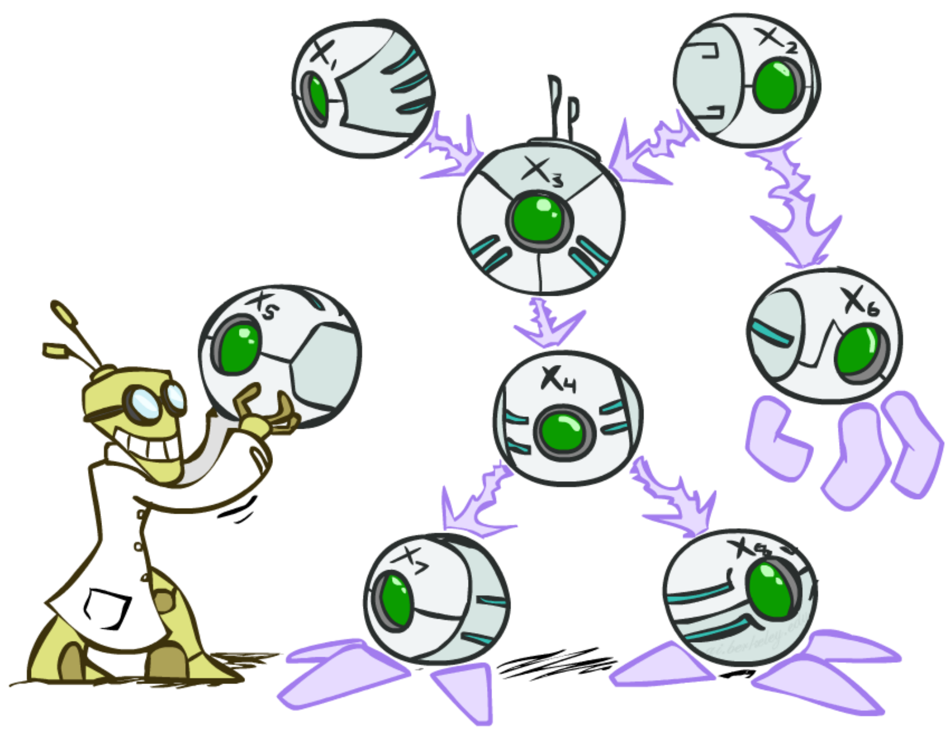
Ve492: Introduction to Artificial Intelligence
Bayesian Networks: Representation



Paul Weng

UM-SJTU Joint Institute

Slides adapted from http://ai.berkeley.edu, CMU, AIMA, UM

Bayes' Nets

- Representation
- Conditional Independences
- Probabilistic Inference

Probabilistic Models

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

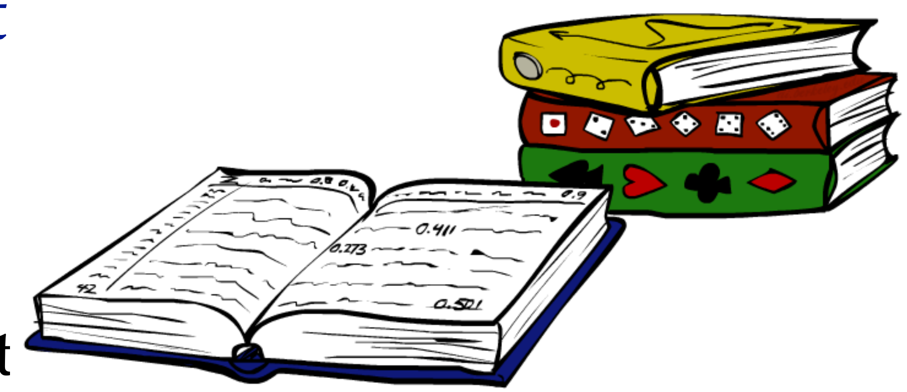
- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
  - "All models are wrong; but some are useful."
    – George E. P. Box

- 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)

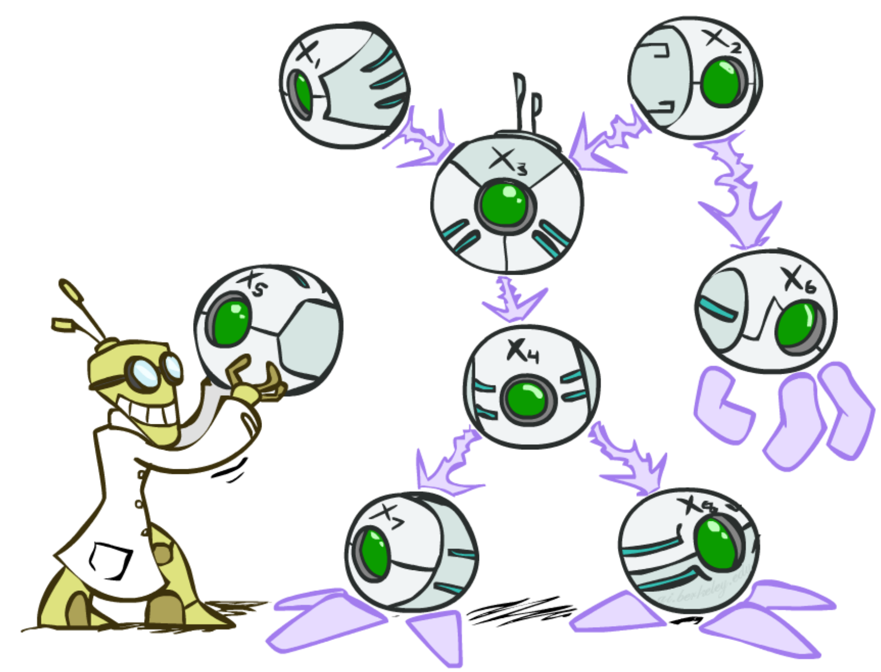
Bayes' Nets: Big Picture

- Joint distribution can be used for inference

- Three problems with directly 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
  - Computational complexity of inference

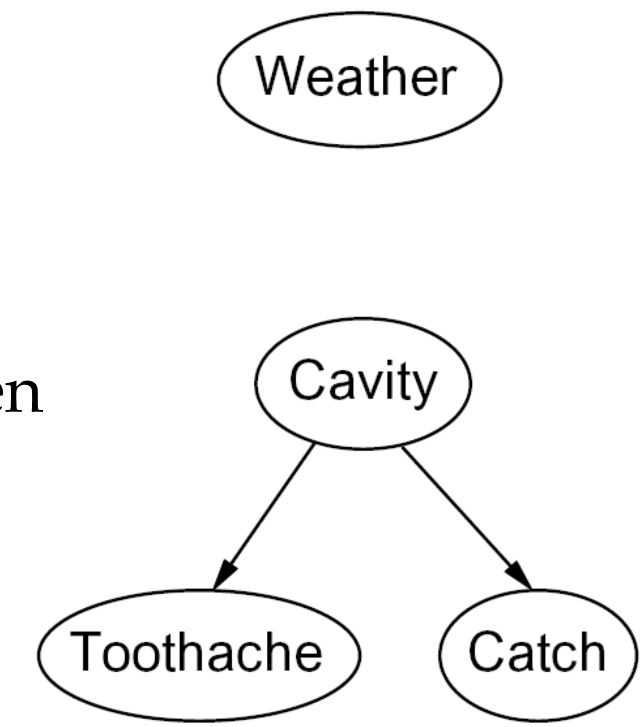


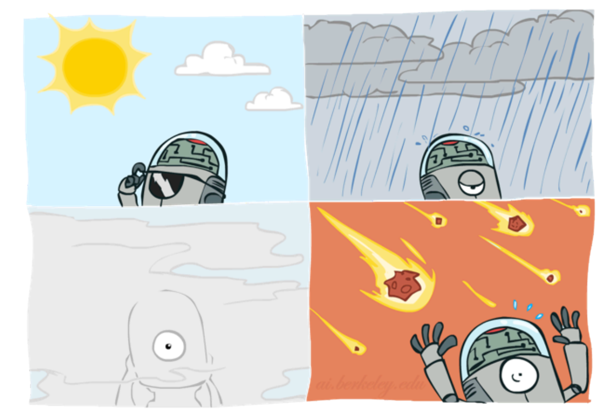
- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - Instance of graphical models
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions

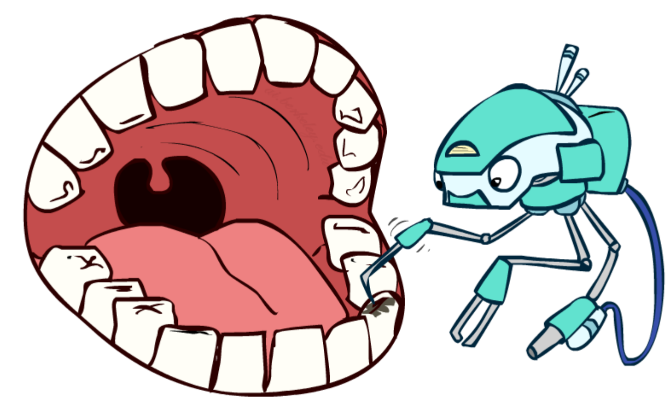


Bayes' Net Notation

- 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!)

