CSEP 573: Artificial Intelligence

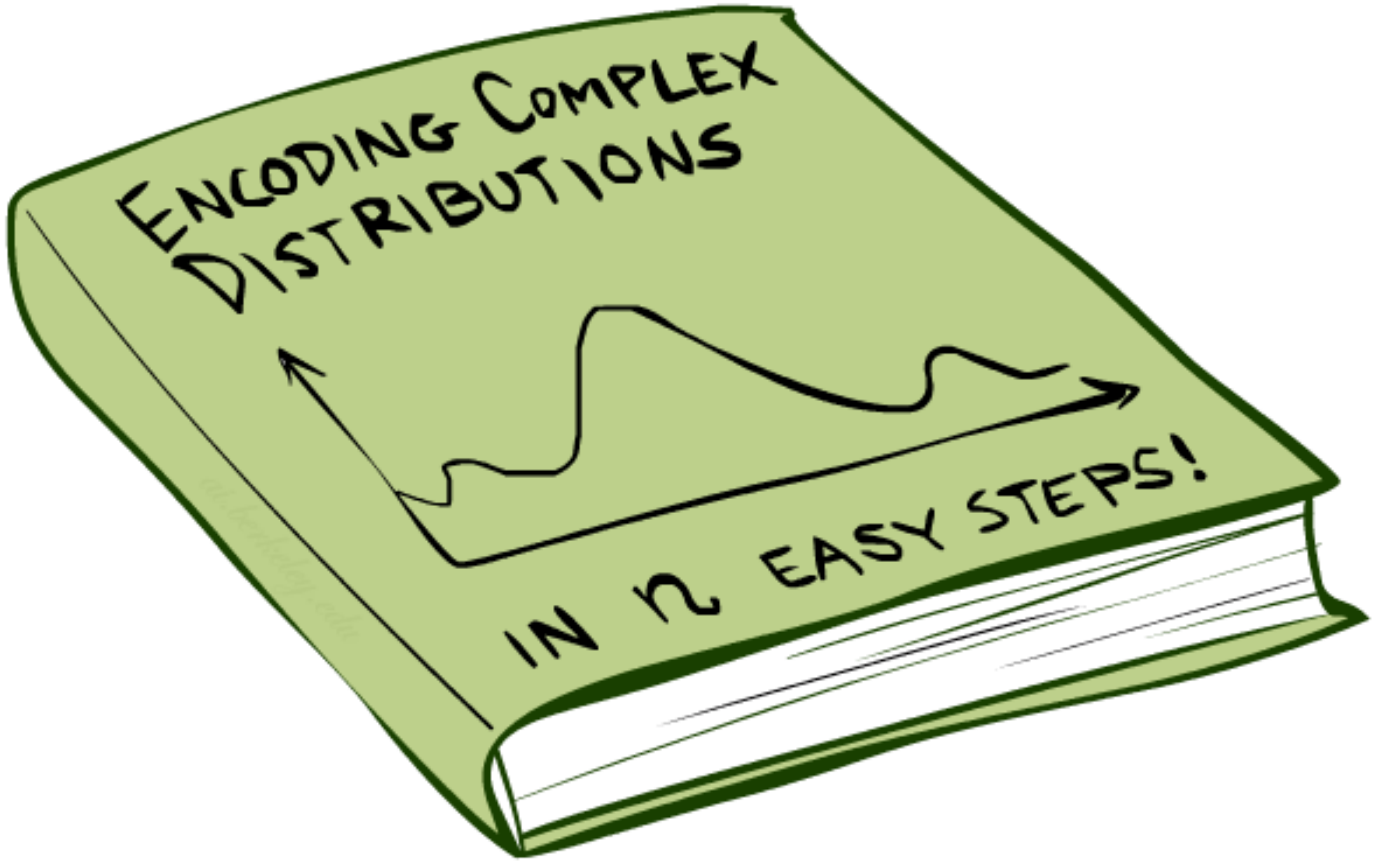
Graphical Models

slides adapted from
Stuart Russel, Dan Klein, Pieter Abbeel from ai.berkeley.edu
And Hanna Hajishirzi, Dan Weld, Luke Zettlemoyer

Reminder: elementary probability

- Basic laws: $0 \leq P(\omega) \leq 1 \qquad \sum_{\omega \in \Omega} P(\omega) = 1$
- Events: subsets of Ω: $\mathbf{P}(A) = \sum_{\omega \in A} P(\omega)$
- Random variable $X(\omega)$ has a value in each ω
  - Distribution $\mathbf{P}(X)$ gives probability for each possible value x
  - Joint distribution $\mathbf{P}(X,Y)$ gives total probability for each combination x,y
- Summing out/marginalization: $P(X=x) = \sum_y P(X=x,Y=y)$
- Conditional probability: $\mathbf{P}(X|Y) = \mathbf{P}(X,Y)/\mathbf{P}(Y)$
- Product rule: $\mathbf{P}(X|Y)\mathbf{P}(Y) = \mathbf{P}(X,Y) = \mathbf{P}(Y|X)\mathbf{P}(X)$
  - Generalize to chain rule: $\mathbf{P}(X_1,..,X_n) = \prod_i \mathbf{P}(X_i \mid X_1,..,X_{i-1})$

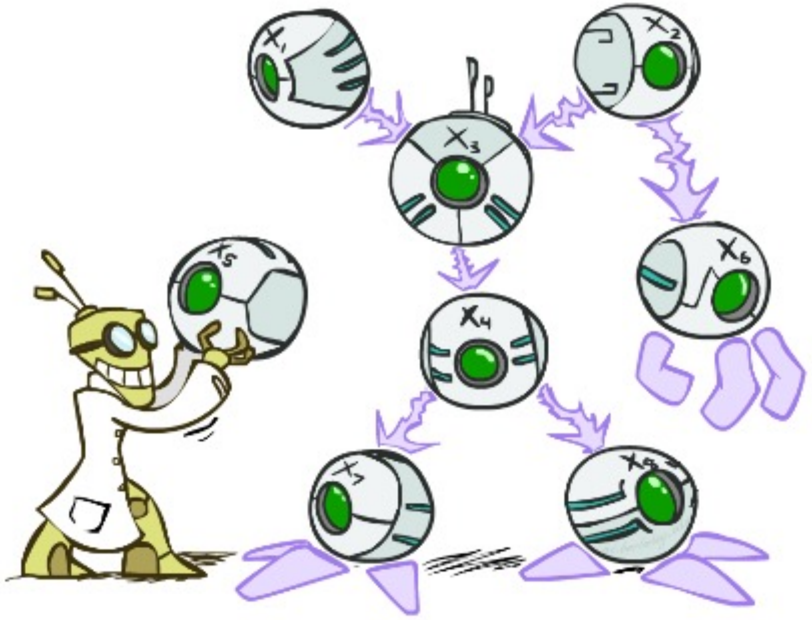
Bayes' Nets: Big Picture



Bayes Nets: Big Picture

- Bayes nets: a technique for describing complex joint distributions (models) using simple, conditional distributions
  - A subset of the general class of graphical models
  - Also called belief networks

- Use local causality/conditional independence:
  - the world is composed of many variables,
  - each interacting locally with a few others



Bayes Nets

Part I: Representation

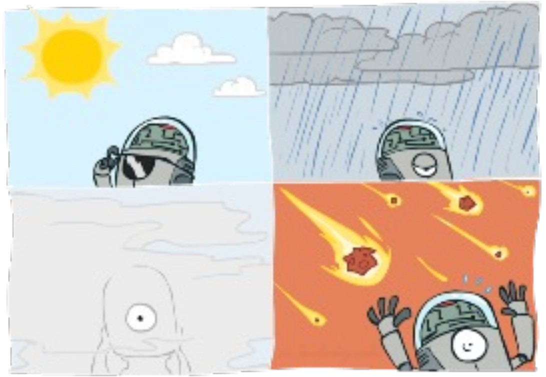
Part II: Independence

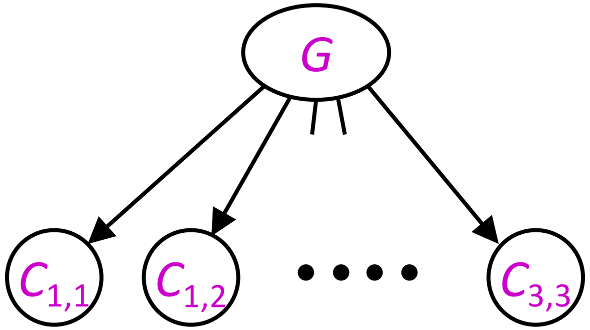
Part III: Exact inference

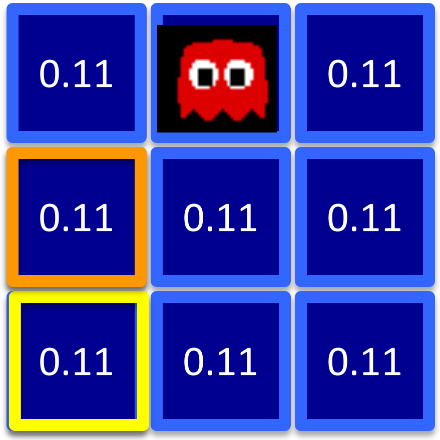
Part IV: Approximate Inference

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 (more on this later)

