Announcements

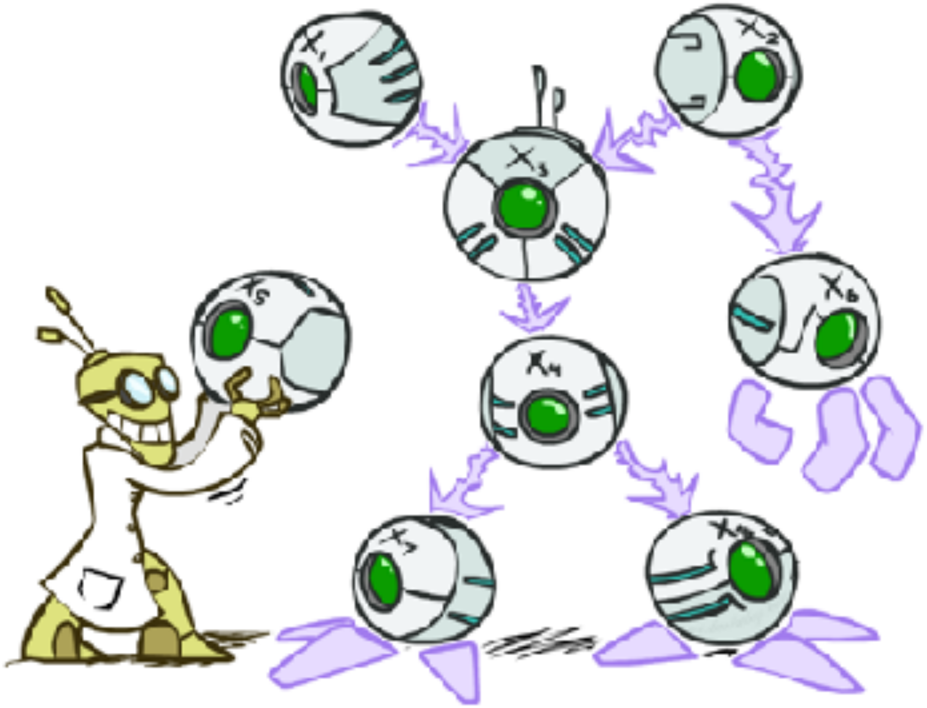
- Midterm
  - Next Wednesday, 8-9:30 pm
  - See Piazza post for location, other details
  - If you need DSP accommodations and haven't been contacted, email rich.zhang@eecs.berkeley.edu ASAP

- Homework 6
  - Due Monday 3/13 at 11:59pm

- Review sessions
  - F 11-1, 306 Soda: Review lecture
  - M 12-3, 540AB Cory: Guerrilla section
  - Exam-prep next week: each is focused on a different subset of topics
    - Schedule will be announced soon on Piazza

CS 188: Artificial Intelligence

Bayes' Nets



Lecturer: Davis Foote --- 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.

Slides on d-separation were created by Davis Foote.]

Conditional Independence

- X is conditionally independent of Y given Z $\quad X \perp\!\!\!\perp Y \mid Z$

  if and only if:

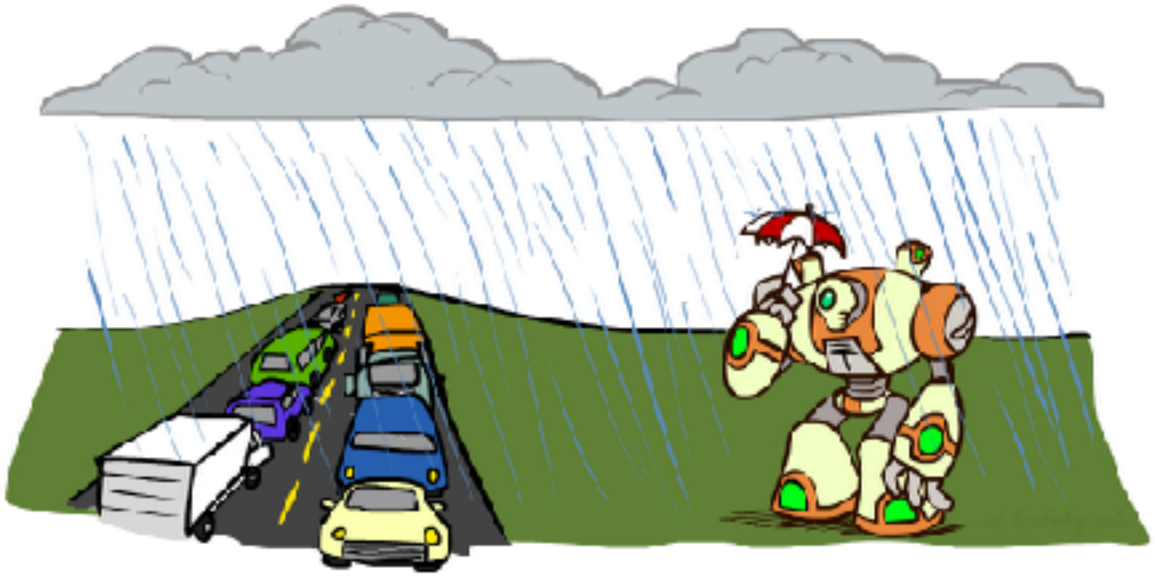
  $$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

  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 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:

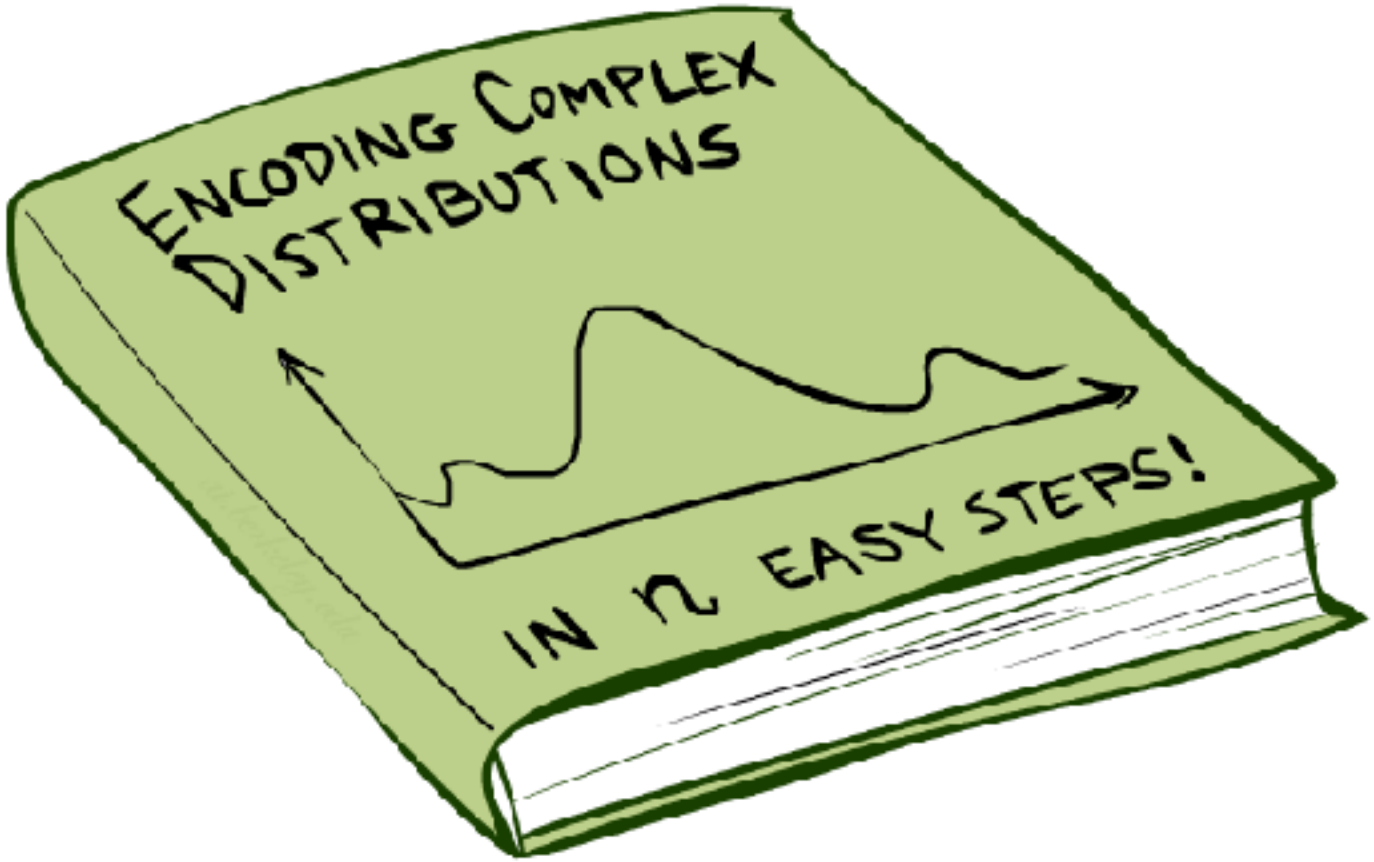
$$P(\text{Traffic}, \text{Rain}, \text{Umbrella}) = P(\text{Rain})P(\text{Traffic}|\text{Rain})P(\text{Umbrella}|\text{Rain}, \text{Traffic})$$