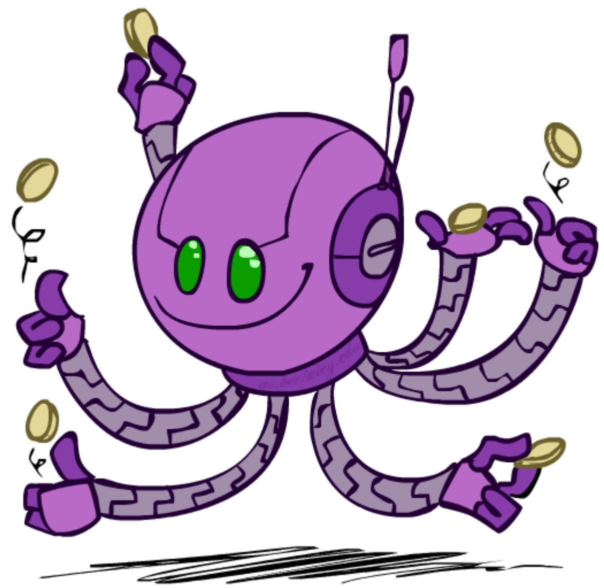






Example: Coin Flips

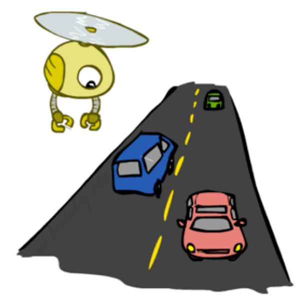
- N independent coin flips



- No interactions between variables: absolute independence

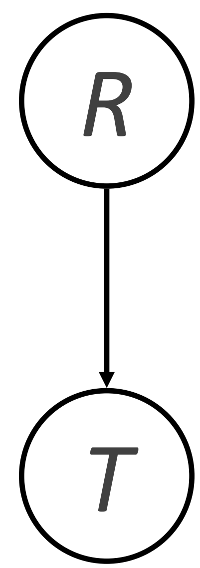
Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic



- Model 1: independence

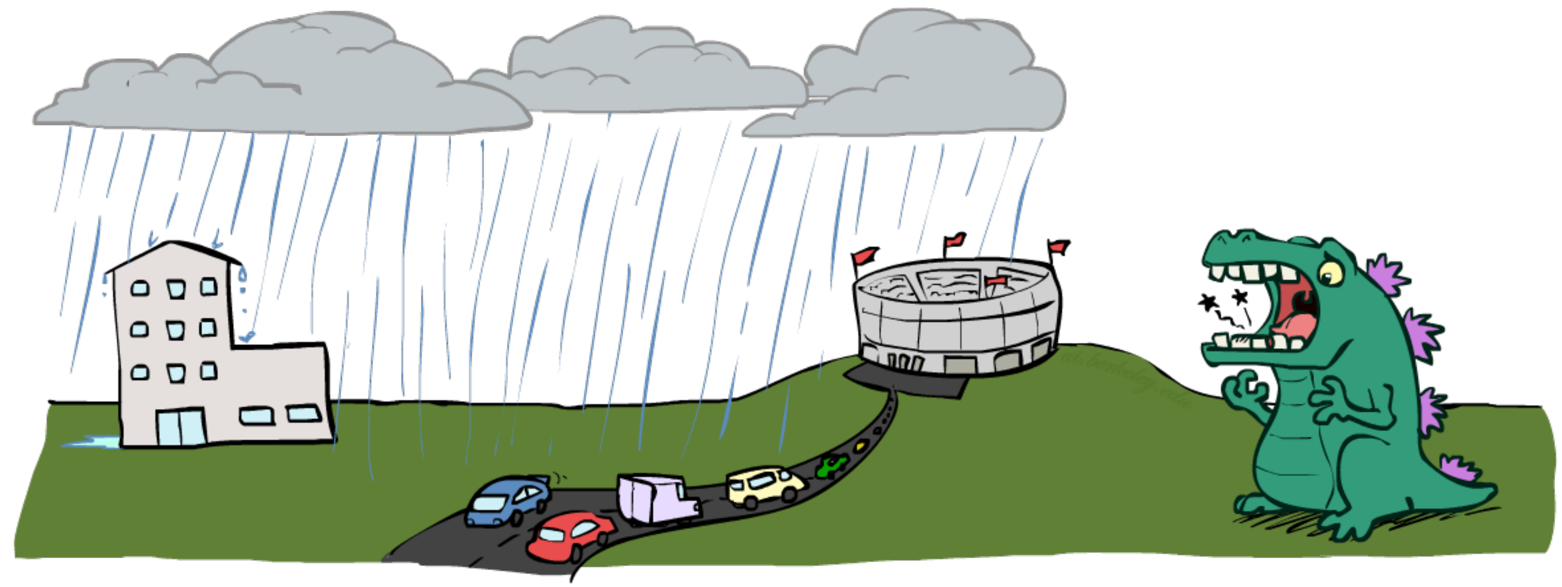
- Model 2: rain causes traffic

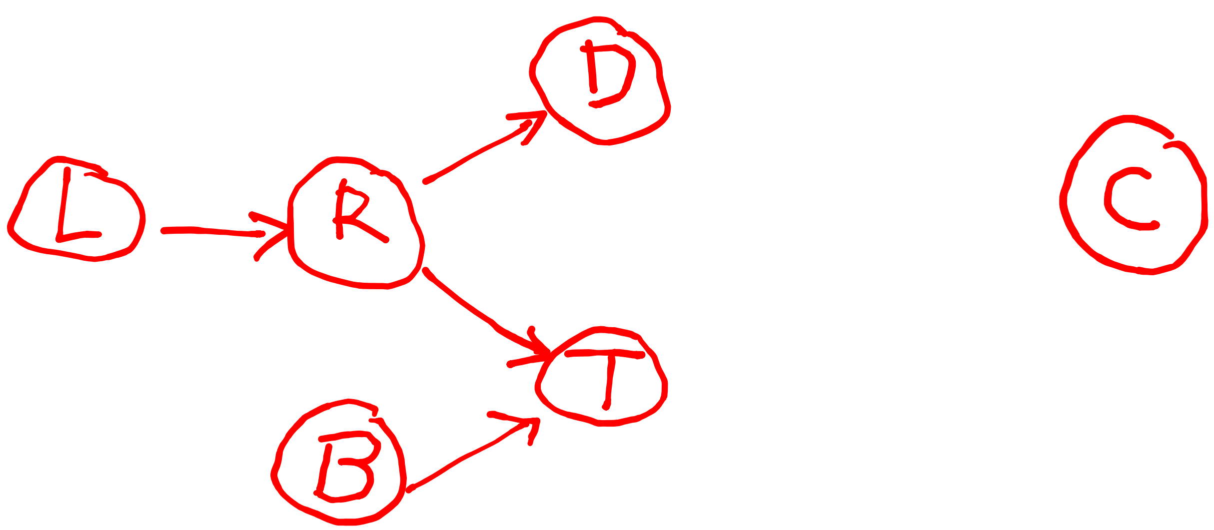


- Why is an agent using model 2 better?

Example: Traffic II

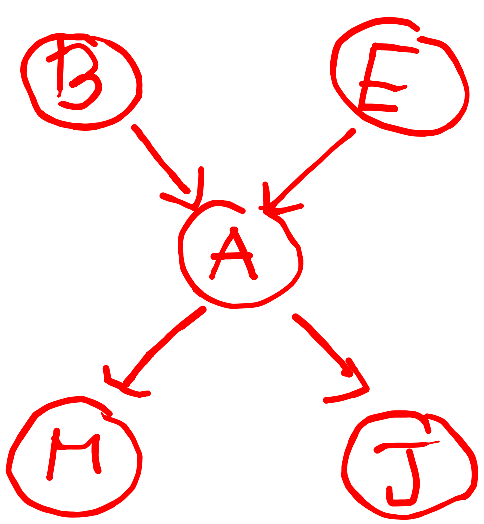
- Let's build a causal graphical model!
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - B: Ballgame
  - C: Cavity