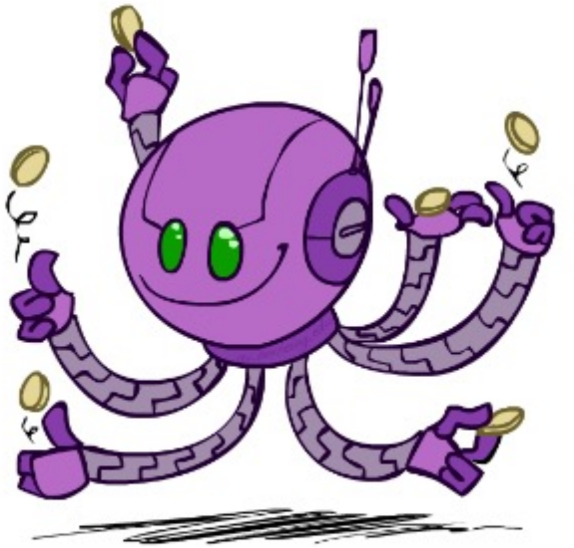






Example Bayes' Net: Coin Flips

- N independent coin flips



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

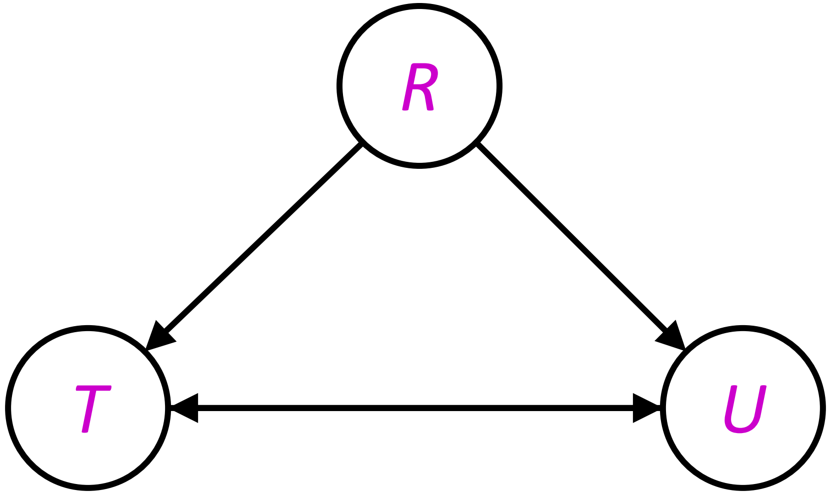
- No interactions between variables: absolute independence

Conditional Independence: Traffic

- What about this domain:
  - Traffic
  - Umbrella
  - Raining

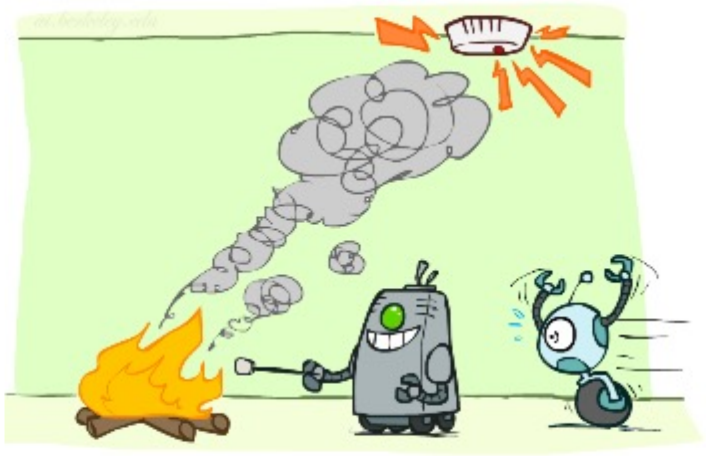
Example Bayes' Net: Traffic

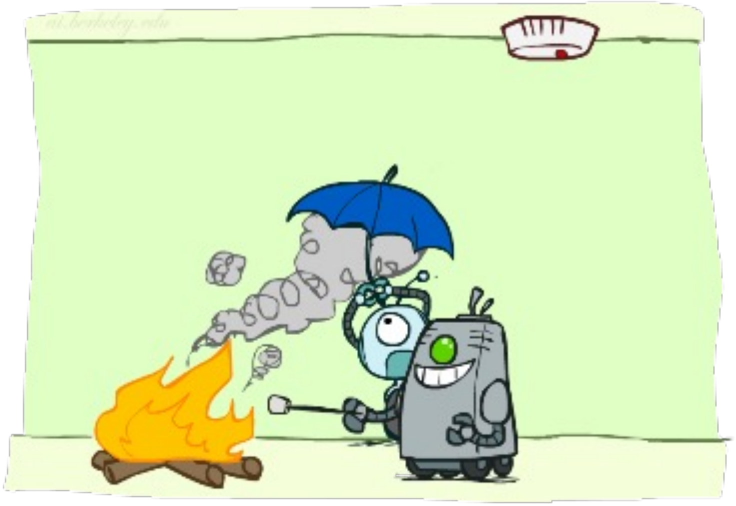
- Variables:
  - T: There is traffic
  - U: I'm holding my umbrella
  - R: It rains



Conditional Independence: Fire

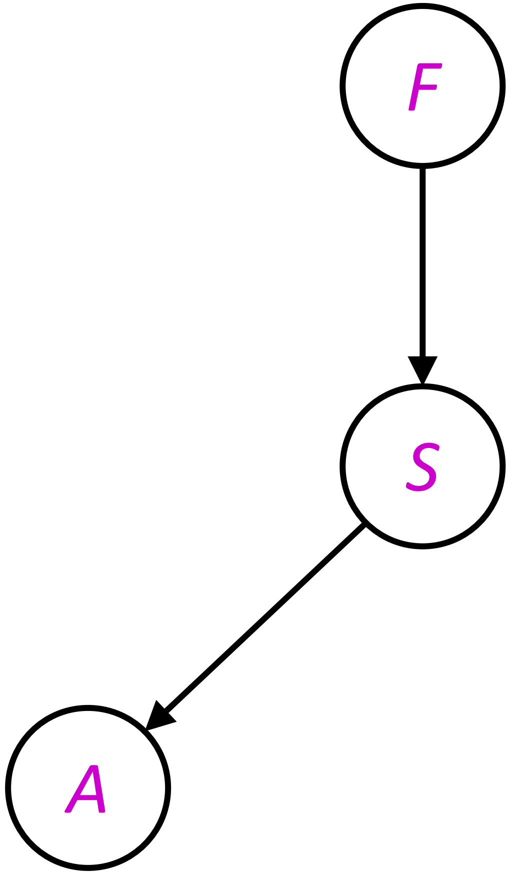
- What about this domain:
  - Fire
  - Smoke
  - Alarm