- 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 help us express conditional independence assumptions

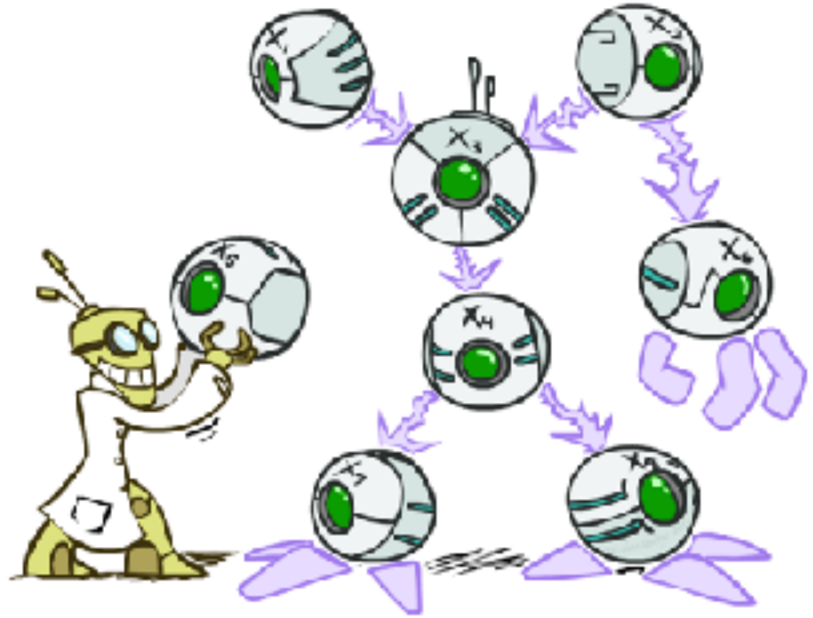
Bayes'Nets: Big Picture



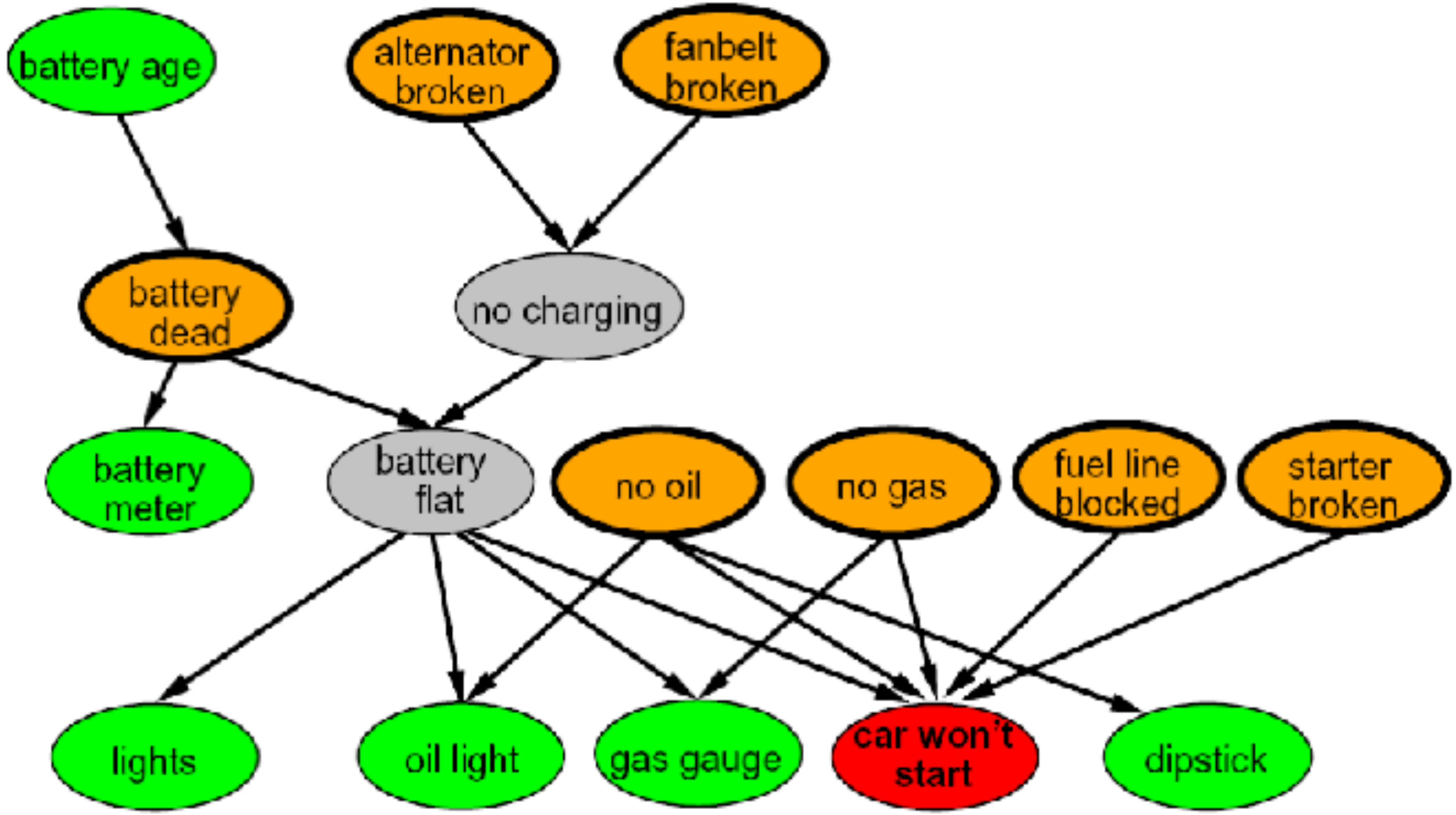
Bayes' Nets: Big Picture

- 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
  - Unintuitive. Where do these probabilities come from?