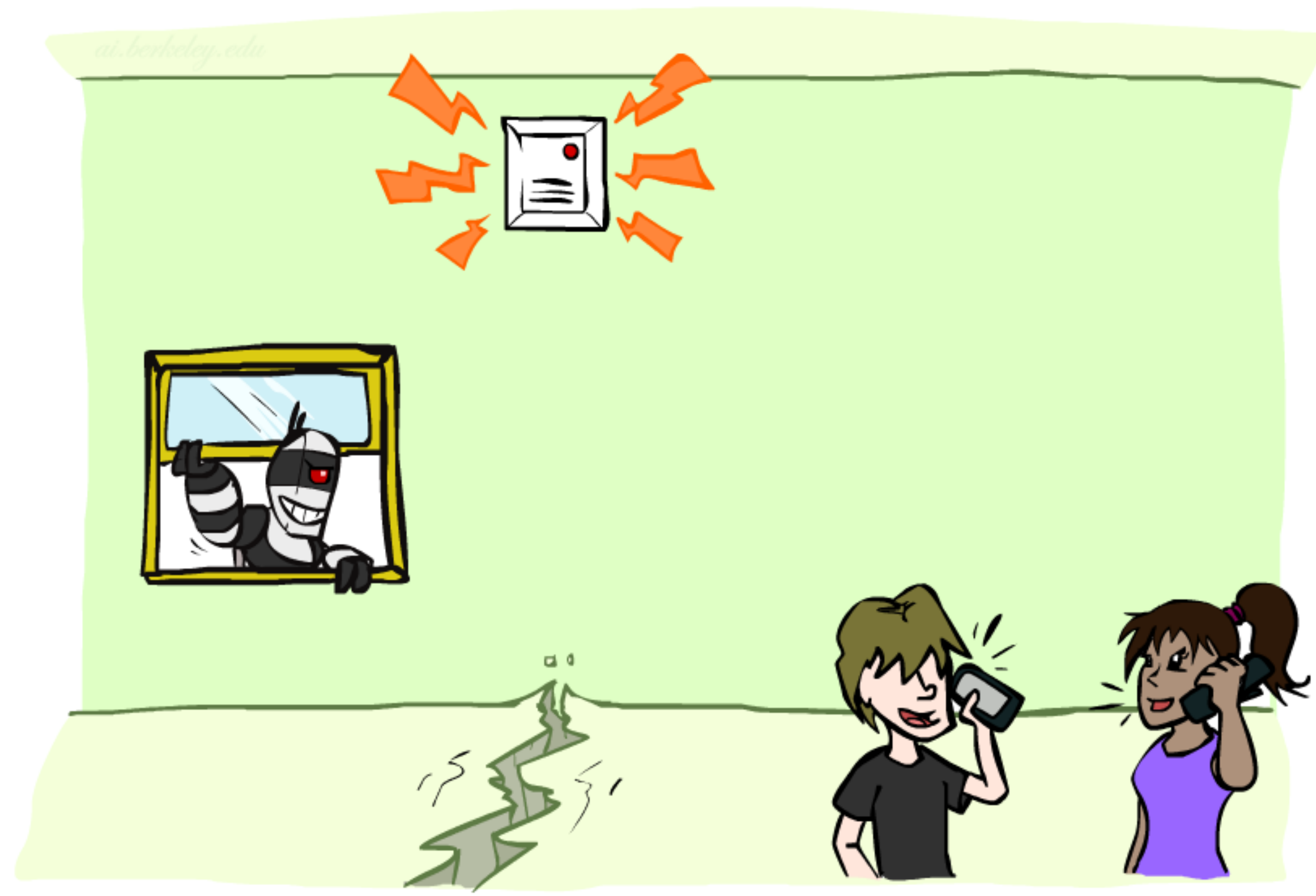




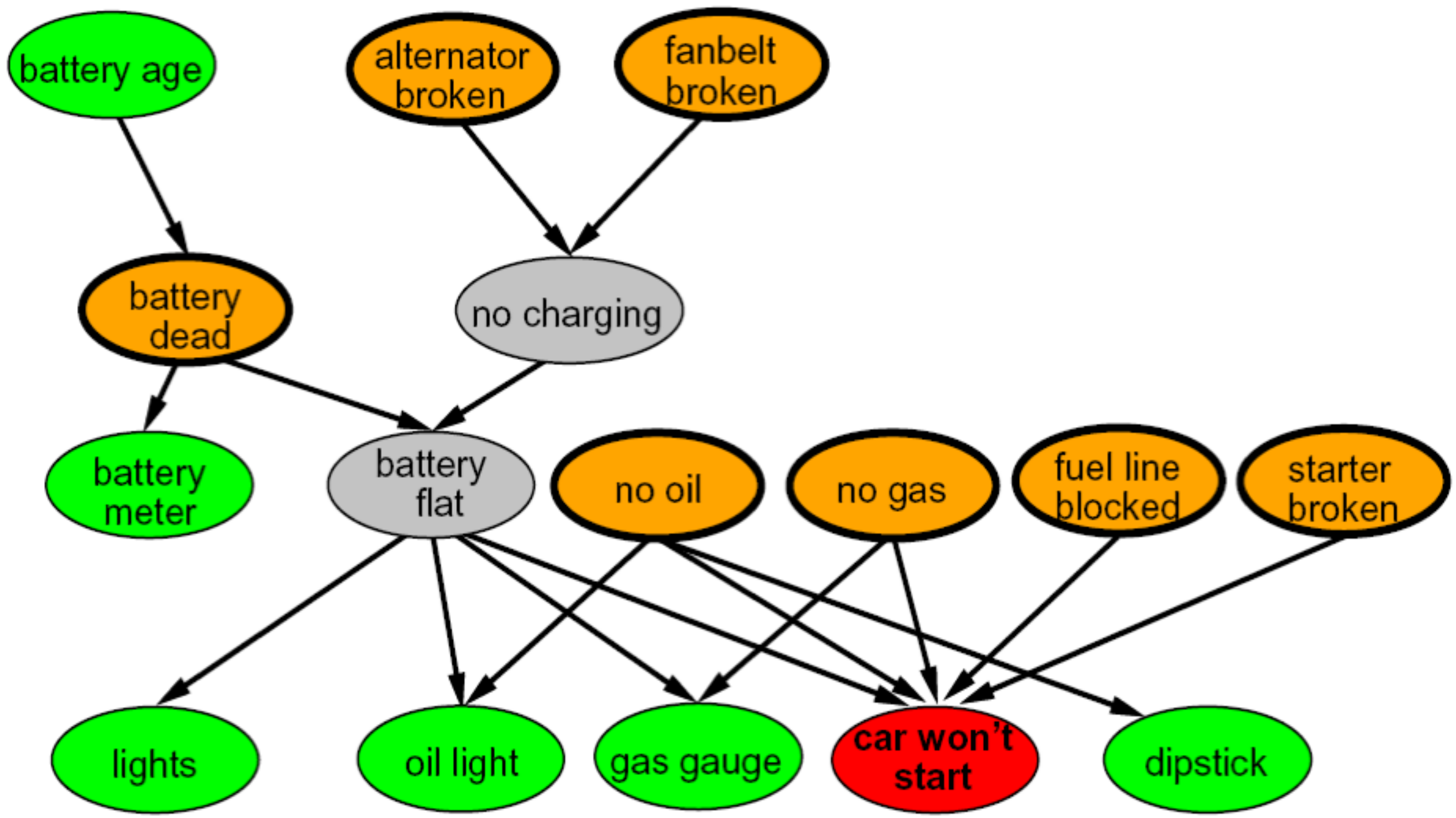
Example: Alarm Network

- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!

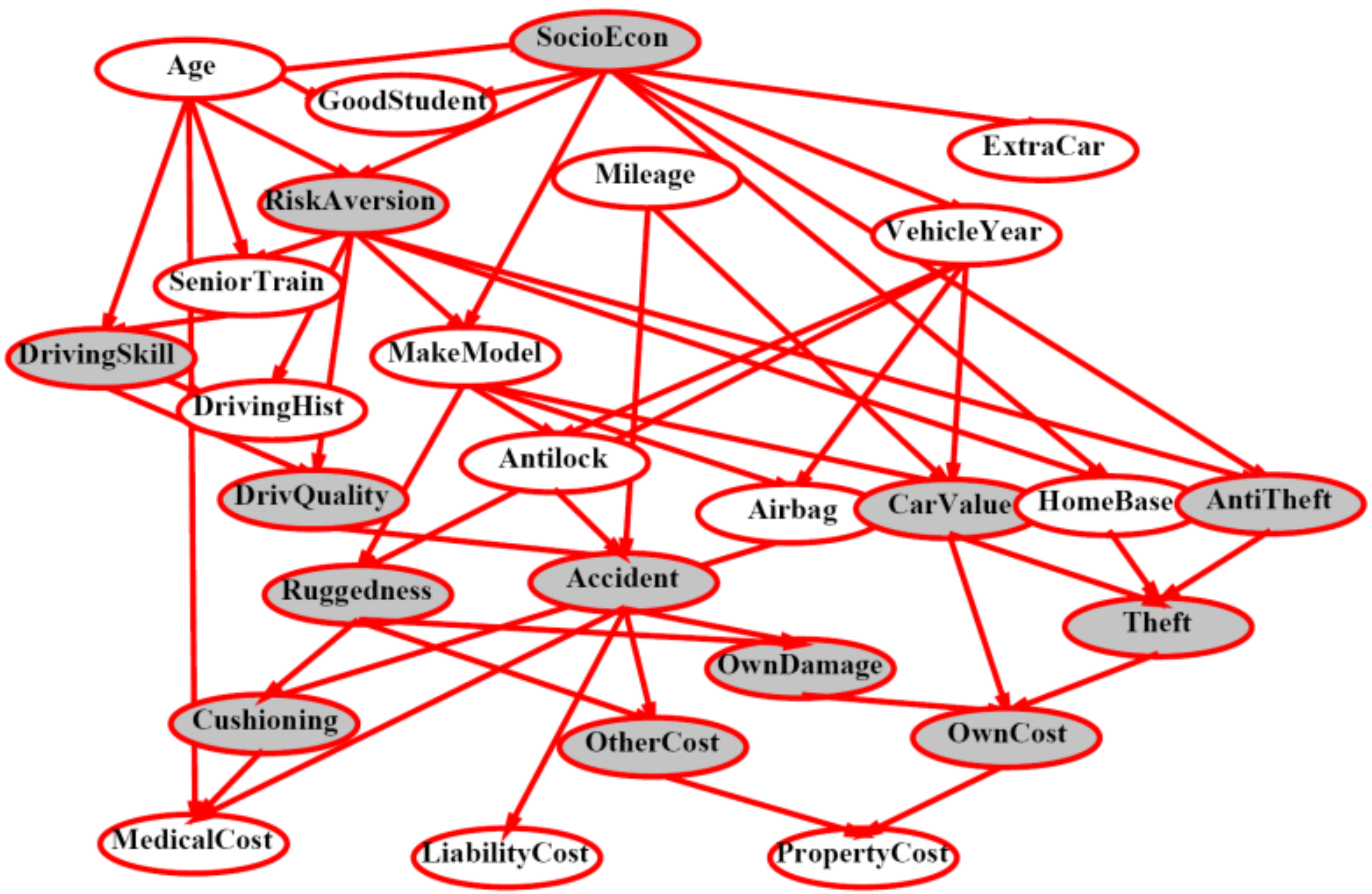




Example Bayes' Net: Car

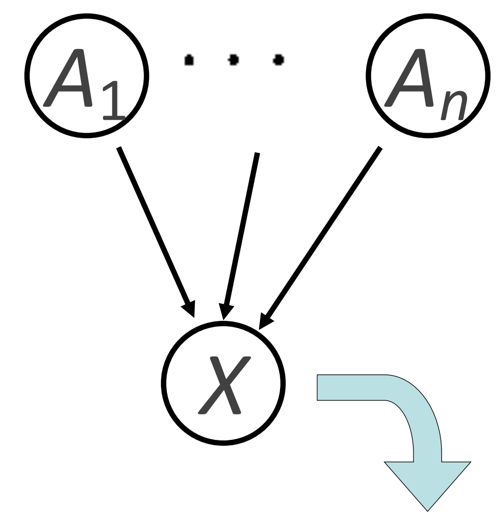


Example Bayes' Net: Insurance



Bayes' Net Definition and Semantics

- A set of nodes, one per variable
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents' values



$$P(X|a_1 \ldots a_n) \qquad P(X|A_1 \ldots A_n)$$

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

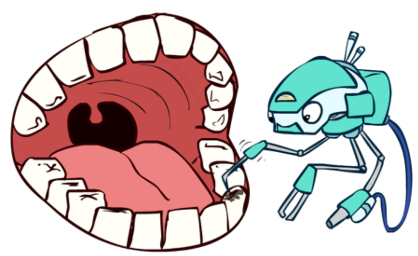
A Bayes net = Topology (graph) + Local Conditional Probabilities

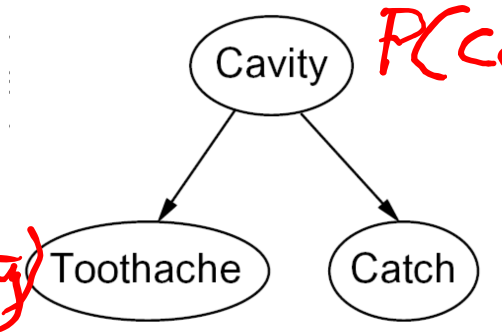
Probabilities in BNs

- 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:

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

  - Example:





P(cavity)

P(t | cavity)

P(Catch | cavity)

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

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

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

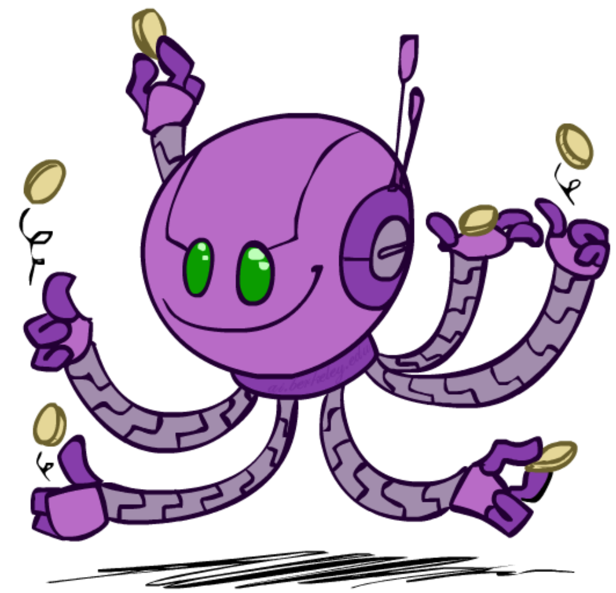
- <u>Assume</u> conditional independences: $P(x_i|x_1, \ldots x_{i-1}) = P(x_i|\mathit{parents}(X_i))$ ✓