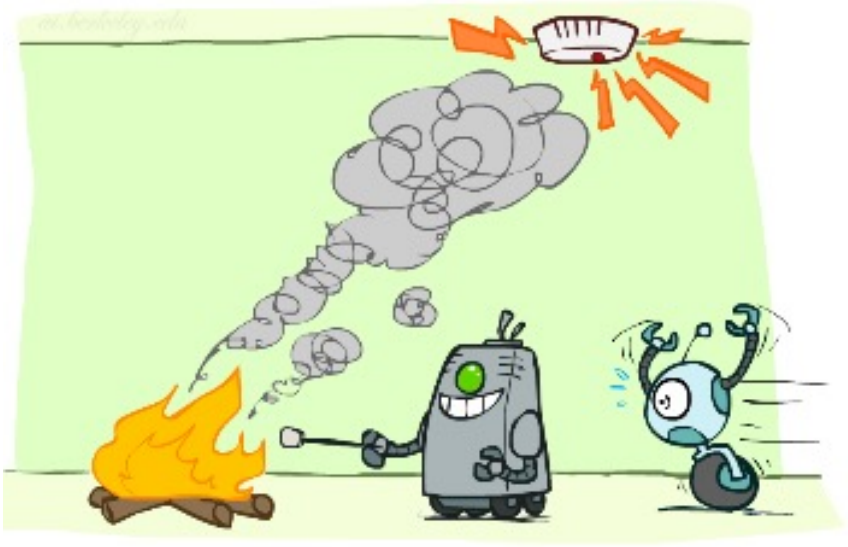




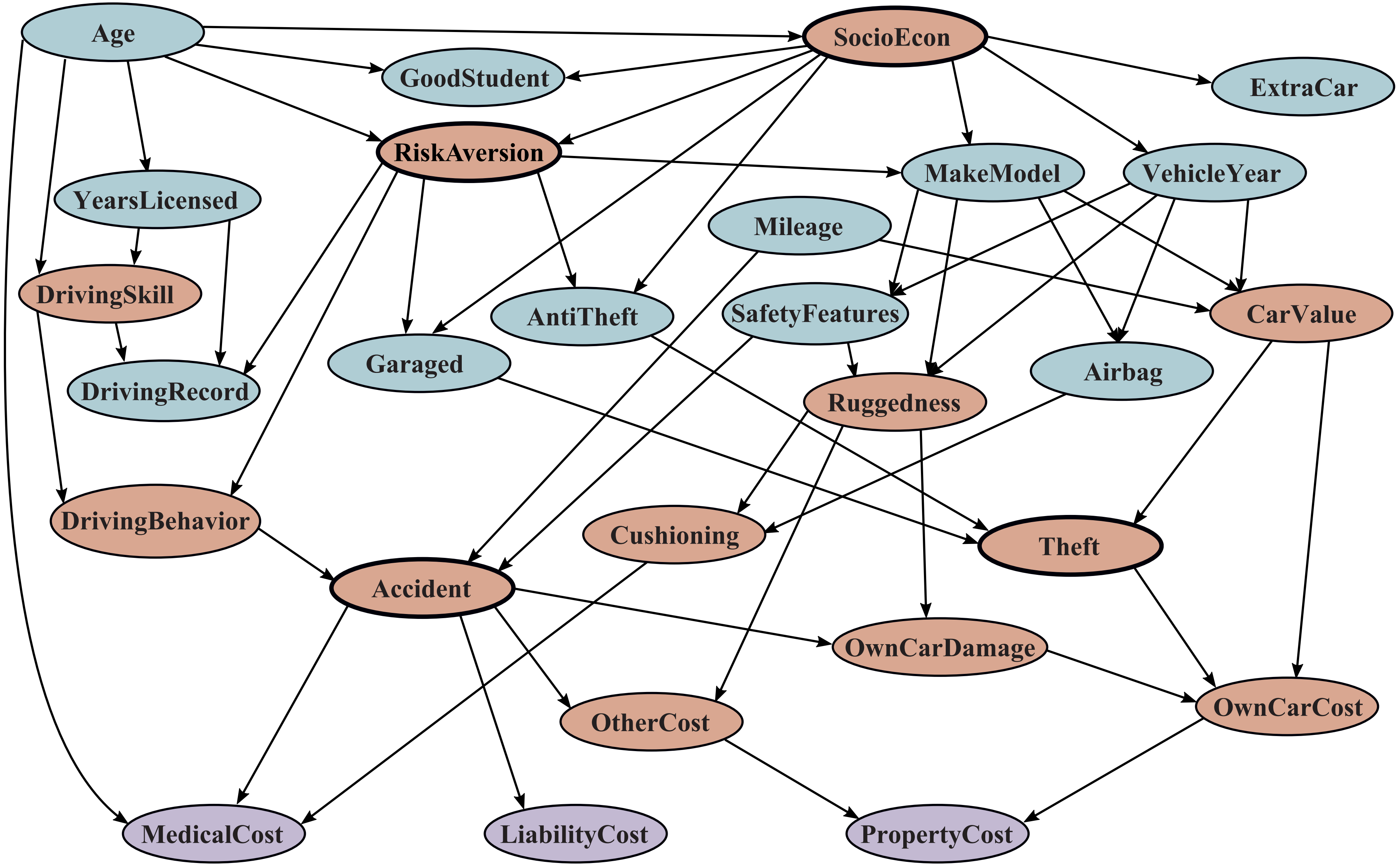
Example Bayes' Net: Smoke alarm

- Variables:
  - F: There is fire
  - S: There is smoke
  - A: Alarm sounds



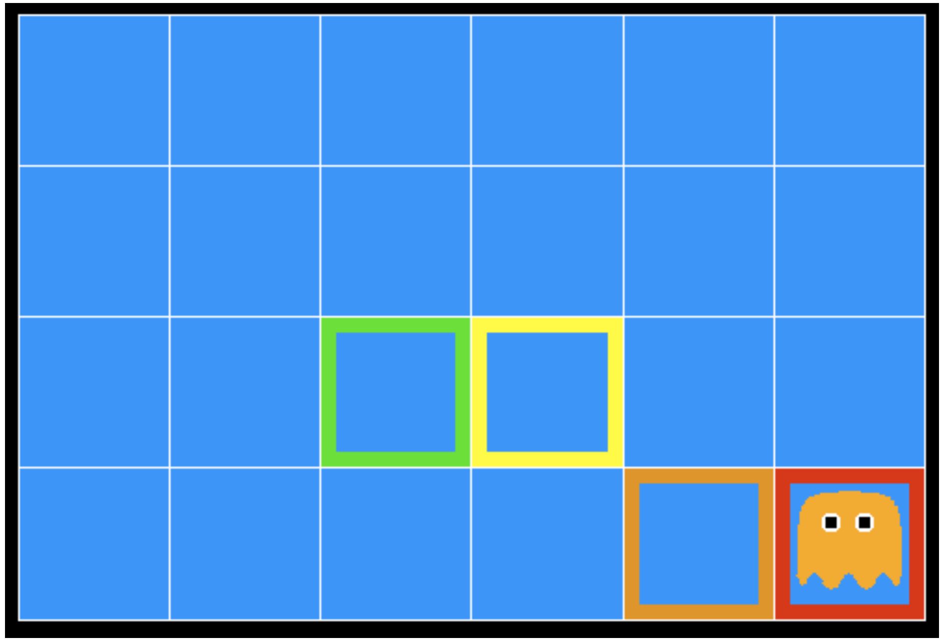


Example Bayes' Net: Car Insurance



Why do conditional independence?-- Ghostbusters

- A ghost is in the grid somewhere
- Sensor readings tell how close a square is to the ghost
  - On the ghost: usually red
  - 1 or 2 away: mostly orange
  - 3 or 4 away: typically yellow
  - 5+ away: often green
- Click on squares until confident of location, then "***bust***"



Video of Demo Ghostbusters with Probability

P(ghost is in this position given all of the evidence that we have seen so far)

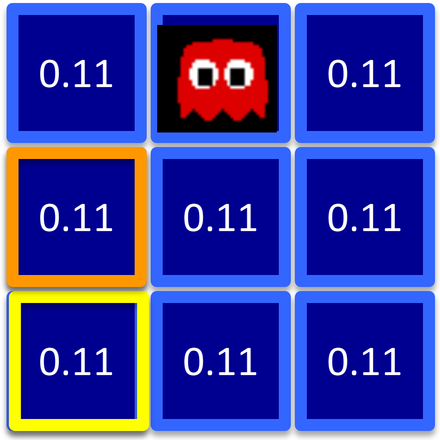
Ghostbusters model

- Variables and ranges:
  - G (ghost location) in $\{(1,1),\ldots,(3,3)\}$
  - $C_{x,y}$ (color measured at square x,y) in {red,orange,yellow,green}

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

- Ghostbuster physics:
  - ***Uniform prior distribution*** over ghost location: $\boldsymbol{P}(G)$
  - ***Sensor model***: $\boldsymbol{P}(C_{x,y} \mid G)$ (depends only on distance to G)
    - E.g. $P(C_{1,1} = \text{yellow} \mid G = (1,1)) = 0.1$

Ghostbusters model, contd.



- **P**(G, $C_{1,1}$, … $C_{3,3}$) has …
  - $9 \times 4^9 = 2{,}359{,}296$ entries!
  - $|G| = 9$, $|C_{i,j}| = 4$; Grid squares times size of each
- Ghostbuster independence:
  - Are $C_{1,1}$ and $C_{1,2}$ independent?
    - E.g., does $P(C_{1,1} = \text{yellow}) = P(C_{1,1} = \text{yellow} \mid C_{1,2} = \text{orange})$?
- Ghostbuster physics again:
  - ***P***$(C_{x,y} \mid G)$ ***depends <u>only</u> on distance to G***
    - So $P(C_{1,1} = \text{yellow} \mid \underline{G = (2,3)}\,) = P(C_{1,1} = \text{yellow} \mid \underline{G = (2,3)}, C_{1,2} = \text{orange})$
    - I.e., $C_{1,1}$ is ***conditionally independent*** of $C_{1,2}$ ***given*** G

Ghostbusters model, contd.

- Apply the chain rule to decompose the joint probability model:

  $\mathbf{P}(G, C_{1,1}, \ldots C_{3,3}) = \mathbf{P}(G)\, \mathbf{P}(C_{1,1} \mid G)\, \mathbf{P}(C_{1,2} \mid G, \cancel{C_{1,1}})\, \mathbf{P}(C_{1,3} \mid G, \cancel{C_{1,1}, C_{1,2}}) \ldots \mathbf{P}(C_{3,3} \mid G, \cancel{C_{1,1}, \ldots, C_{3,2}})$

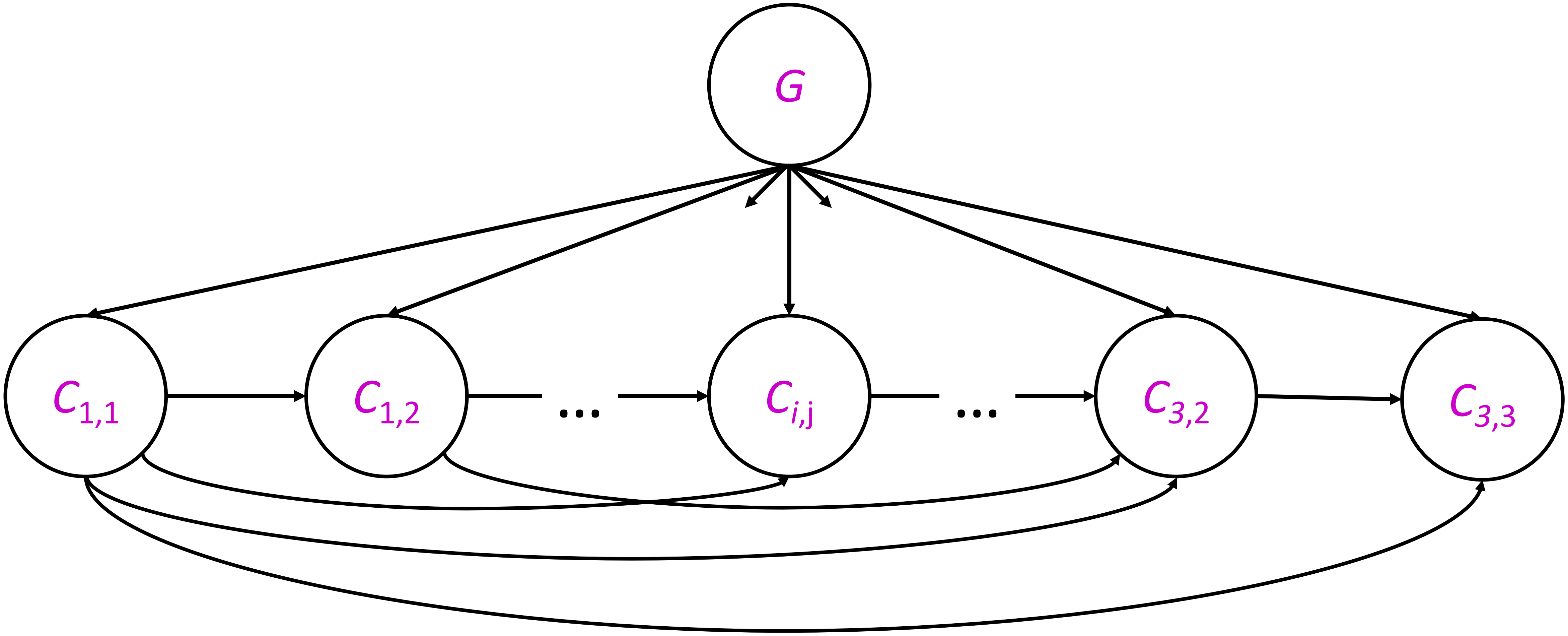
- Now simplify using conditional independence:

  $\mathbf{P}(G, C_{1,1}, \ldots C_{3,3}) = \mathbf{P}(G)\, \mathbf{P}(C_{1,1} \mid G)\, \mathbf{P}(C_{1,2} \mid G)\, \mathbf{P}(C_{1,3} \mid G) \ldots \mathbf{P}(C_{3,3} \mid G)$

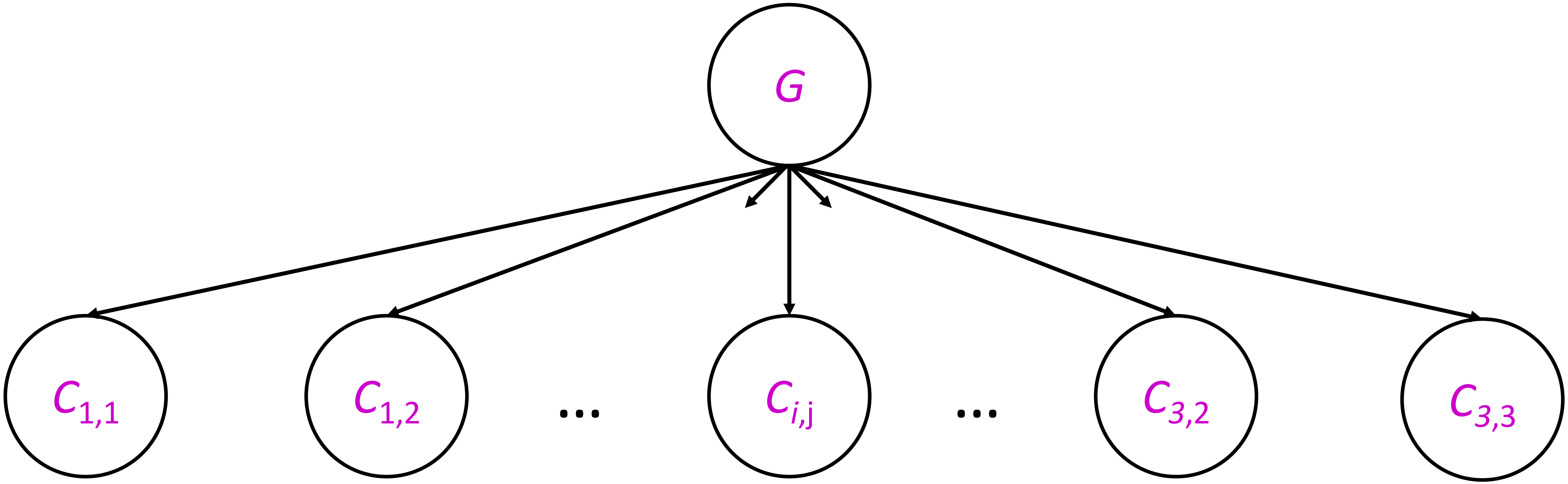
- I.e., conditional independence properties of ghostbuster physics simplify the probability model from ***exponential*** to ***quadratic*** in the number of squares
  - $|\mathbf{P}(C_{i,i} \mid G)| = 4 \times 9$ rather than $|\mathbf{P}(C_{3,3} \mid G, C_{1,1}, \ldots, C_{3,2})| = 4 \times 9 \times 4^8$
  - In total: $9 + 9 \times (4 \times 9) = 333$ entries, before was $9 \times 4^9 = 2{,}359{,}296$ entries

- This is called a ***Naïve Bayes*** model:
  - One discrete query variable (often called the ***class*** or ***category*** variable)
  - All other variables are (potentially) evidence variables
  - Evidence variables are all conditionally independent given the query variable

Ghostbusters Full Joint



Ghostbusters Naïve Bayes

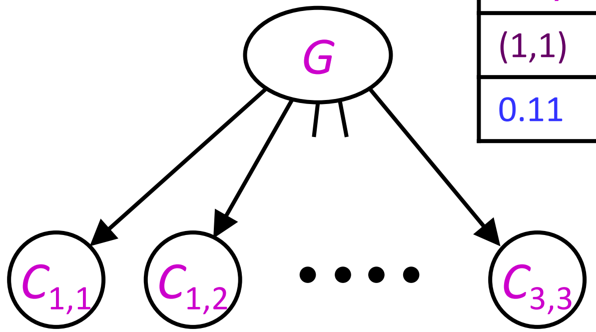


Bayes Net Syntax and Semantics

Bayes' Net Syntax

- A set of nodes, one per variable X_i
- A directed, acyclic graph
- A conditional distribution for each node given its ***parent variables*** in the graph
  - ***CPT*** (conditional probability table)

    each row is a distribution for child given values of its parents



P(G)			
(1,1)	(1,2)	(1,3)	…
0.11	0.11	0.11	…

G	**P**($C_{1,1}$ | ***G***)			
	g	y	o	r
(1,1)	0.01	0.1	0.3	0.59
(1,2)	0.1	0.3	0.5	0.1
(1,3)	0.3	0.5	0.19	0.01
…				

Bayes net = Topology (graph) + Local Conditional Probabilities

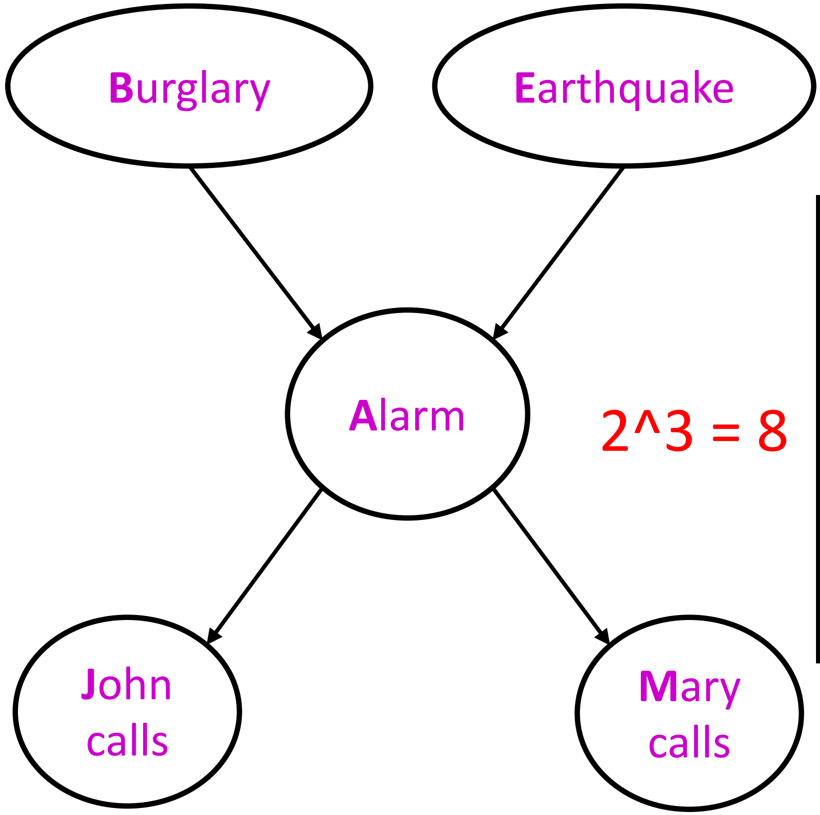
Example: Alarm Network

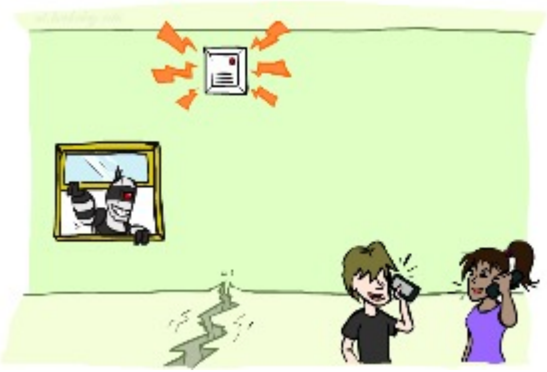
P(B)	
true	false
0.001	0.999

2

2

P(E)	
true	false
0.002	0.998





B	E	P(A|B,E)	
		true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

2^3 = 8

A	P(J|A)	
	true	false
true	0.9	0.1
false	0.05	0.95

2^2 = 4

2^2 = 4

A	P(M|A)	
	true	false
true	0.7	0.3
false	0.01	0.99

Factor size of each CPT:

$d \prod d_i$

Parent range sizes: $d_1,\ldots,d_k$

Child range size: d

Each table row must sum to 1

General formula for sparse BNs

- Suppose
  - n variables
  - Maximum range size is d
  - Maximum number of parents is k
- Full joint distribution has size $O(d^n)$
- Bayes net has size $O(n \cdot d^k)$
  - Linear scaling with n as long as causal structure is local
- *Often* $O(n \cdot d^k) << O(d^n)$