- 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

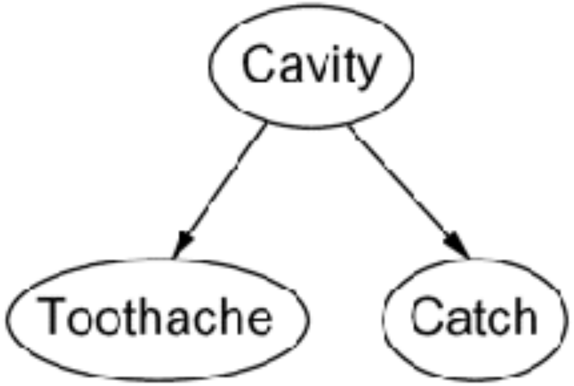


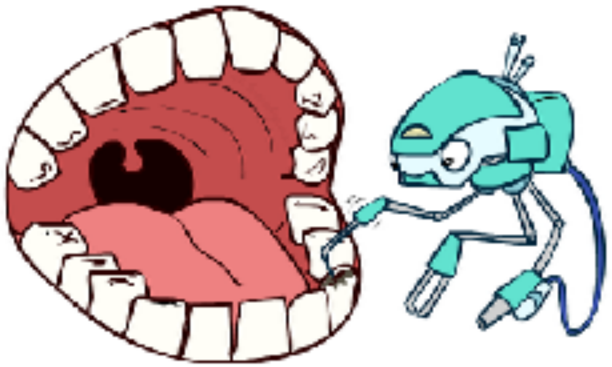
Example Bayes' Net: Car



Graphical Model Notation

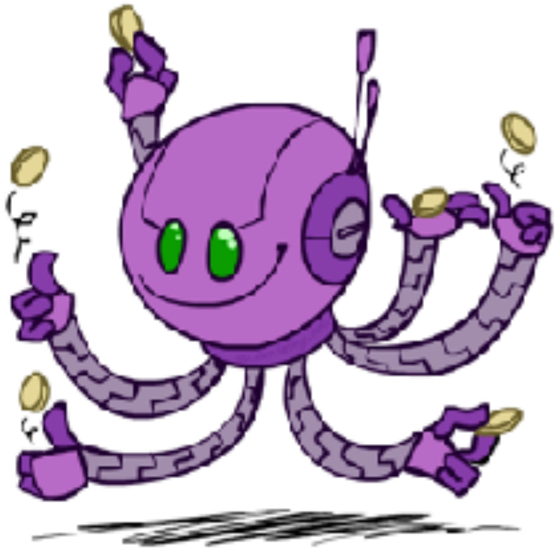
- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions
  - Indicate “direct influence” between variables
  - Formally: encode conditional independence assumptions (more later)
- For now: imagine that arrows mean direct causation (in general, they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





Example: Coin Flips

- N independent coin flips

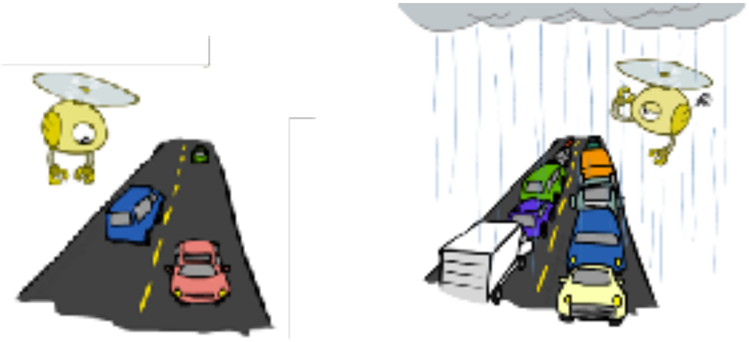


$$X_1 \qquad X_2 \qquad \cdots \qquad X_n$$

- No interactions between variables: absolute independence

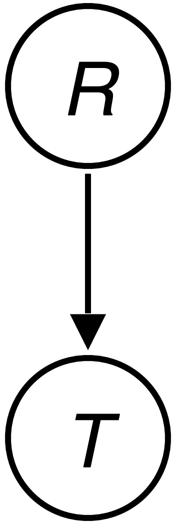
Example: Traffic

- 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!

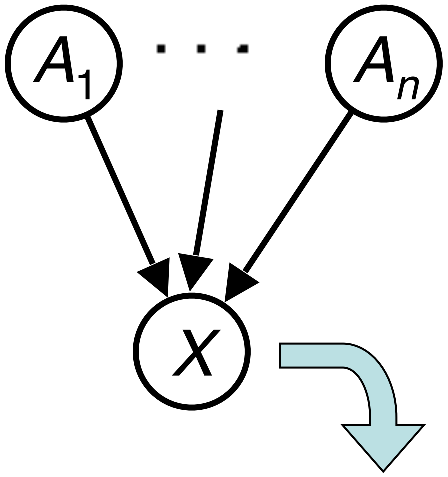
Bayes' Net Semantics

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

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

  - CPT: conditional probability table



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

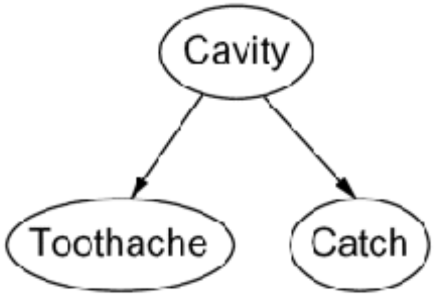
A Bayes net = Topology (graph) + Local Conditional Probabilities

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:



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

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

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

Example: Coin Flips

X_1 X_2 ... X_n

$P(X_1)$

h	0.5
t	0.5

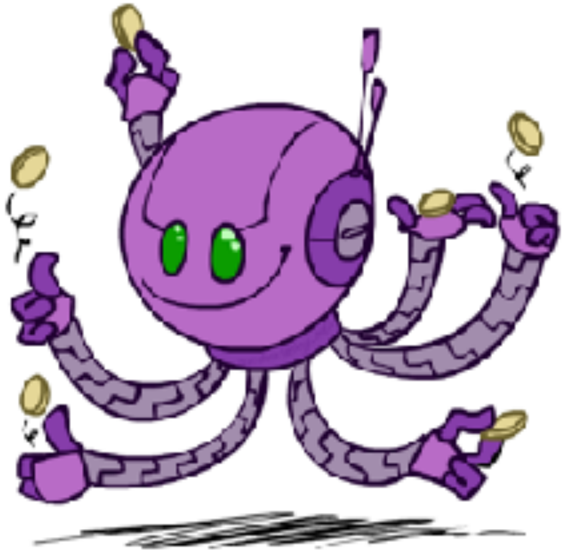
$P(X_2)$

h	0.5
t	0.5

...

$P(X_n)$

h	0.5
t	0.5

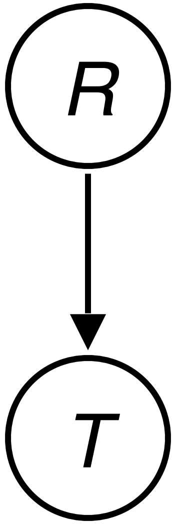


$$P(h, h, t, h) =$$

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

Note: Now that we know the true meaning of an edge, we can see that it doesn't necessarily mean causality. We could add an edge between coins and still have a valid BN representing the same distribution; just because X_1 and X_2 are independent doesn't mean we can't write $P(X_2 \mid X_1)$.