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

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

Example: Coin Flips

X_1 $\quad$ X_2 $\quad \cdots \quad$ X_n



$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

$\cdots$

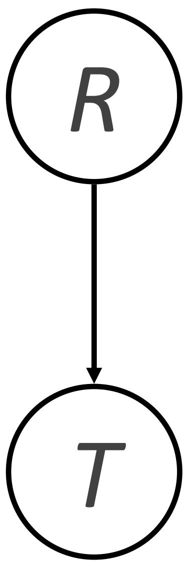
$P(X_n)$

h	0.5
t	0.5

$P(h, h, t, h) =$ $0.5 \times 0.5 \times 0.5 \times 0.5$

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

Example: Traffic



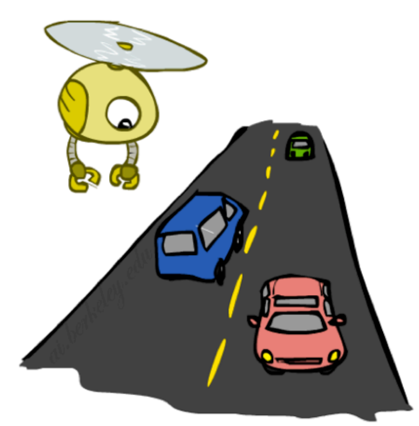
$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

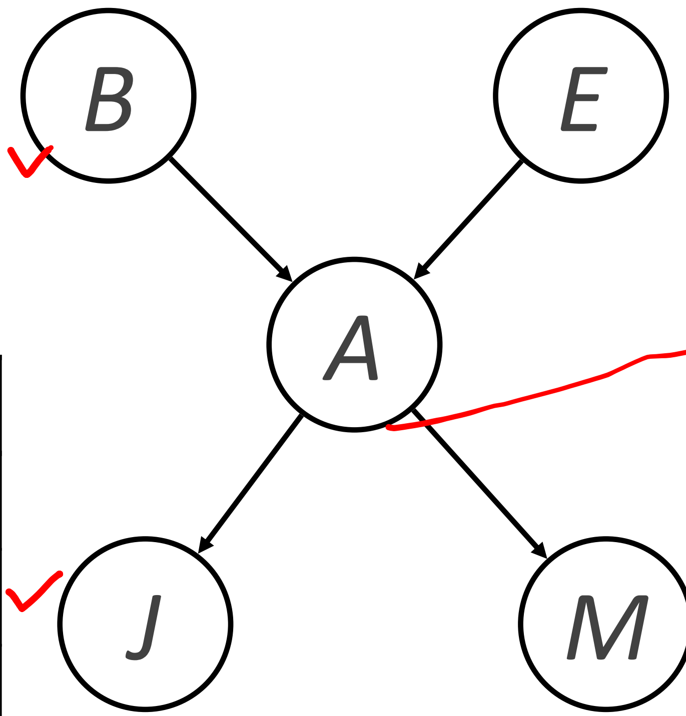
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$$P(+r, -t) = \frac{1}{4} \times \frac{1}{4}$$

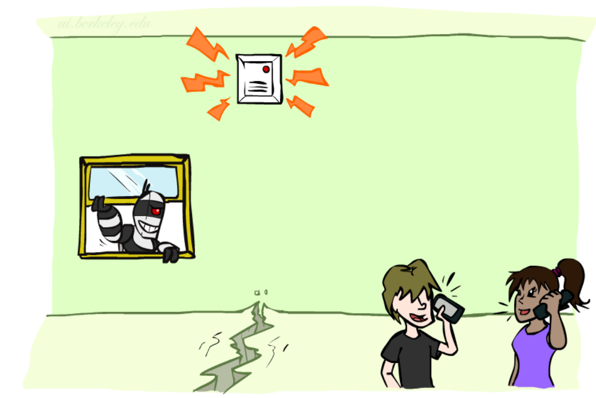


Quiz: Alarm Network

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

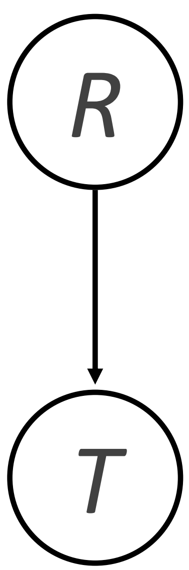
$$P(+b, -e, +a, -j, +m) =$$

$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$

$= 0.0000656684$

Example: Traffic

- Causal direction



$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

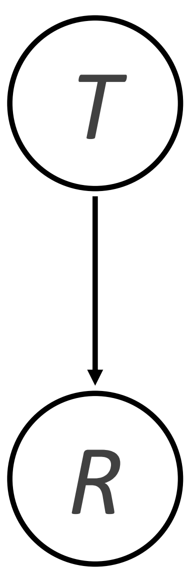
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$P(T, R)$

+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

$P(R|T)$

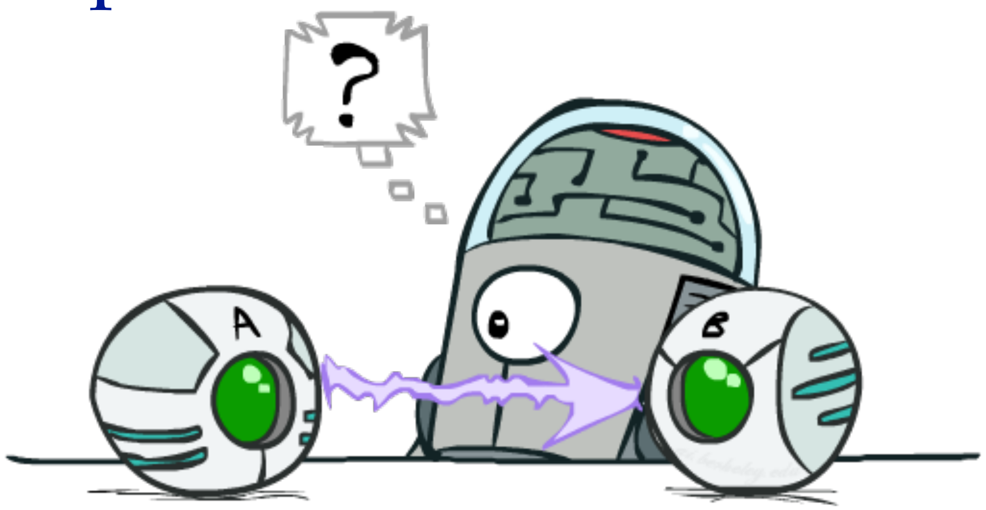
+t	+r	1/3
	-r	2/3
-t	+r	1/7
	-r	6/7

$P(T, R)$

+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

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

Size of a Bayes' Net

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

  2^N

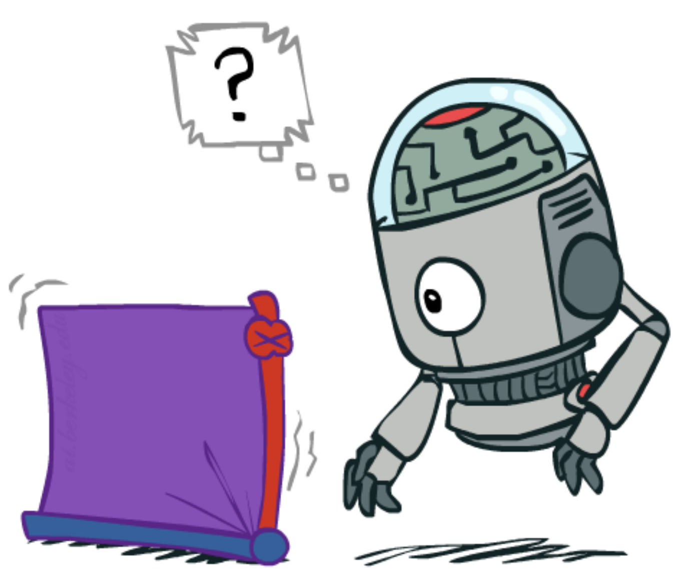
- How big is an N-node net if nodes have up to k parents?

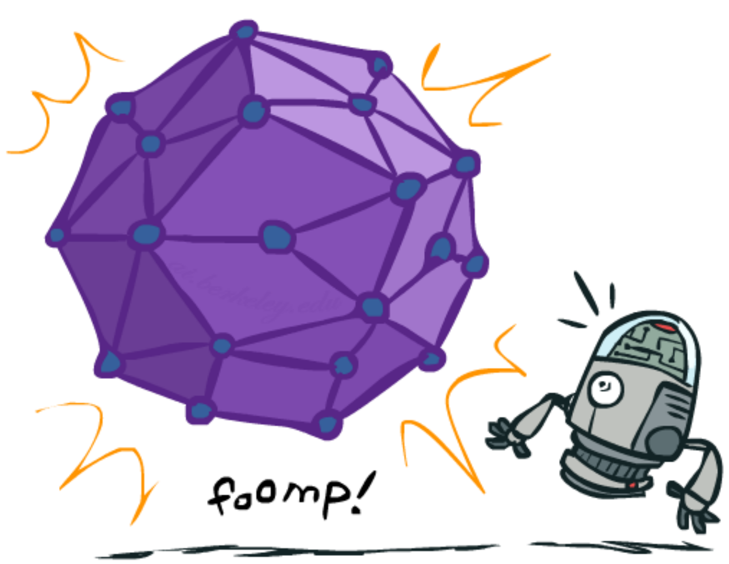
  $O(N * \underbrace{2^{k+1}}_{\text{CPT}})$

- Both give you the power to calculate

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

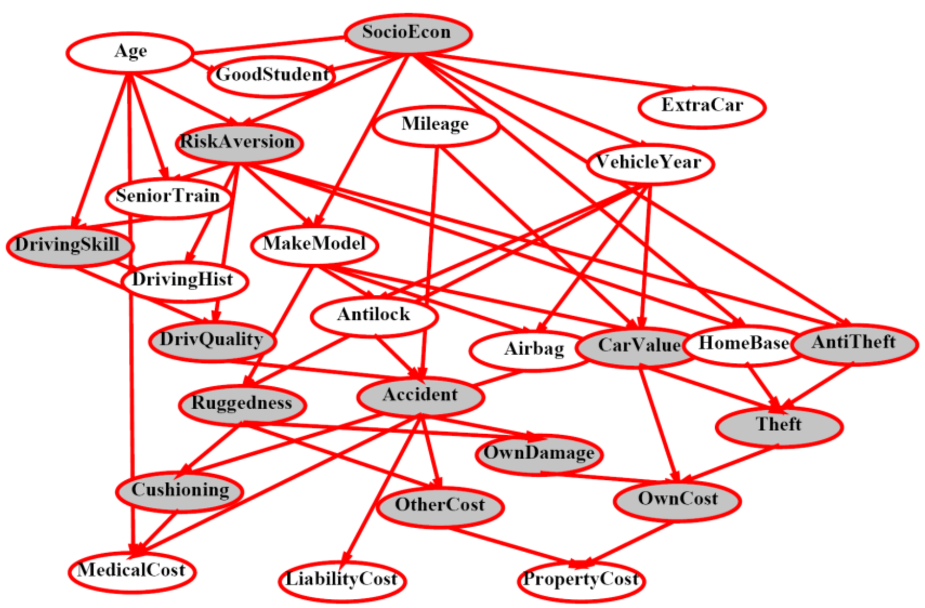
- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also faster to answer queries (coming)





Bayes' Nets

- 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 | e)?
  - Representation: given a BN graph, what kinds of distributions can it encode?
  - Modeling: what BN is most appropriate for a given domain?