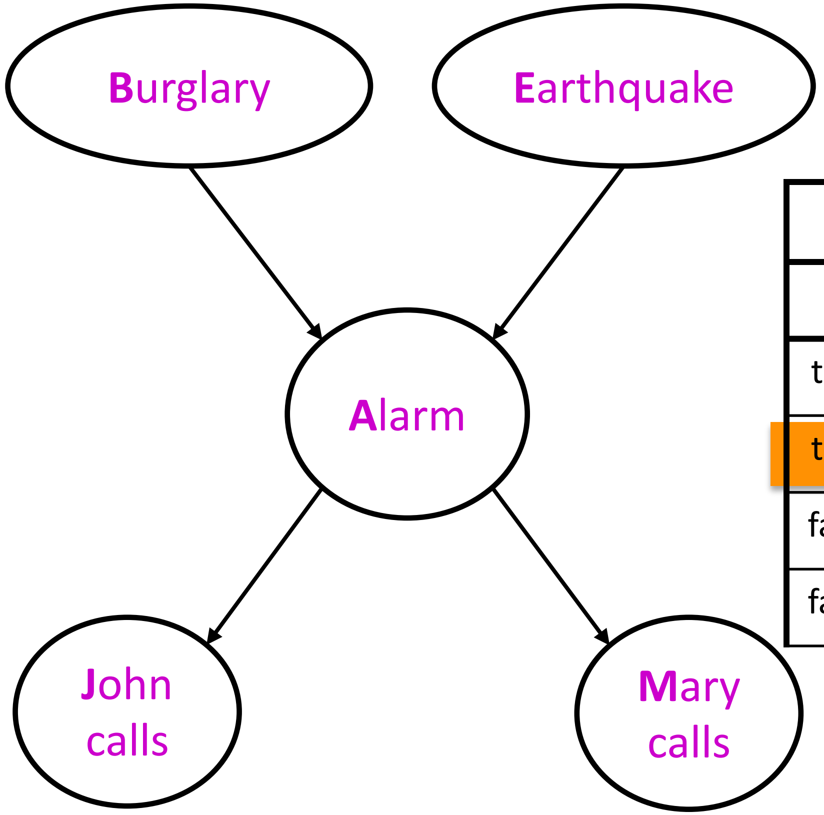
Bayes net global semantics

- Bayes nets encode joint distributions as product of conditional distributions on each variable:

$$\mathbf{P}(X_1,..,X_n) = \prod_i \mathbf{P}(X_i \mid Parents(X_i))$$

Example

P(B)	
true	false
0.001	0.999



P(E)	
true	false
0.002	0.998

B	E	P(A|B,E)	
		true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

A	P(J|A)	
	true	false
true	0.9	0.1
false	0.05	0.95

A	P(M|A)	
	true	false
true	0.7	0.3
false	0.01	0.99

$P(b, \neg e, a, \neg j, \neg m) =$

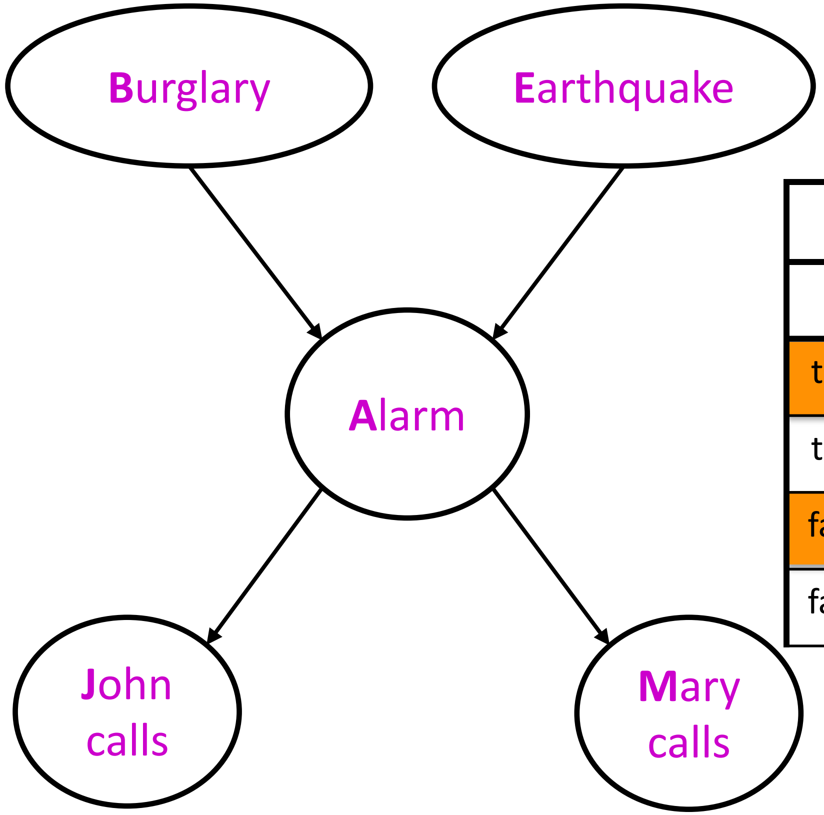
$P(b)\ P(\neg e)\ P(a|b,\neg e)\ P(\neg j|a)\ P(\neg m|a)$

$= .001 \times .998 \times .94 \times .1 \times .3$

$= .000028$

Example: Your turn

P(B)	
true	false
0.001	0.999



P(E)	
true	false
0.002	0.998

B	E	**P**(A|B,E)	
		true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

A	**P**(J|A)	
	true	false
true	0.9	0.1
false	0.05	0.95

A	**P**(M|A)	
	true	false
true	0.7	0.3
false	0.01	0.99

P(B, e, a, j, m) =

P(B) P(e) P(a|B,e) P(j|a) P(m|a)

$= \langle .001, .999 \rangle \times .002 \times \langle .95, .29 \rangle \times .9 \times .7$

$= \langle .001 * .002 * .95 * .9 * .7,$

$.999 * .002 * .29 * .9 * .7 \rangle$

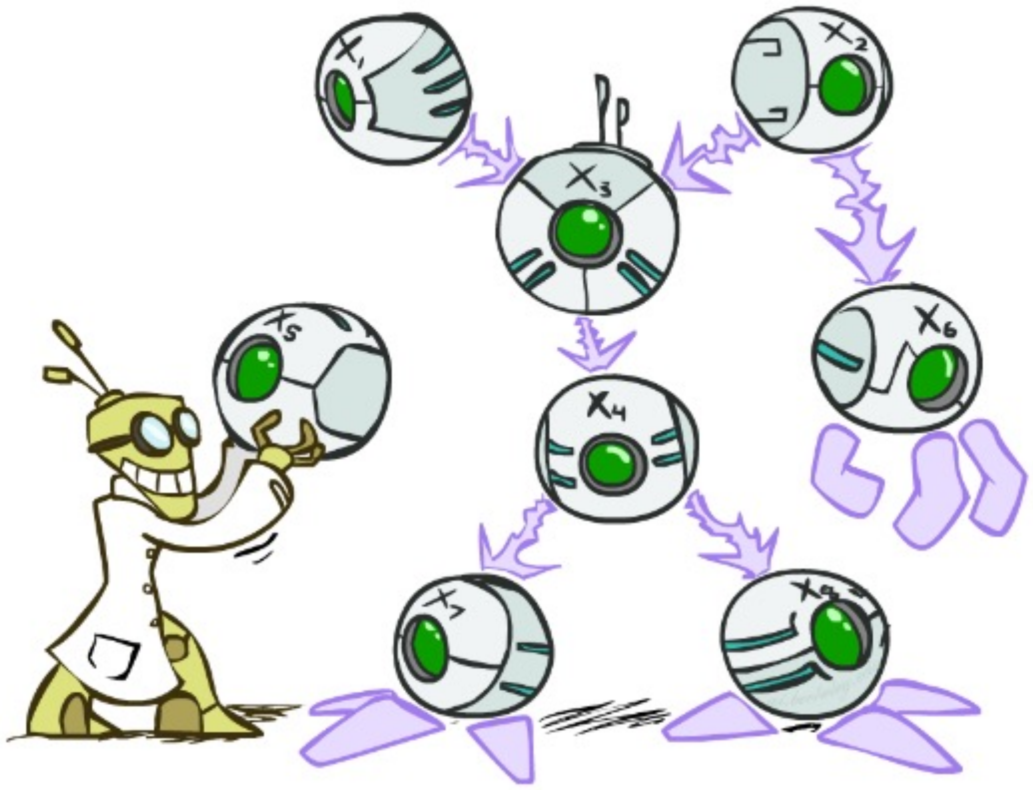
$= \langle .00000120, .000365 \rangle$

Question

- Which of the following does a Bayes' net model explicitly?
  - The joint probability distribution?
  - The conditional probability distribution?
- Is one of the following more expressive than the other?
  - The joint probability distribution
  - The conditional probability distribution
- Why do we use Bayes' nets?

Summary

- Independence and conditional independence are important forms of probabilistic knowledge
- Bayes net encode joint distributions efficiently by taking advantage of conditional independence
  - Global joint probability = product of local conditionals
- Next: more on independence
- Then: how to answer queries, i.e., compute conditional probabilities of queries given evidence



Bayes Nets

✓ Part I: Representation

Part II: Independence

Part III: Exact inference

- Enumeration (always exponential complexity)
- Variable elimination (worst-case exponential complexity, often better)
- Inference is NP-hard in general