Example: Traffic



$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

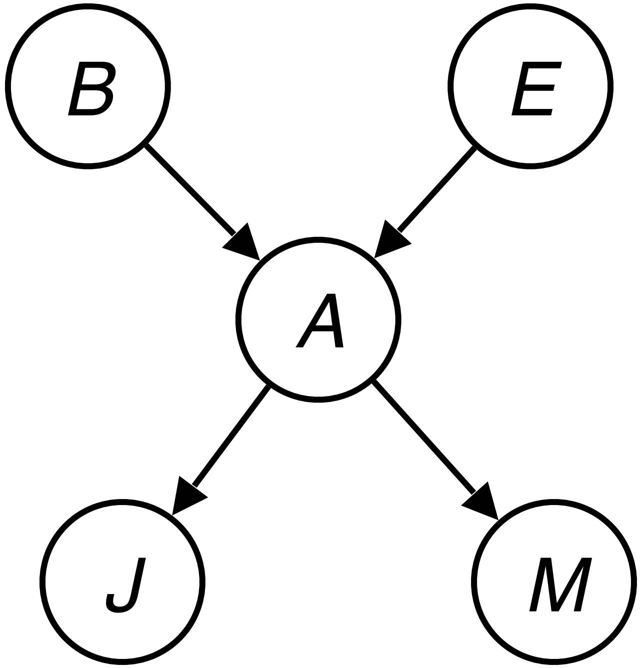
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

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

On the point of causality: it’s also possible to compute P(T) and P(R | T) using Bayes’ rule, so we could even have a Bayes’ net with the edge reversed to represent this distribution.

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

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

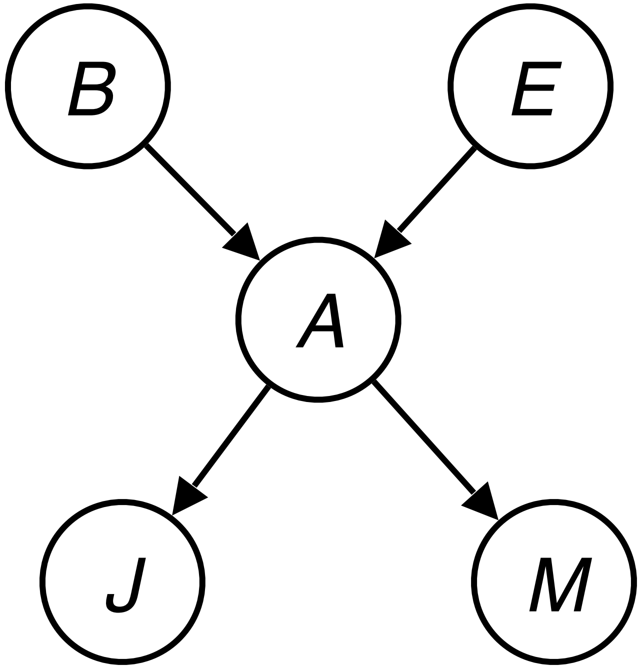
A	M	$P(M \mid A)$
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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

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

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

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

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

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

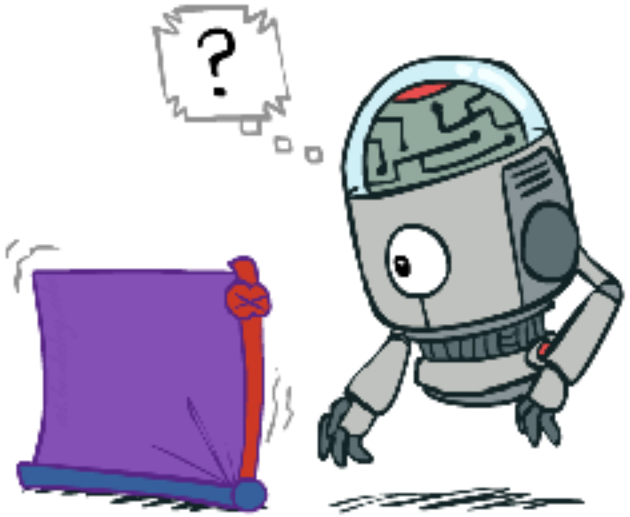
$$P(+b, -e, +a, -j, +m) =$$
$$P(+b)P(-e)P(+a|+b,-e)P(-j|+a)P(+m|+a) =$$
$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

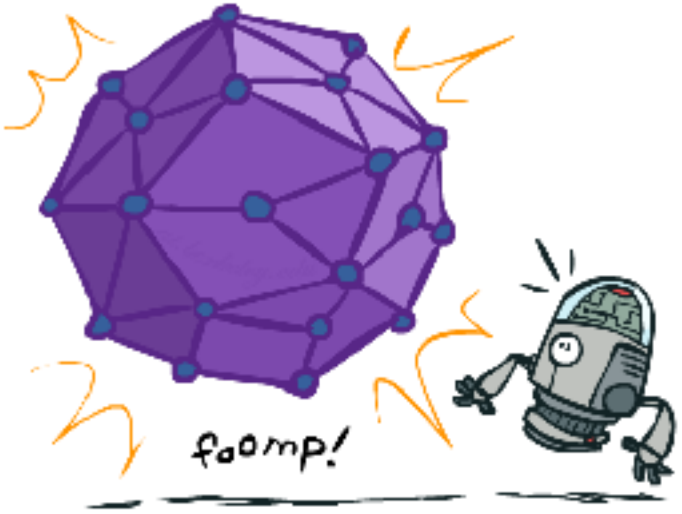
Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?
  - 2^N
- How big is an N-node net if nodes have up to k parents?
  - $O(N * 2^{k+1})$
- Both give you the power to calculate

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

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





CS 188: Artificial Intelligence

Bayes' Nets: Independence

Lecturer: Davis Foote --- 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.

Slides on d-separation were created by Davis Foote.]

Bayes' Nets

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

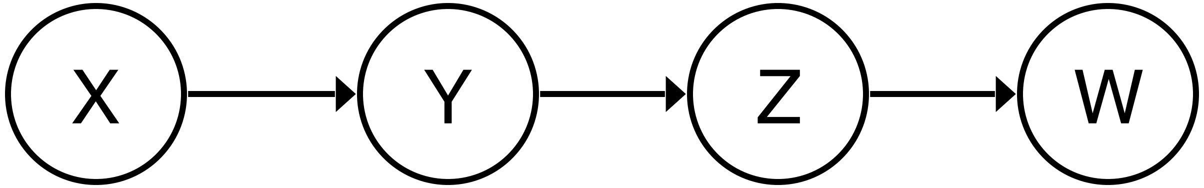
Bayes Nets: Assumptions

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

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

- Beyond above "chain rule → Bayes net" conditional independence assumptions
  - Often additional conditional independences
  - They can be read off the graph
- Why do we care?
  - Modeling: understand assumptions made when choosing a Bayes net graph
  - Interpretation: see the consequences of your choice of model