Bayes' Nets

- ✓ Representation
- Conditional Independences
- Probabilistic Inference

Conditional Independence

- X and Y are independent if

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

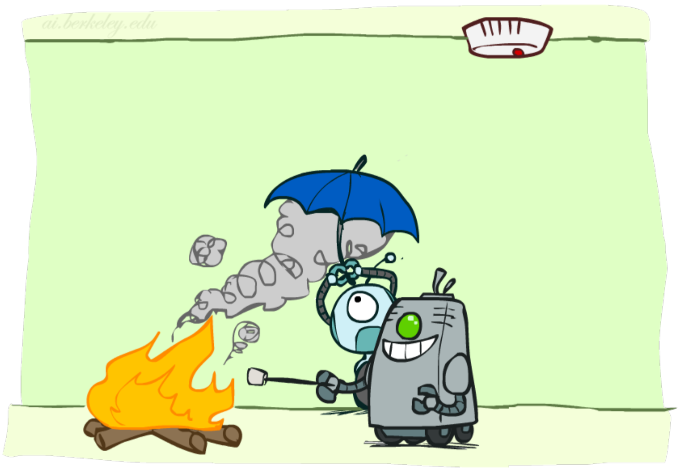
- X and Y are conditionally independent given Z

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

- (Conditional) independence is a property of a joint distribution

- Example:

$$Alarm \perp\!\!\!\perp Fire | Smoke$$



Bayes' Nets: Assumptions

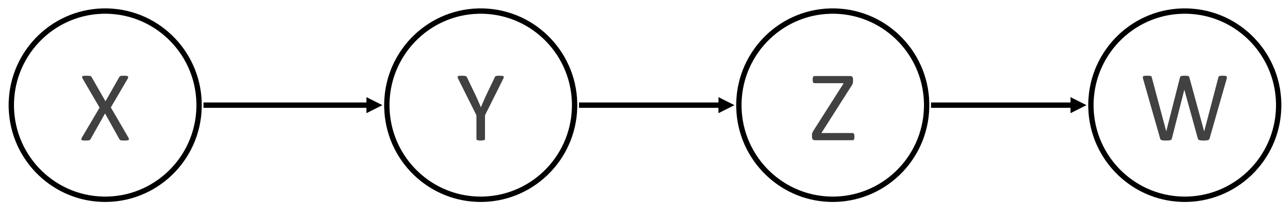
- Assumptions we are required to make to define the Bayes' net when given the graph:

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

- Beyond the above 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:



$P(X)$ $P(Y|X)$ $P(Z|Y)$ $P(W|Z)$

- Additional implied conditional independence assumptions?

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

$X \perp\!\!\!\perp W \mid Z$

$Y \perp\!\!\!\perp W \mid Z$

$X \perp\!\!\!\perp W \mid Y$

Independence in a BN

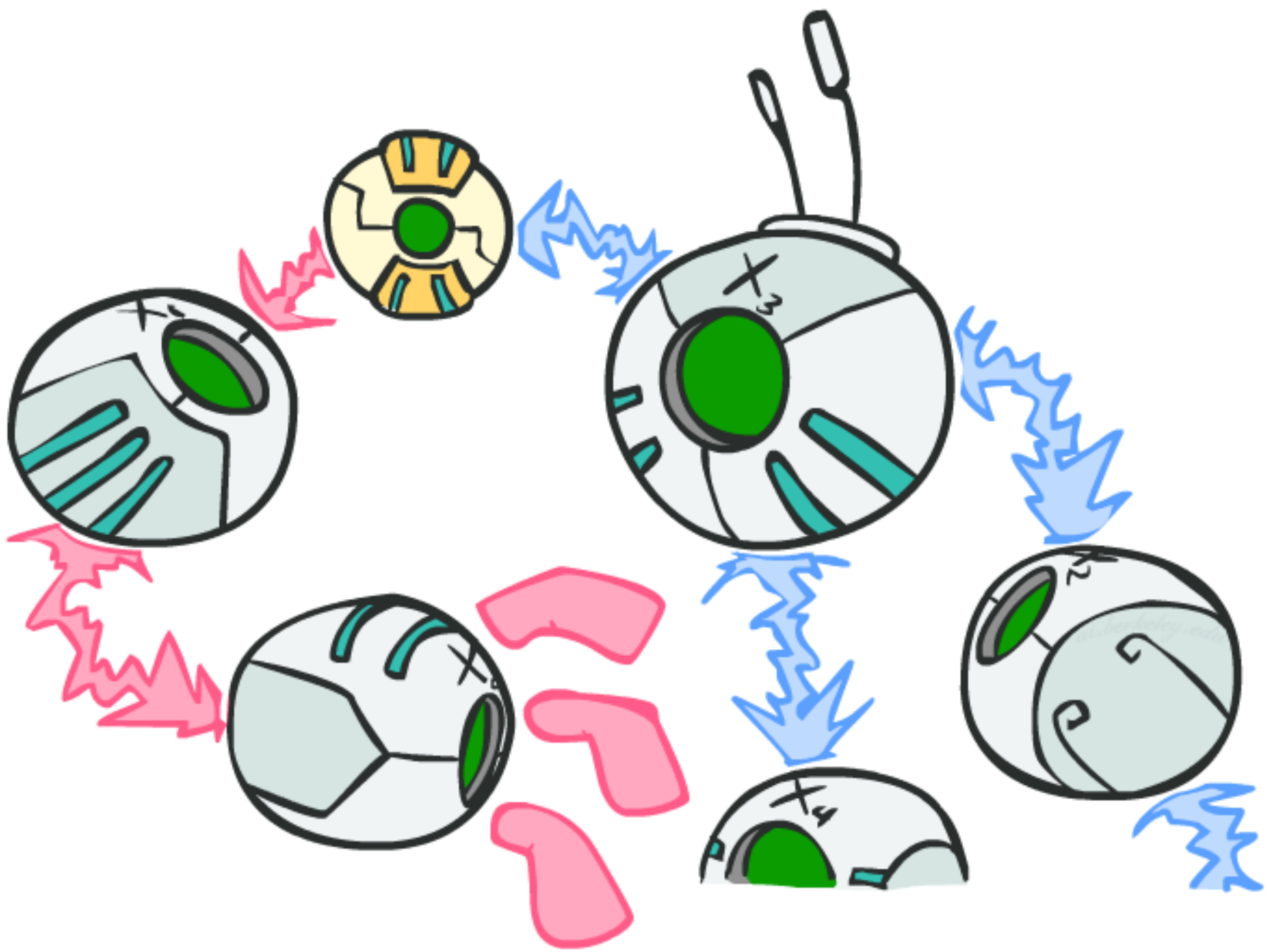
- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counter example
  - Example:

$$X \rightarrow Y \rightarrow Z$$

$P(X \mid Z) = P(X)$?

  - Question: are X and Z necessarily independent?
    - Answer: No, e.g., 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

Serial Chains

- This configuration is a "serial chain"

$X \to Y \to Z$

X: Low pressure Y: Rain Z: Traffic

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

- Is X always 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.
  - Counter-example:
    - Low pressure => rain => traffic, high pressure => no rain => no traffic
  - In numbers:

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

$P(+z) \neq P(+z \mid +x)$

$P(+x) = 0.5 = P(-x)$

Serial Chains

- This configuration is a "serial chain"

X: Low pressure Y: Rain Z: Traffic

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

- Is X always independent of Z given Y?