Part IV: Approximate Inference

Conditional independence in BNs

- Compare the Bayes net global semantics

$$\mathbf{P}(X_1,..,X_n) = \prod_i \mathbf{P}(X_i \mid Parents(X_i))$$

  with the chain rule identity

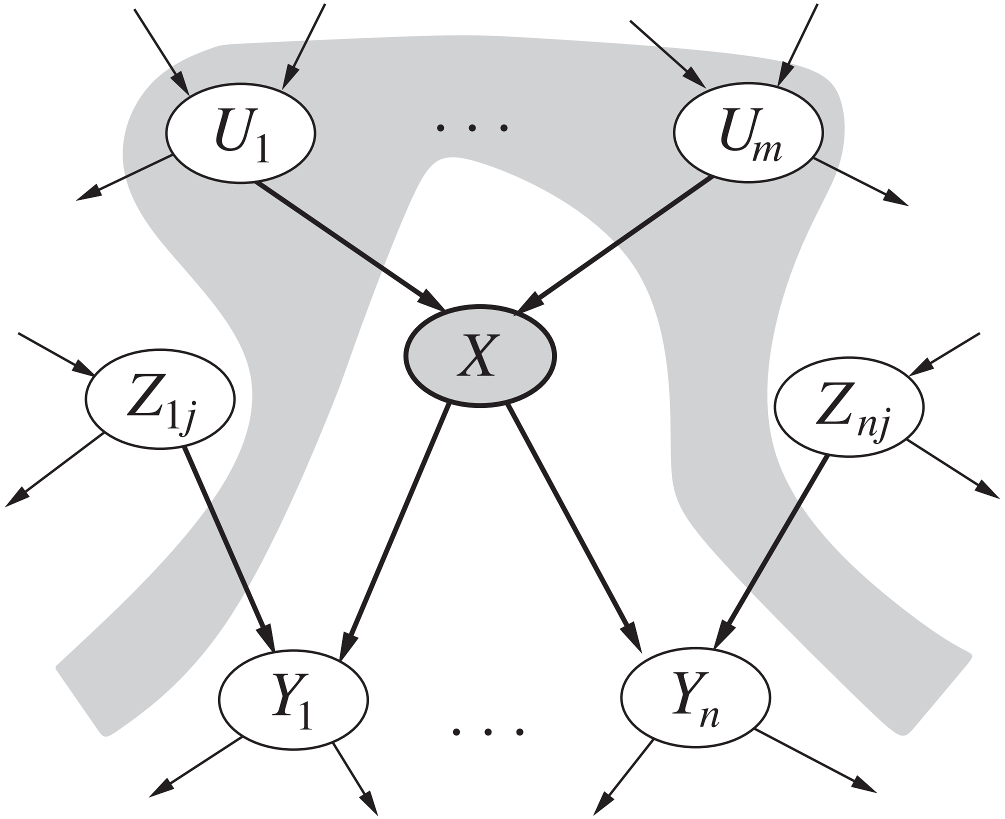
$$\mathbf{P}(X_1,..,X_n) = \prod_i \mathbf{P}(X_i \mid X_1,\ldots,X_{i-1})$$

- Assume (without loss of generality) that $X_1,..,X_n$ sorted in topological order according to the graph (i.e., parents before children), so $Parents(X_i) \subseteq X_1,\ldots,X_{i-1}$

- So the Bayes net asserts conditional independences $\mathbf{P}(X_i \mid X_1,\ldots,X_{i-1}) = \mathbf{P}(X_i \mid Parents(X_i))$
  - To ensure these are valid, choose parents for node X_i that "shield" it from other predecessors

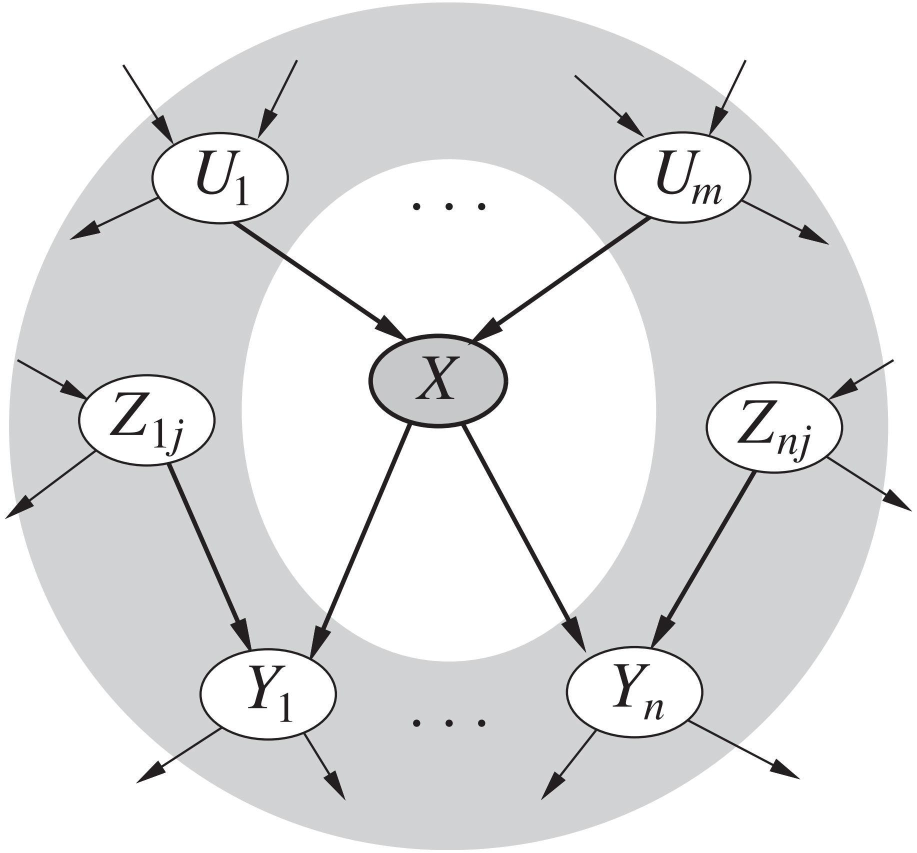
Conditional independence semantics

- ***Every variable is conditionally independent of its non-descendants given its parents***
- Conditional independence semantics <=> global semantics



Markov blanket

- A variable's Markov blanket consists of parents, children, children's other parents
- ***Every variable is conditionally independent of all other variables given its Markov blanket***



Reminder: Conditional Independence

- X and Y are independent if

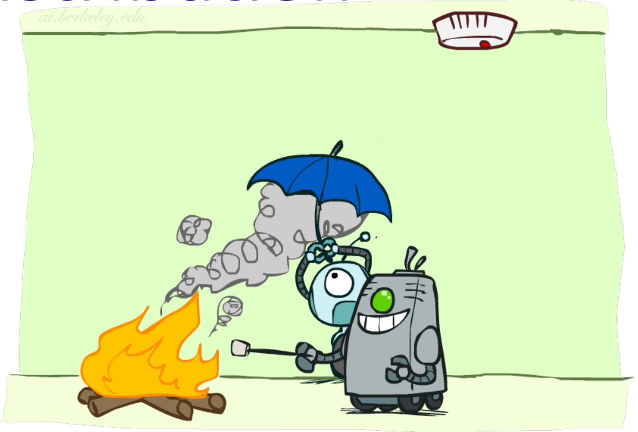
$$\forall x, y \;\; P(x, y) = P(x)P(y) \dashrightarrow X \perp\!\!\!\perp Y$$

- X and Y are conditionally independent given Z

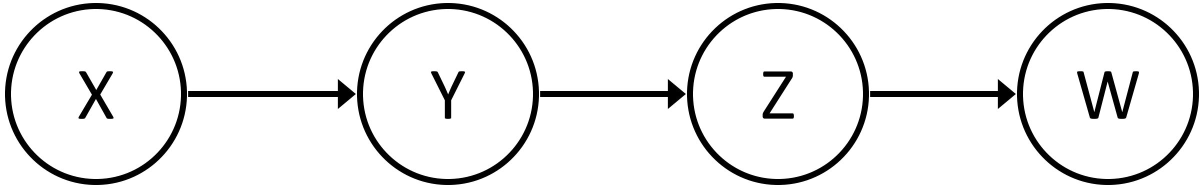
$$\forall x, y, z \;\; P(x, y|z) = P(x|z)P(y|z) \dashrightarrow X \perp\!\!\!\perp Y|Z$$

- (Conditional) independence is a property of a distribution

- Example: $Alarm \perp\!\!\!\perp Fire|Smoke$



Example



- Conditional independence assumptions directly from simplifications in chain rule:

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

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

- Additional implied conditional independence assumptions?

Example

$$\begin{tikzcd} X \arrow[r] & Y \arrow[r] & Z \arrow[r] & W \end{tikzcd}$$

- Conditional independence assumptions directly from simplifications in chain rule:

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

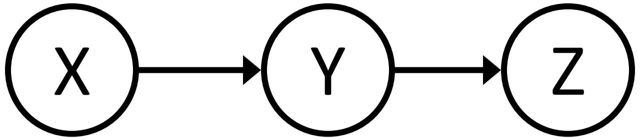
$$W \perp\!\!\!\perp \{X, Y\} | Z$$

- Additional implied conditional independence assumptions?