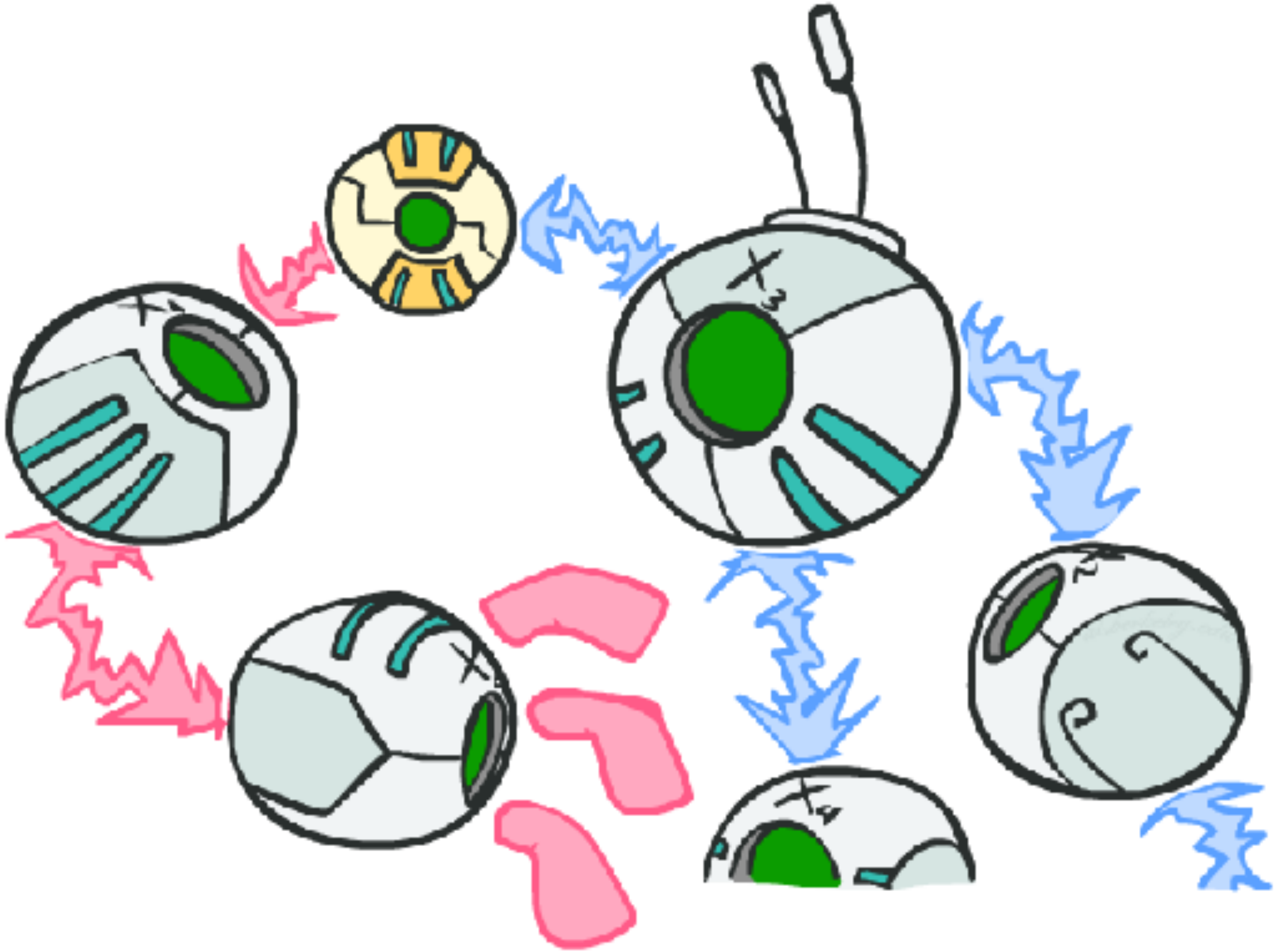
Example



- Conditional independence assumptions directly from simplifications in chain rule:

- Additional implied conditional independence assumptions?

D-separation



Independence in a BN

- Question of the day: are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counterexample
  - Let’s look at some examples
    - (turns out any larger example can be reduced to these smaller ones we are about to cover)

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!*
  - Counterexample:
    - Low pressure causes rain causes traffic, high pressure causes no rain causes no traffic
    - In numbers:

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

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?

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

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

$$= P(z|y)$$

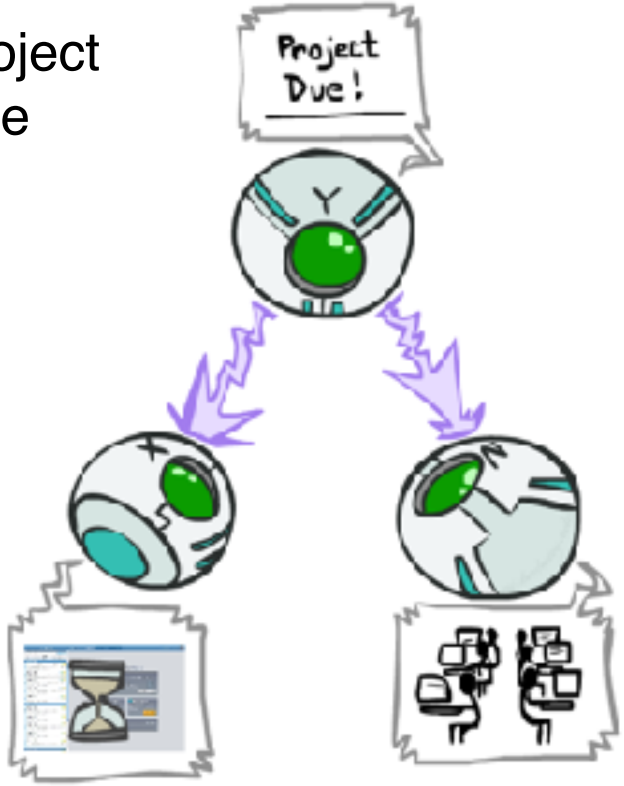
Yes!

- Evidence along the chain “blocks” the influence

Common Cause

- This configuration is a "common cause"

Y: Project due



X: Forums busy

Z: Lab full

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

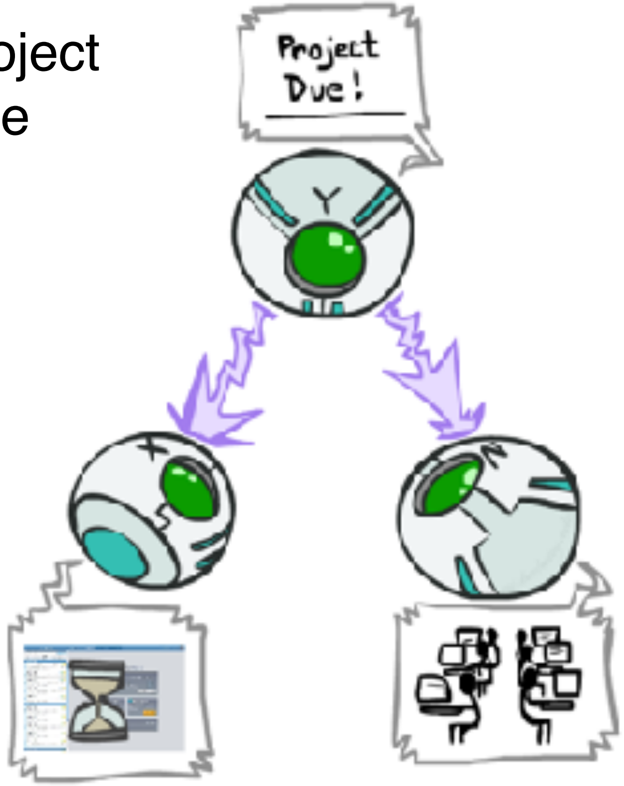
- Guaranteed X independent of Z ? *No!*
  - Counterexample:
    - Project due causes both forums busy and lab full
    - In numbers:

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

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?

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

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

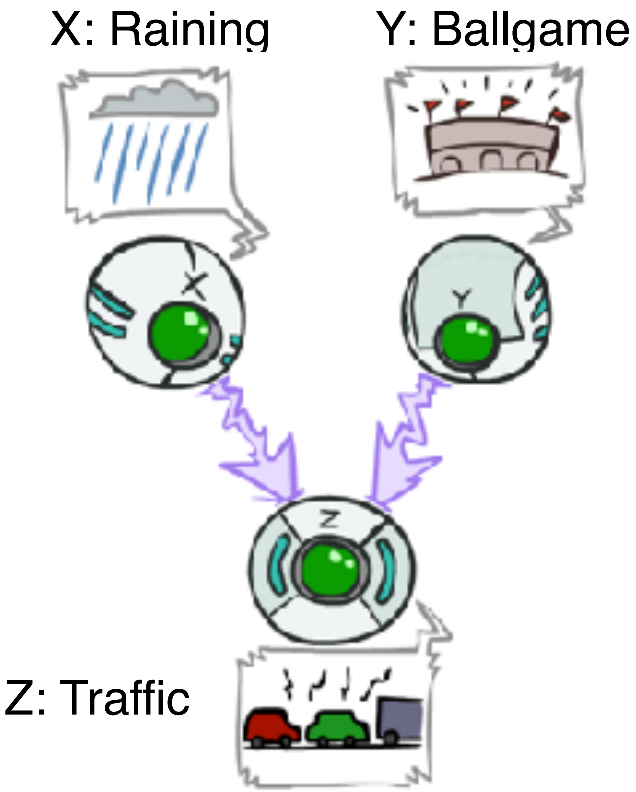
$$= P(z|y)$$

Yes!