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

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

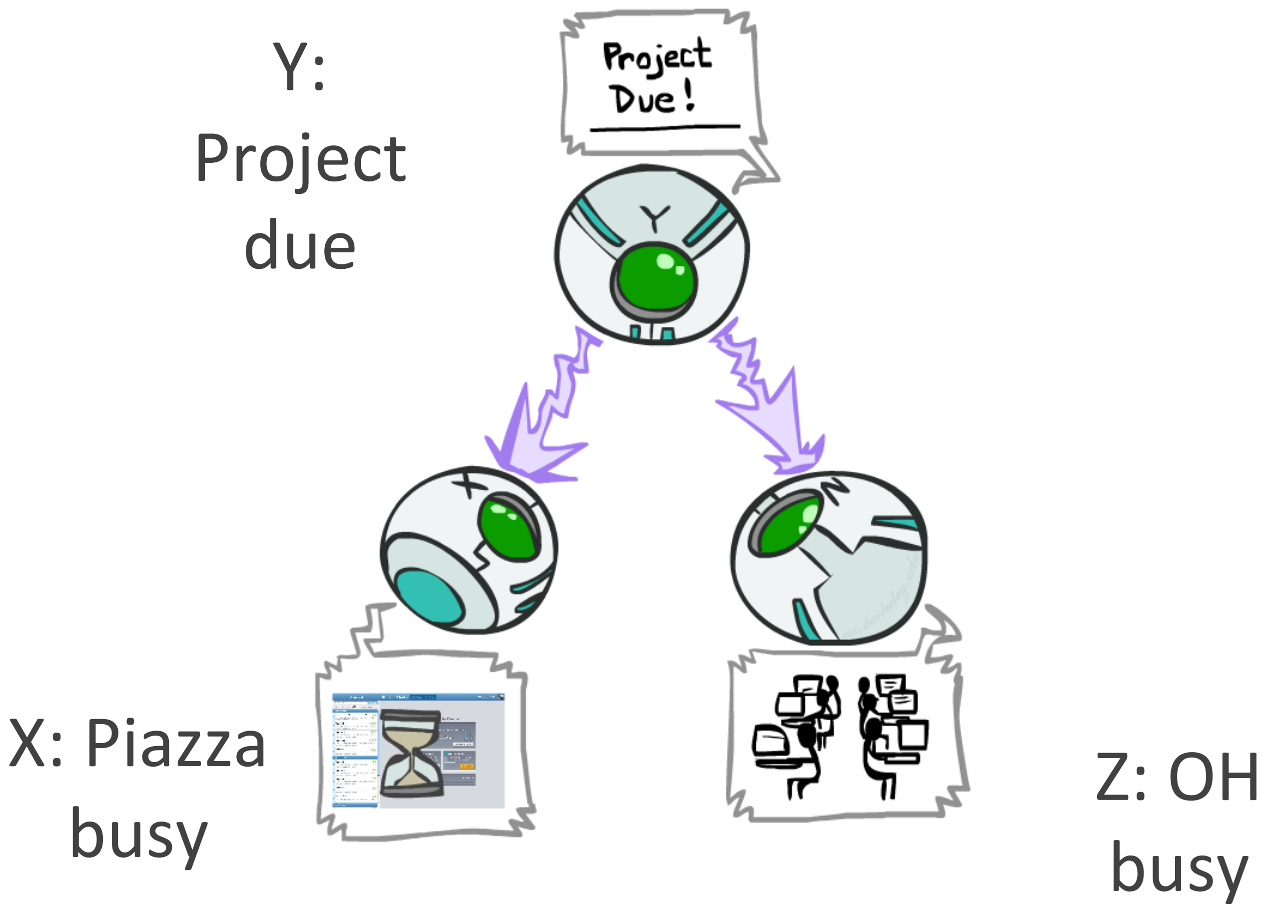
$$= P(z|y)$$

Yes!

- Evidence along the chain "blocks" the influence

Divergent Chain

- This configuration is a "divergent chain"



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

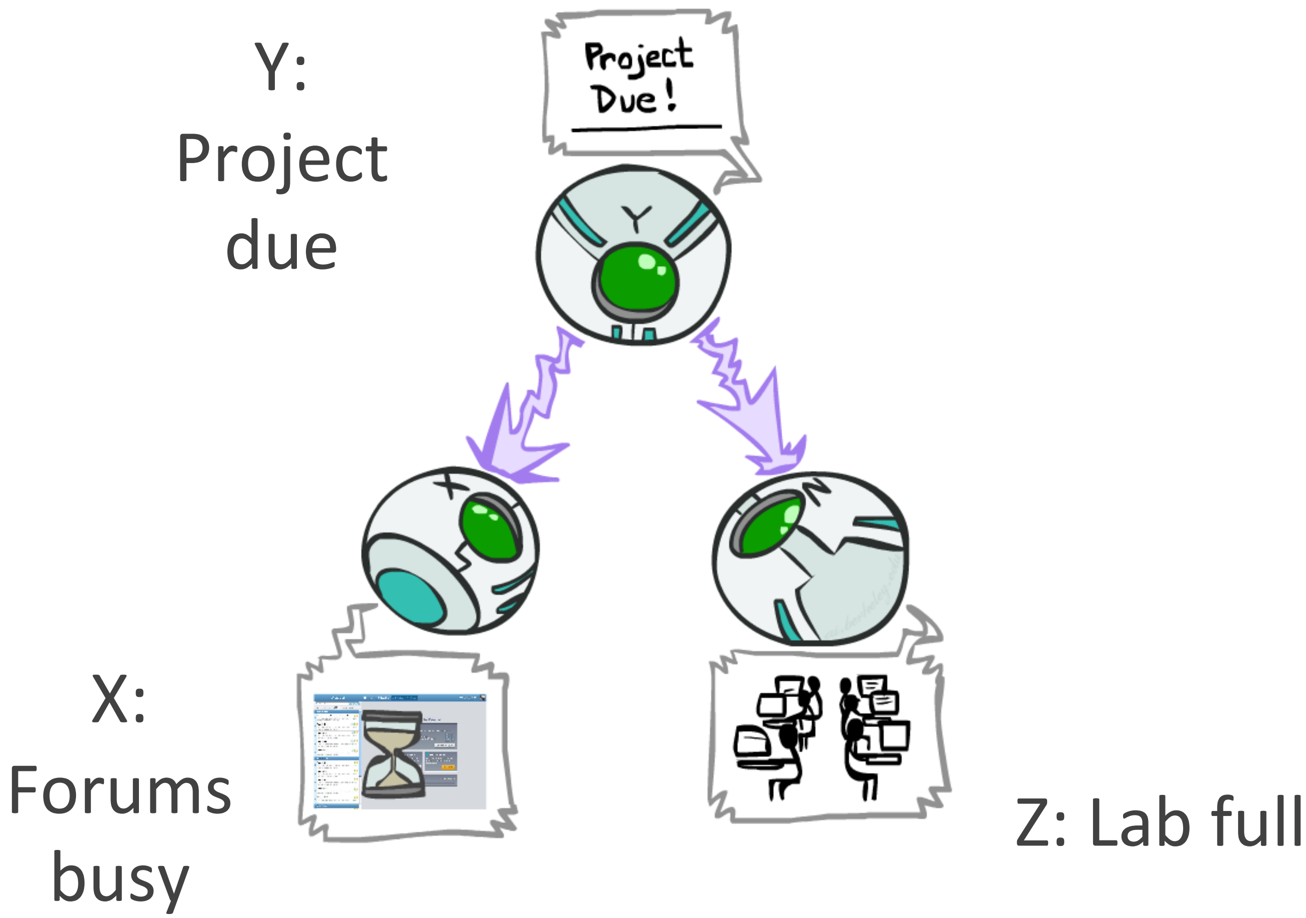
- Is X always independent of Z ?
  - *No!*
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
  - Counter-example:
    - Project due => Piazza busy and OH busy
    - In numbers:

P(+x | +y) = 1, P(-x | -y) = 1,
P(+z | +y) = 1, P(-z | -y) = 1

$P(z | x) = P(z)$?

Divergent Chain

- This configuration is a "divergent chain"



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

- Guaranteed X and Z independent given Y?

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

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

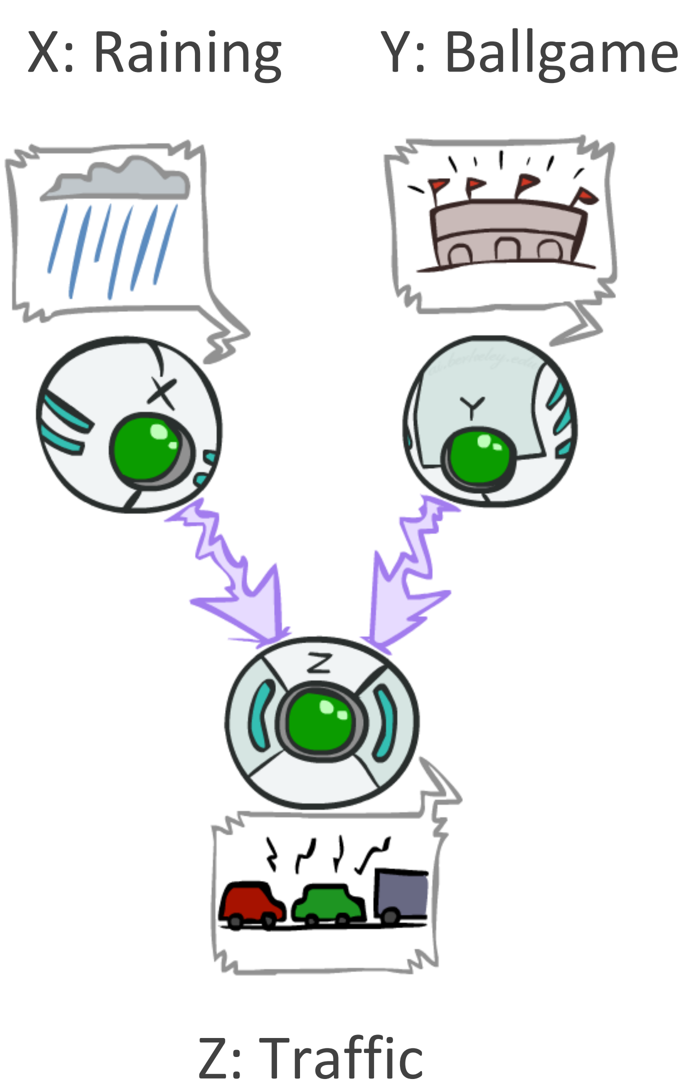
$$= P(z|y)$$

Yes!

- Observing the cause blocks influence between effects.

Convergent Chain

- Last configuration: "convergent chain" (v-structures)

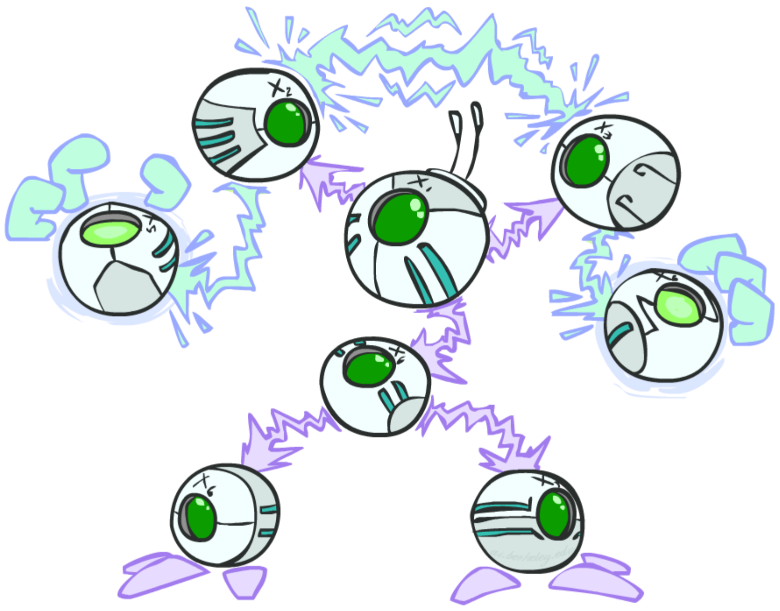


- Are X and Y independent?
  - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
  - Still need to prove they must be (try it!)
- Are X and Y independent given Z?
  - *No*: seeing traffic puts the rain and the ballgame in competition as explanation.
- This is backwards from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