$$W \perp\!\!\!\perp X | Y$$

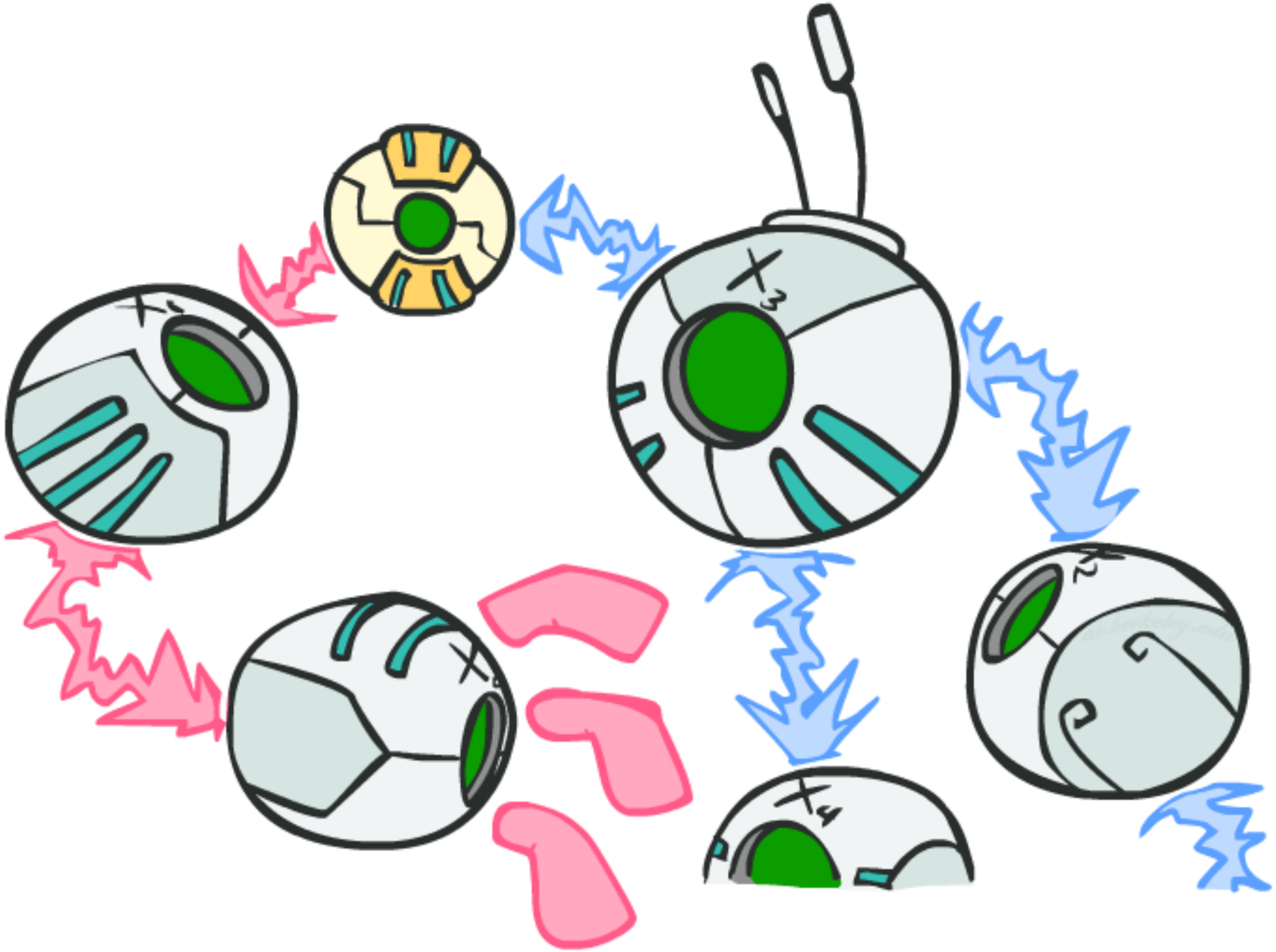
Independence in a Bayes' Net

- Important question about a Bayes' Net:
  - 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

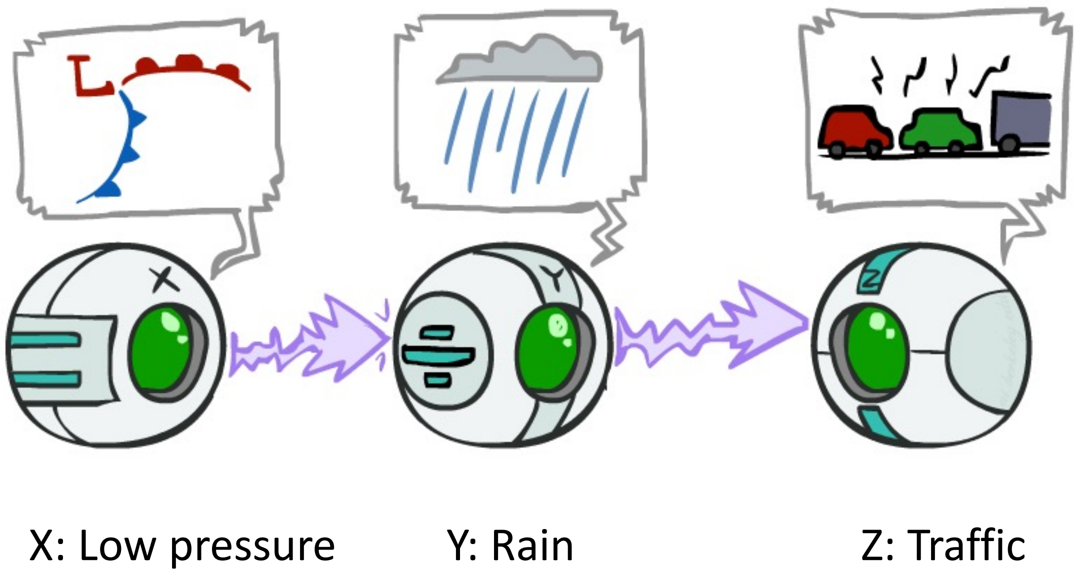


D-separation: Outline

- Study independence properties for triples
  - Why 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"



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

- 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:
    - 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

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

- Guaranteed X independent of Z given Y?

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

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

$$= P(z|y)$$

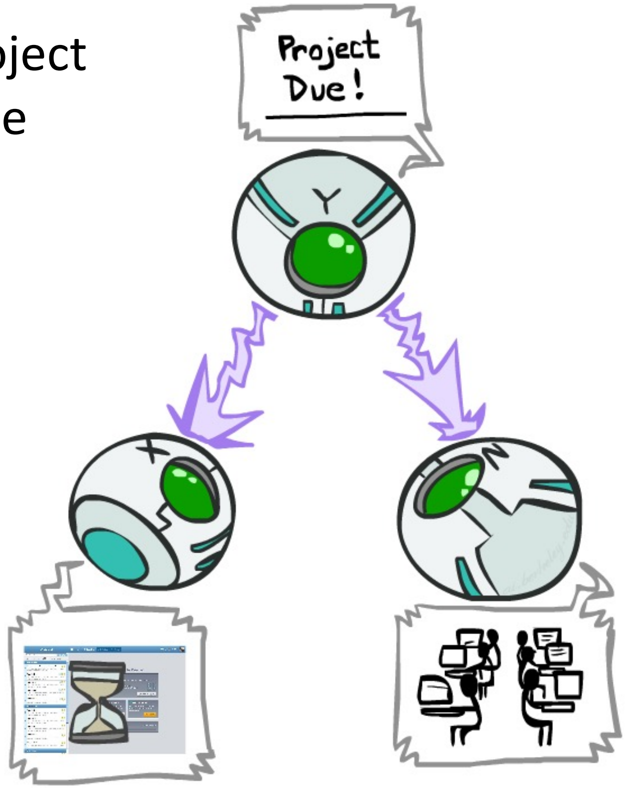
Yes!

- Evidence along the chain "blocks" the influence

Common Causes

- This configuration is a "common cause"

Y: Project due



X: Forums busy

Z: Lab full

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

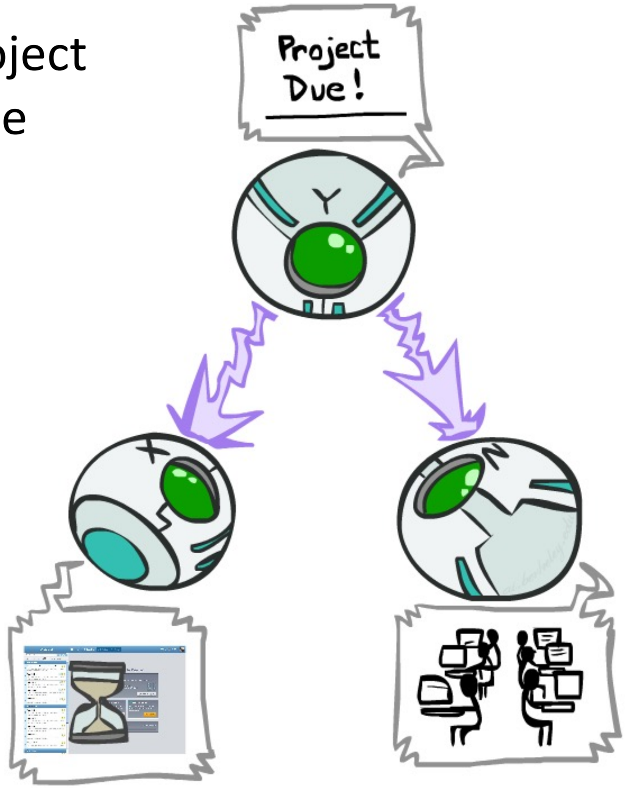
- Guaranteed X independent of Z ?
- *No!*
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
  - Example:
    - Project due causes both forums busy and lab full
    - In numbers:

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?

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

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

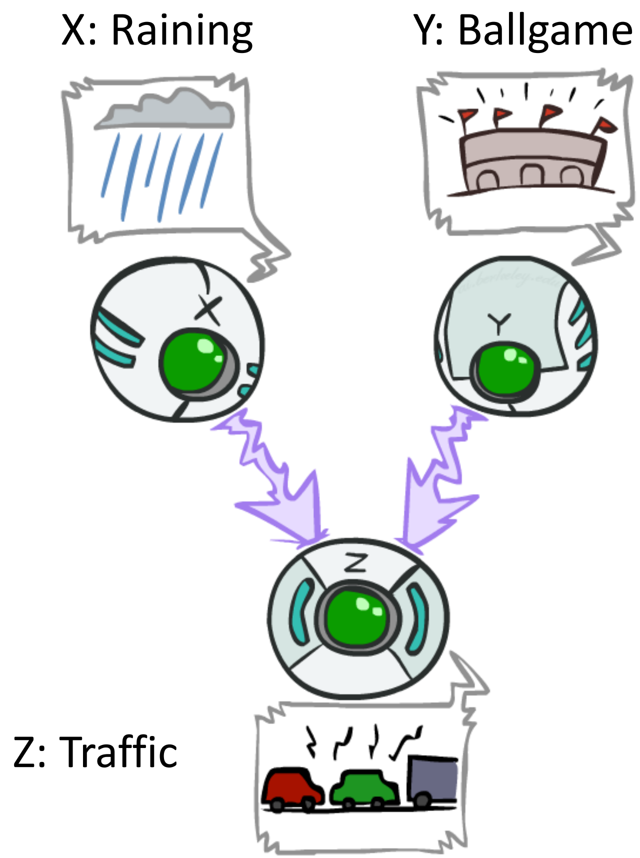
$$= P(z|y)$$

Yes!

- Observing the cause blocks influence between effects.