- Observing the cause blocks influence between effects.

Common Effect

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

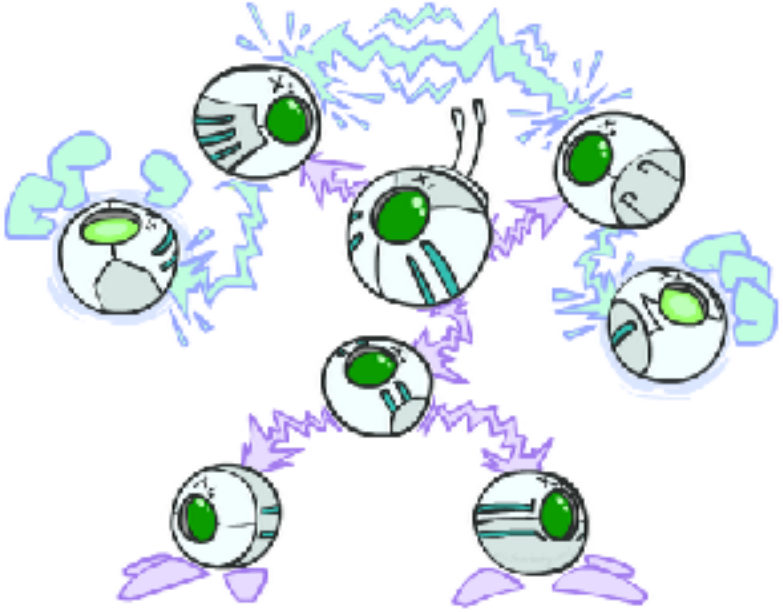


- Are X and Y independent?
  - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
  - Still need to prove they must be (try it!)
- Are X and Y independent given Z?
  - *No*: seeing traffic puts the rain and the ballgame in competition as explanation.
  - Can be shown with numbers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

The General Case

The General Case

- General question: in a given BN, are two variables independent (given evidence)?
- Solution: analyze the graph

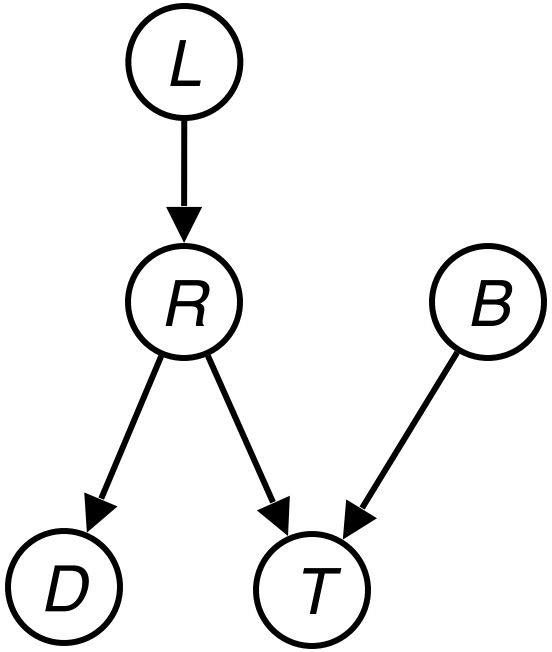


Reachability

- 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, call them "connected"
  - If variables are **not** connected, i.e. "separated," they can't influence each other through any path in the graph
  - Hence they are conditionally independent given the shaded nodes!

- 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"
  - How can we fix this?

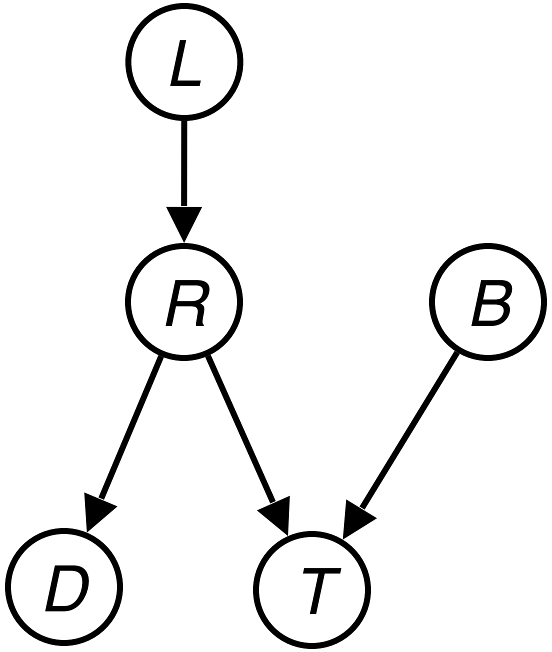


Fixing Reachability

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

- Attempt 2:
  - if two nodes are connected by an undirected path not blocked by a shaded node…
  - …**except** for v-structures, where the rules are reversed and the node **must** be shaded…
  - … call them "d-connected."
  - If variables are **not** connected, i.e. "d-separated," they can't influence each other through any path in the graph
  - Hence they are conditionally independent given the shaded nodes!

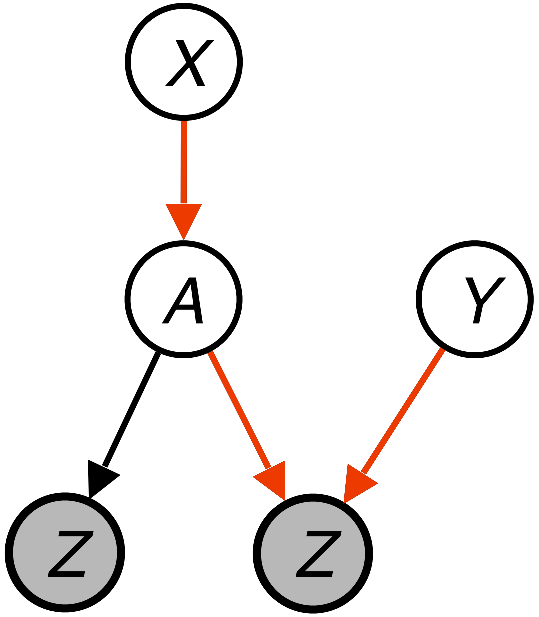
- Let's formalize this into an algorithm



D-Separation

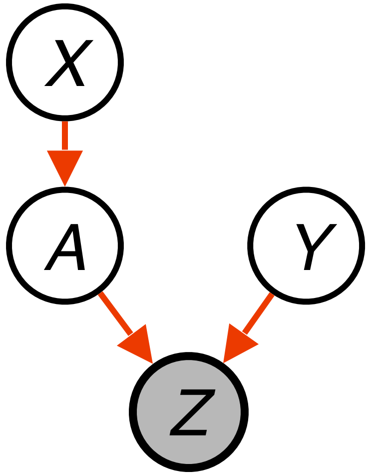
- Goal: determine whether tw nodes es X and Y d-separated ıdent given a set of conditional variables Z

- Steps:
  1. Shade all nodes corresponding to variables in set Z
  2. Enumerate all candidate paths (ignoring direction of edges) from X to Z
  3. For each path, check if it “d-connects” X and Y
  4. If no path d-connects them, they are d-separated!