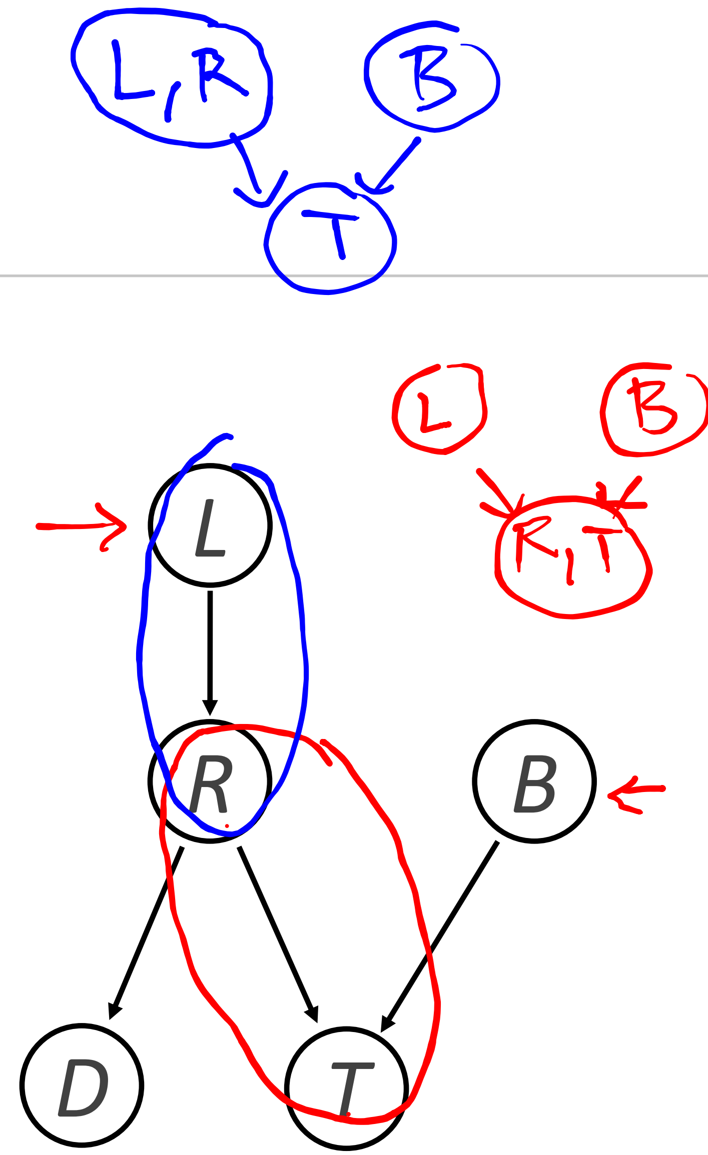


- **Solution**: analyze the graph

- Any complex example can be broken into repetitions of the three canonical cases

Reachability

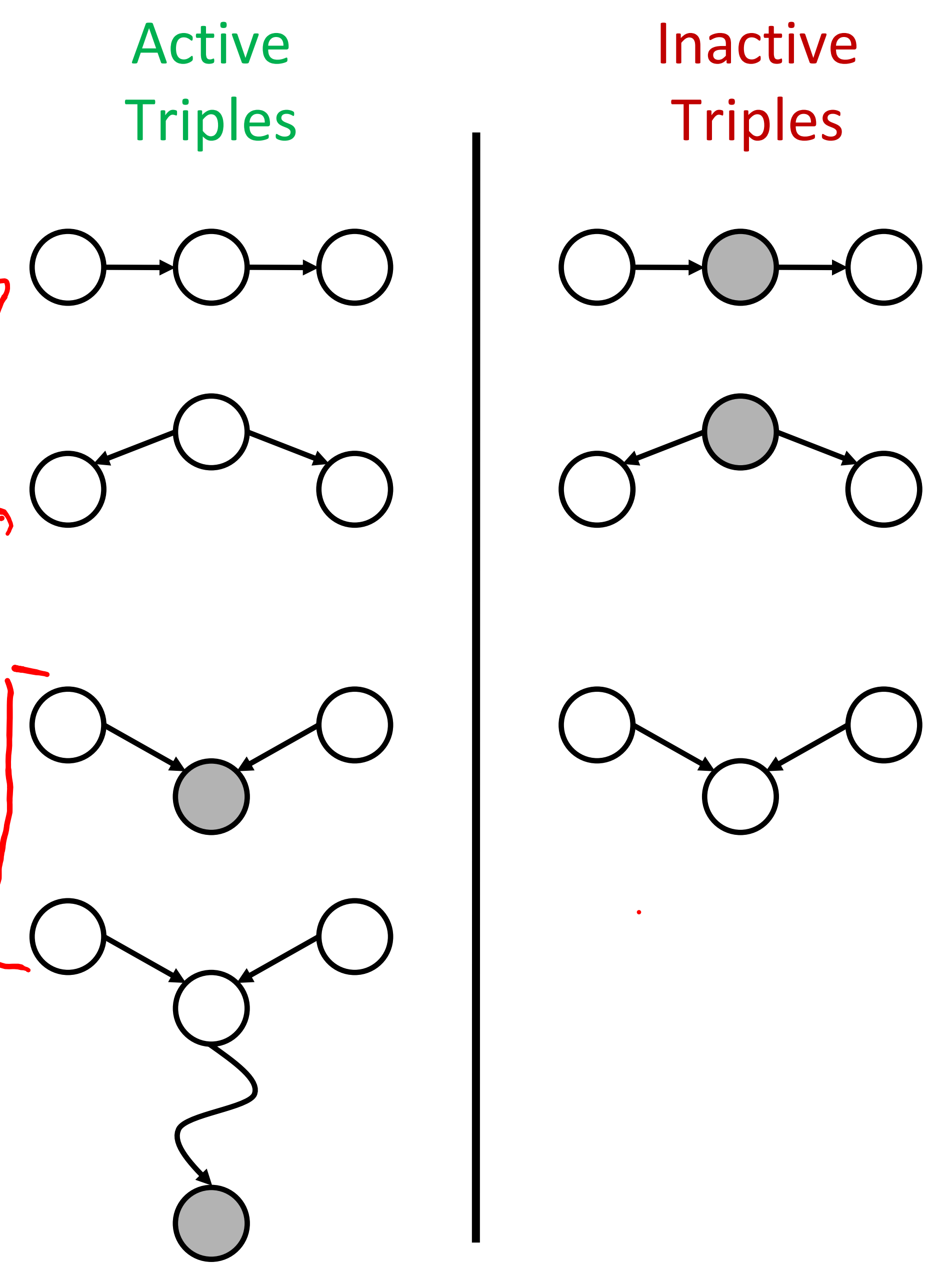
- **Recipe**: shade evidence nodes, look for paths in the resulting graph
- **Attempt 1**: if two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent
- Almost works, but not quite
  - Where does it break?
  - Answer: the v-structure at T doesn't 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:
  - Serial chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
  - Divergent chain $A \leftarrow B \rightarrow C$ where B is unobserved
  - Convergent chain (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



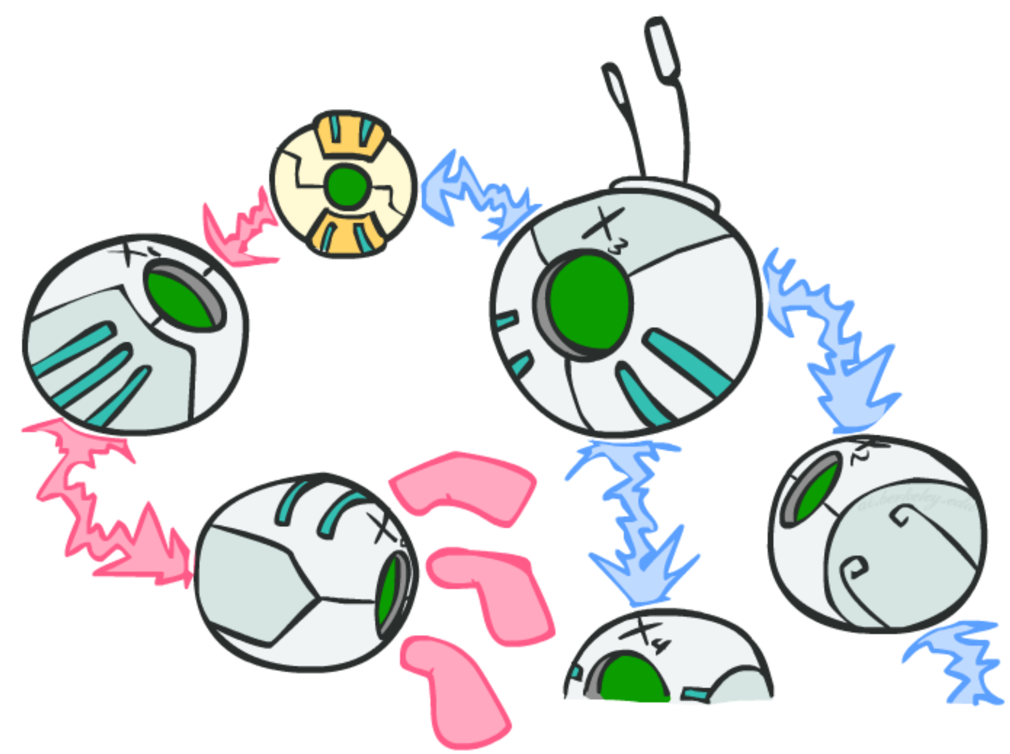
D-Separation

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

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

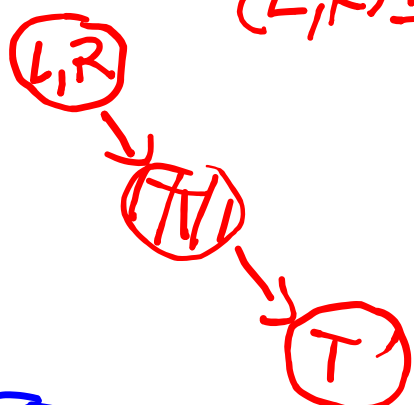
  - Otherwise (i.e. if all paths are inactive), then independence is guaranteed

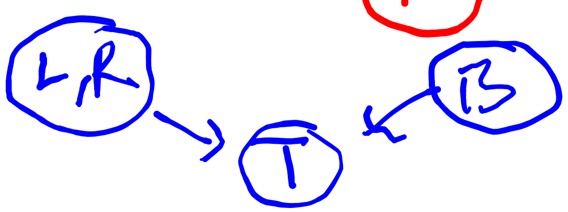
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