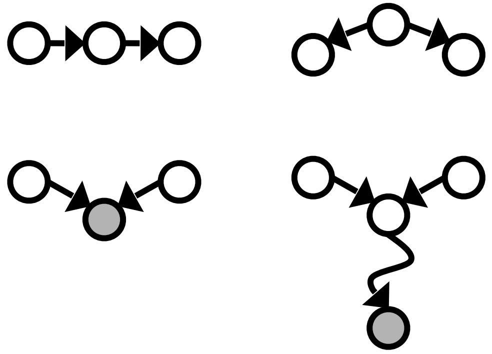
$Z = \{Z_1, Z_2\}$

Active / Inactive Paths

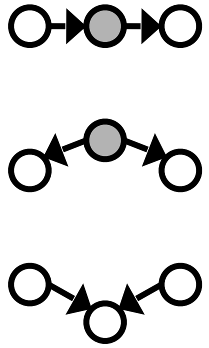


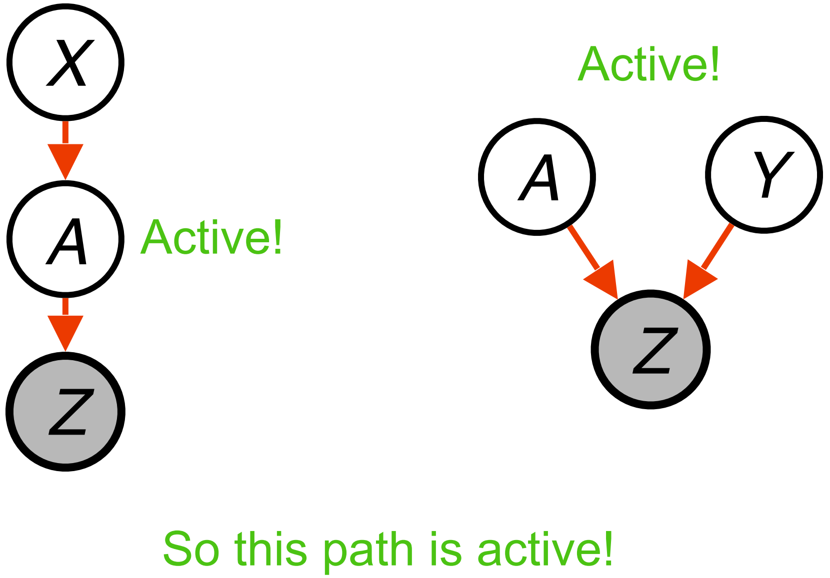
- Question: How can we formalize the “check if this path d-connects X and Y” step?
  - Call a path that d-connects X and Y an “**active path**”
  - Any path in a graph can be decomposed into triples of variables
- A path is active if and only if each triple is active
  - A triple is active or not depending on the middle node (see table)

Active Triples



Inactive Triples

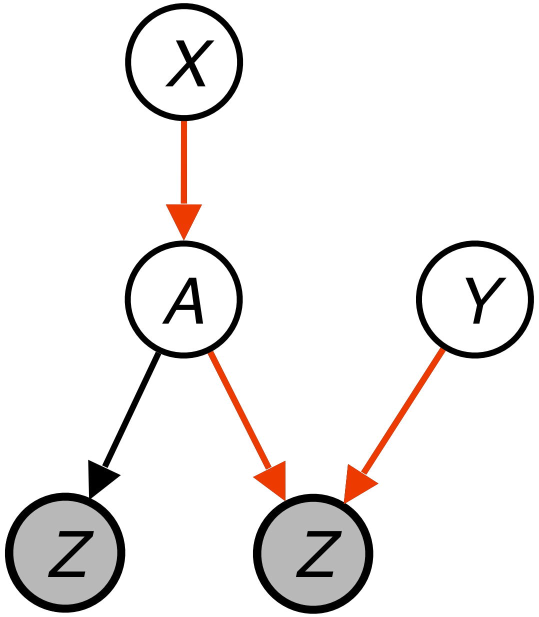




D-Separation: Full picture

- Goal: determine whether two variables X and Y are conditionally independent given a set of conditional variables Z

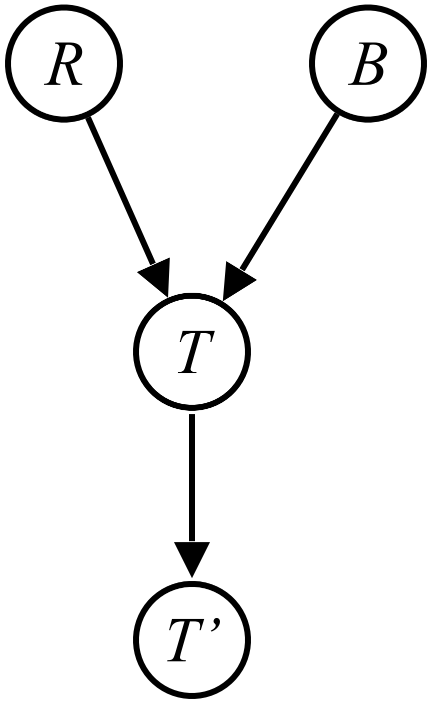
- Steps:
  1. Shade all nodes corresponding to variables in set Z
  2. Enumerate all candidate paths (ignoring direction of edges) from X to Z
  3. For each path:
     1. Decompose the path into triples
     2. If all triples are active, this path is active and d-connects them
  4. If no path is active, they are d-separated



$Z = \{Z_1, Z_2\}$

Example

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



Example

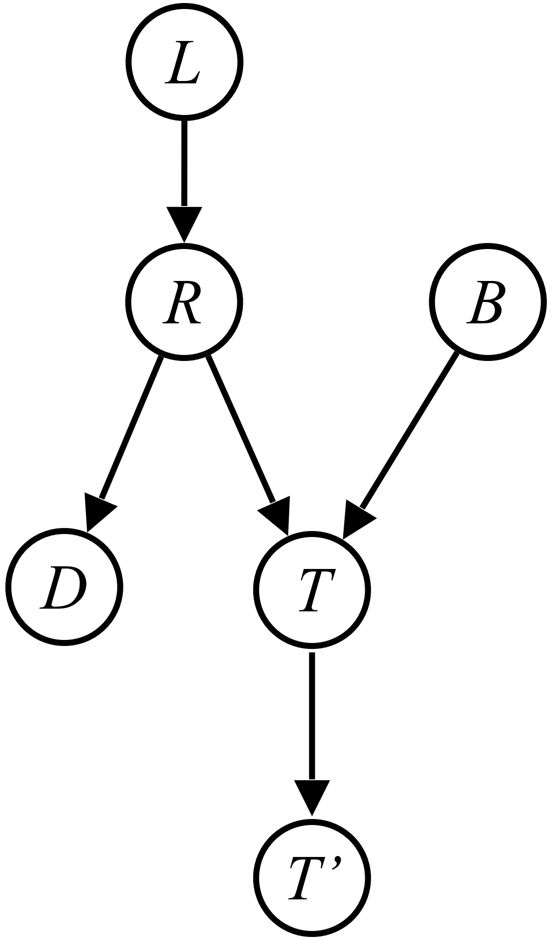
$L \perp\!\!\!\perp T' | T$ Yes

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

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

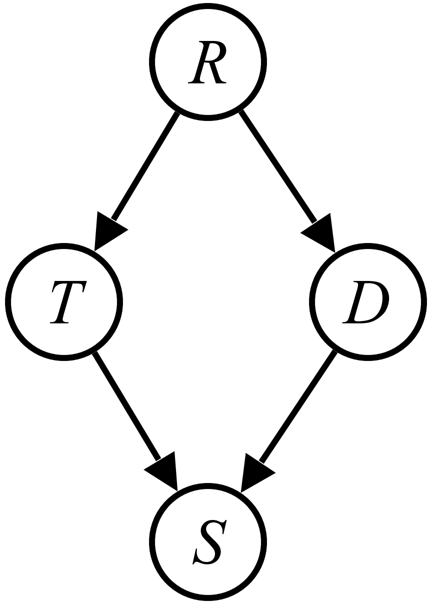
$L \perp\!\!\!\perp B | T'$ No

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



Example

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



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

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

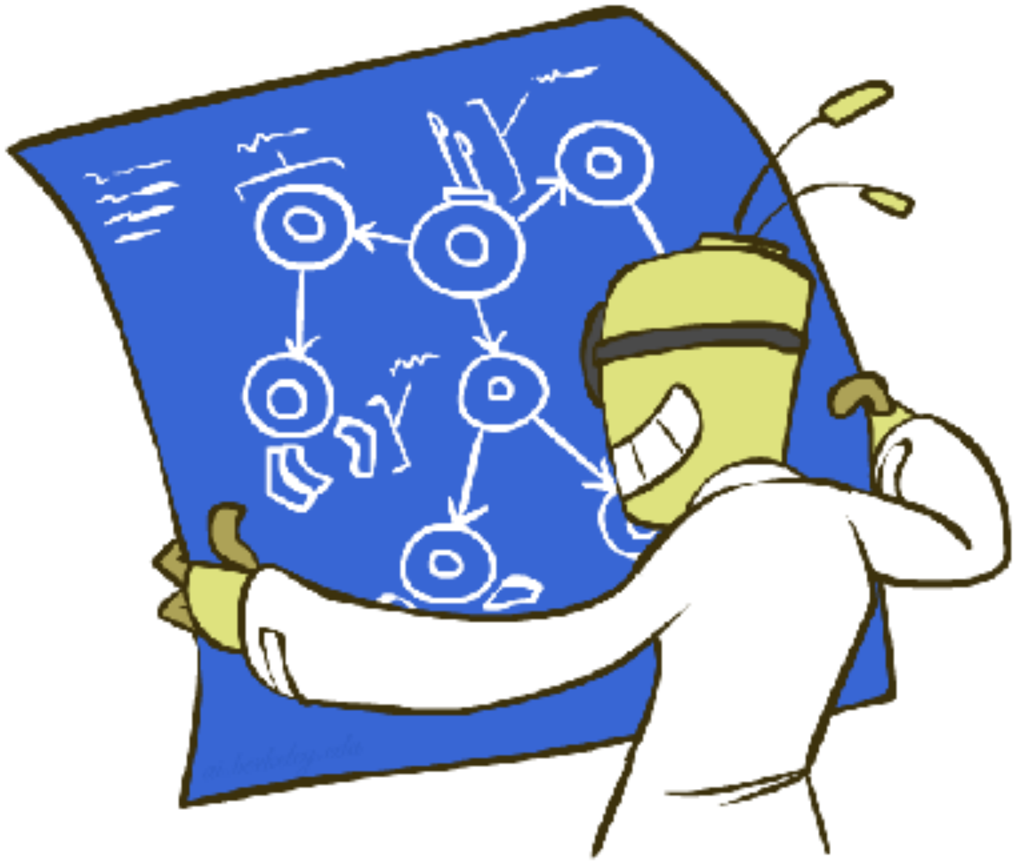
$T \perp\!\!\!\perp D | R, S$ No

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

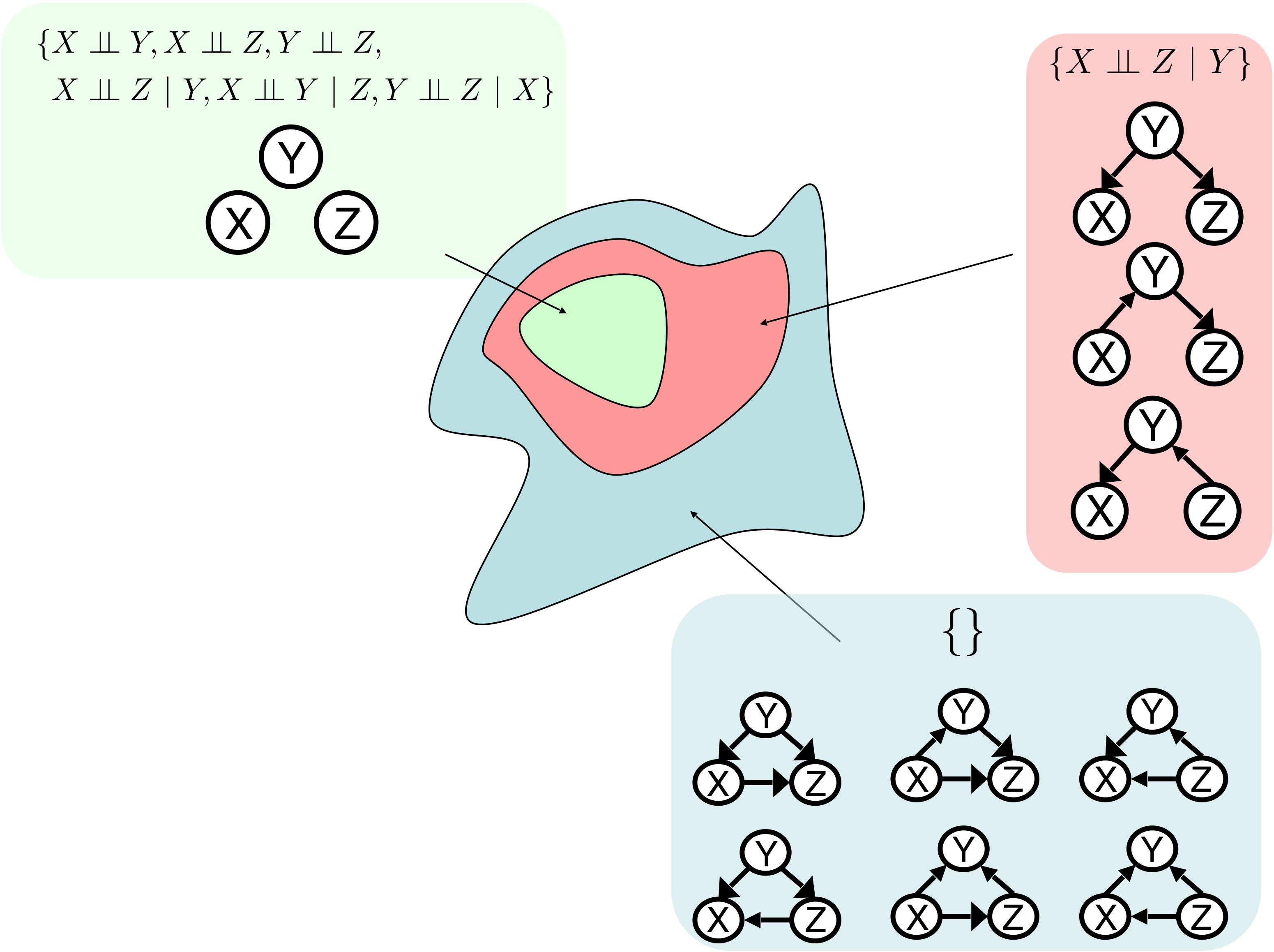
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

- 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
  - Sampling (approximate)
- Learning Bayes' Nets from Data