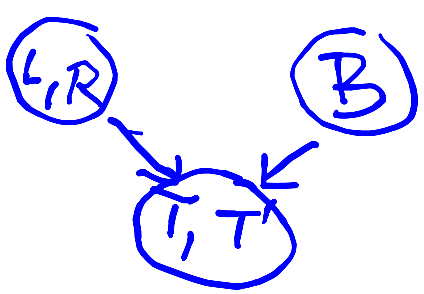


Example

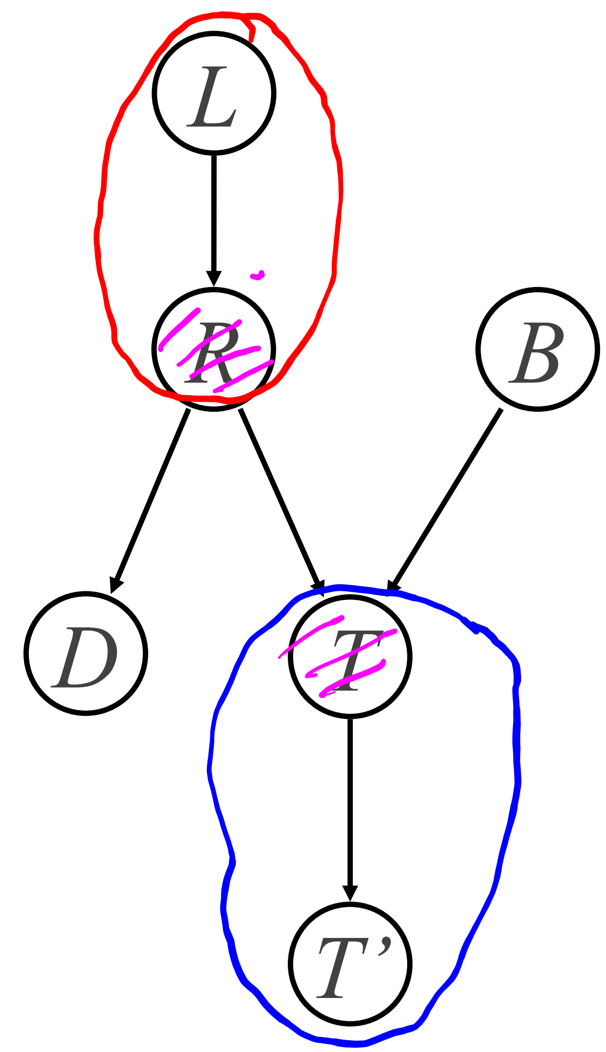
$(L,R) \perp\!\!\!\perp T' \mid T$





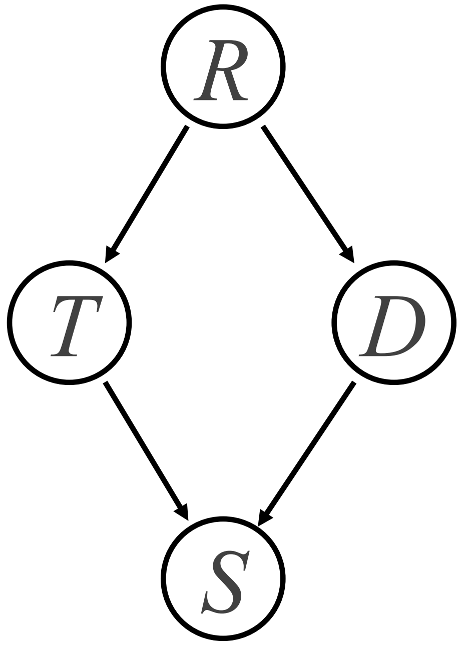


$L \perp\!\!\!\perp T' \mid \underline{T}$	Yes
$L \perp\!\!\!\perp B$	Yes
$L \perp\!\!\!\perp B \mid \underline{T}$	No
$L \perp\!\!\!\perp B \mid T'$	No
$L \perp\!\!\!\perp B \mid T, R$	Yes



Quiz: Conditional Independence

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



- Questions:

$T \perp\!\!\!\perp D$ No

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

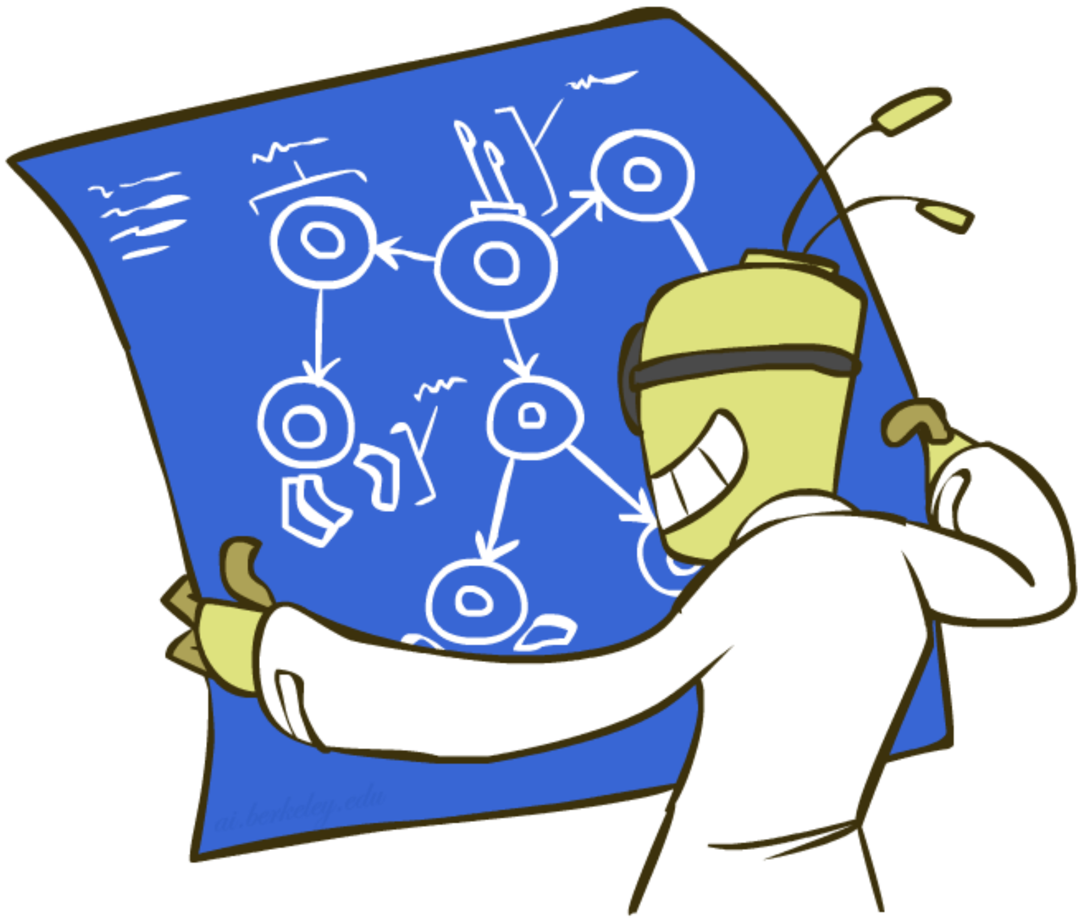
$T \perp\!\!\!\perp D | R, S$ No

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

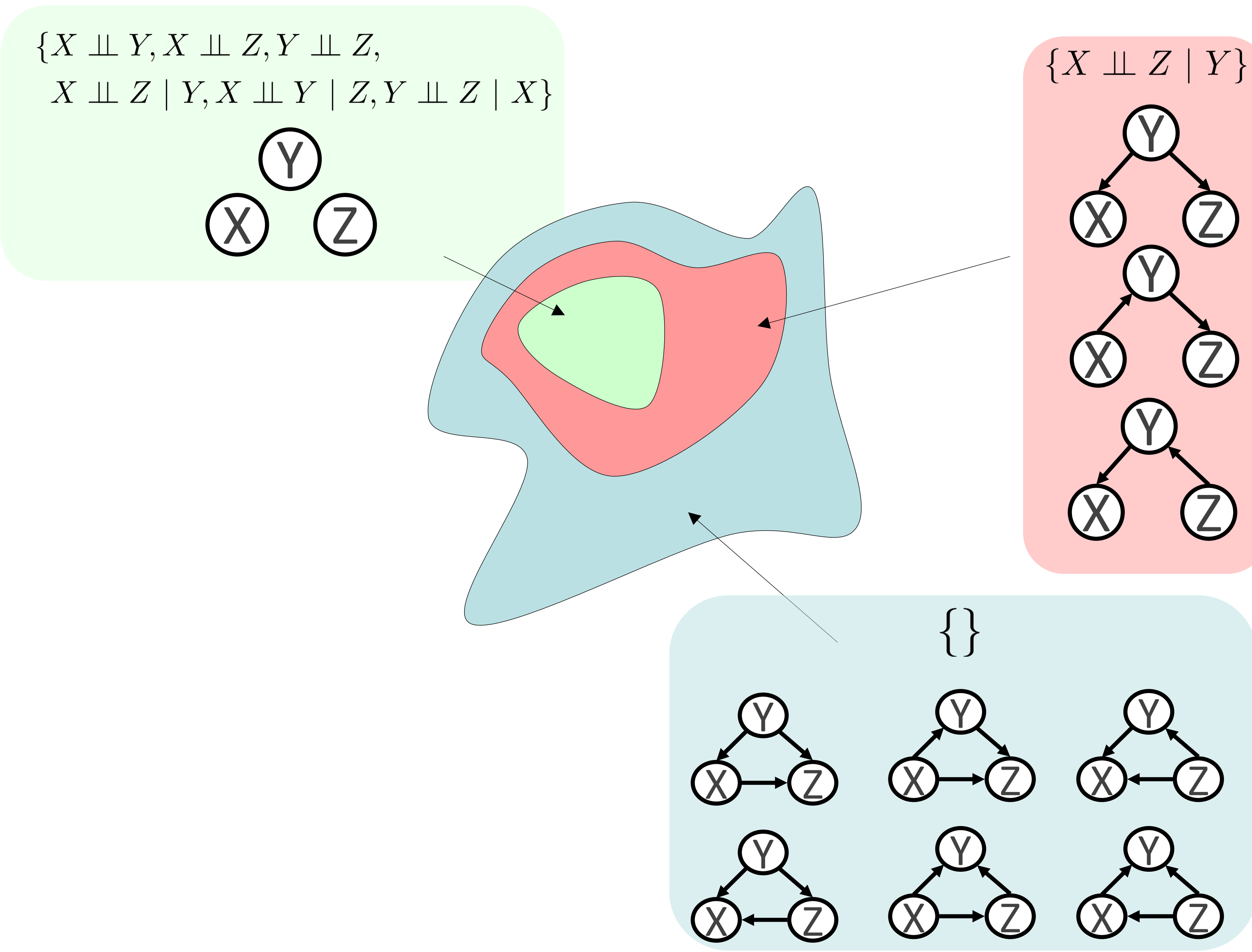
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

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



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
  - Approximate inference (sampling)