Common Effect

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

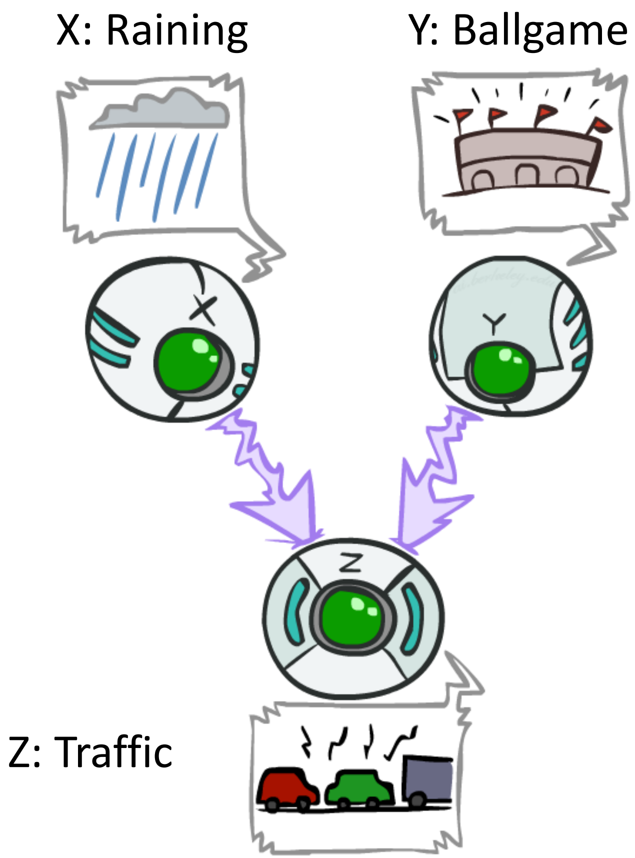


- Are X and Y independent?
  - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
- Proof:

$$P(x, y) = \sum P(x, y, z)$$

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
  - (Proved previously)
- 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.

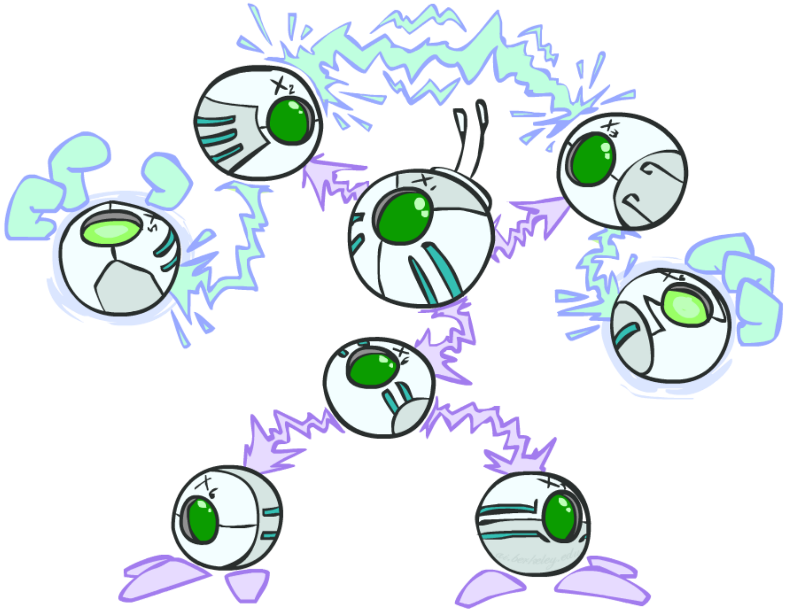
The General Case

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

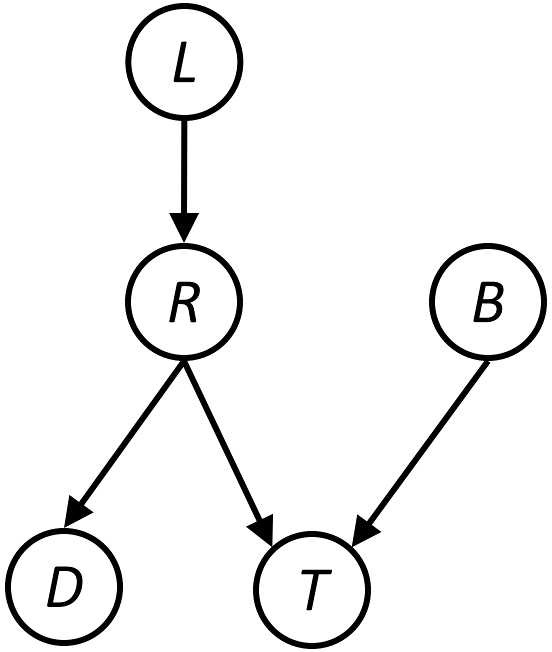


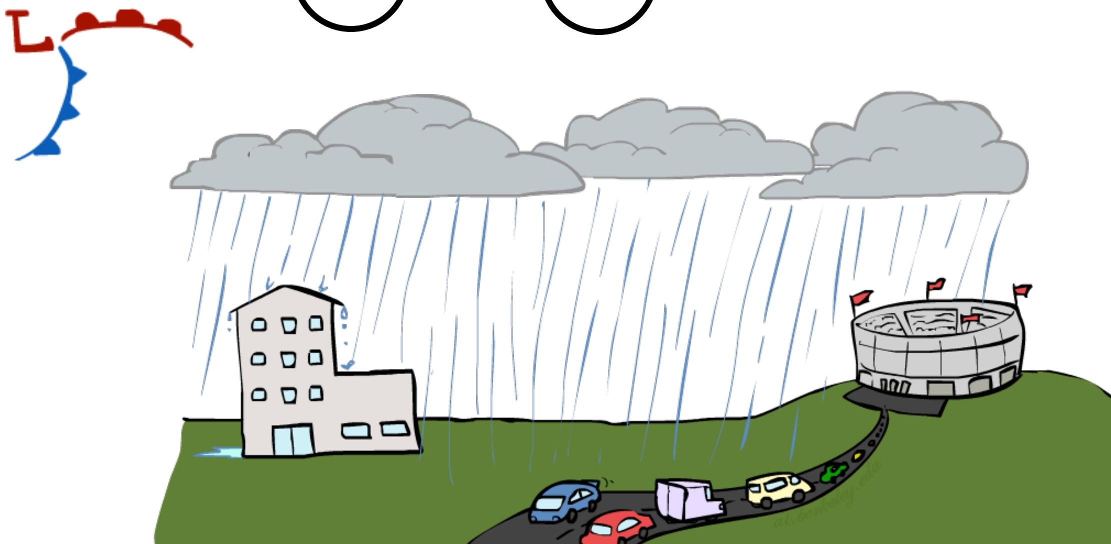
Reachability

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

- Attempt 1: if two nodes are **not** connected* they are conditionally independent
  - *There does not exist an undirected path between them, excluding those blocked by a shaded node.

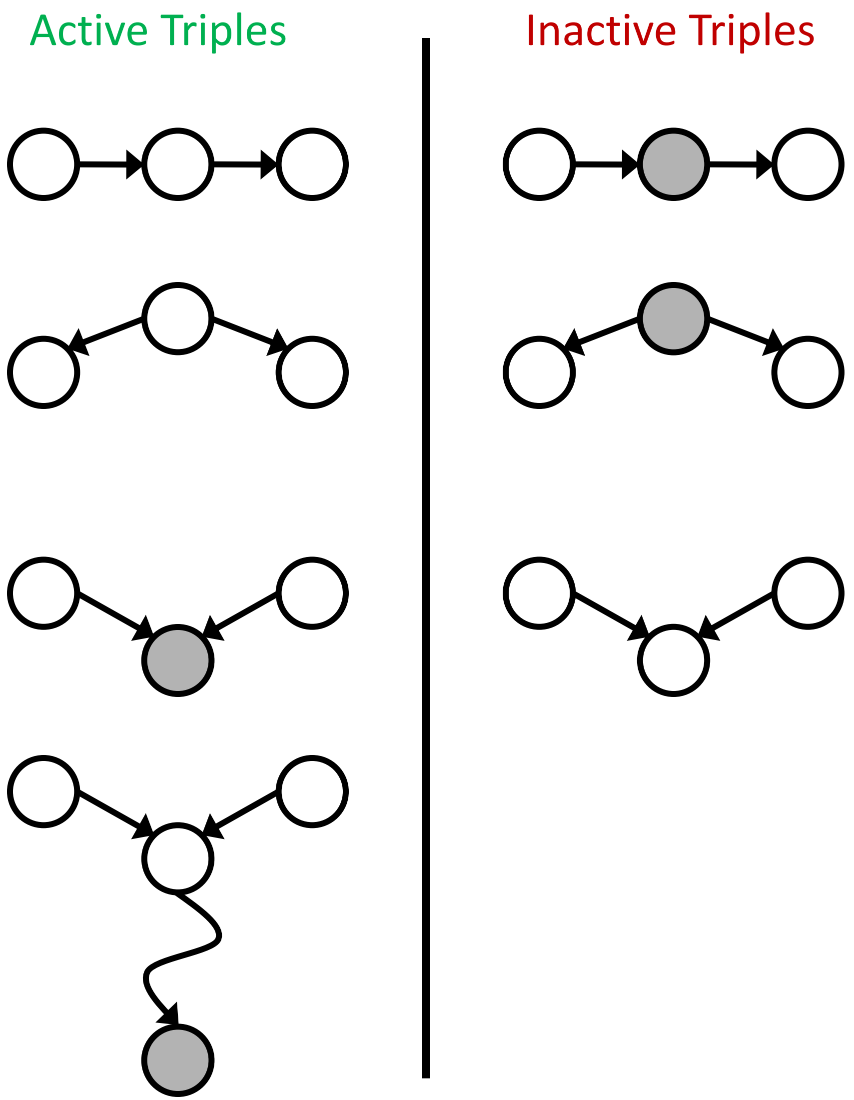
- Almost works, but not quite
  - Where does it break?
  - Answer: the v-structure at T doesn't count as a link in a path unless "active"





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 -> B -> C where B is unobserved (either direction)
  - **Common cause:** A <- B -> C where B is unobserved
  - **Common effect:** (aka v-structure)
    A -> B <- C where B *or one of its descendants* is observed
- All it takes to block a path is a single inactive segment



D-Separation

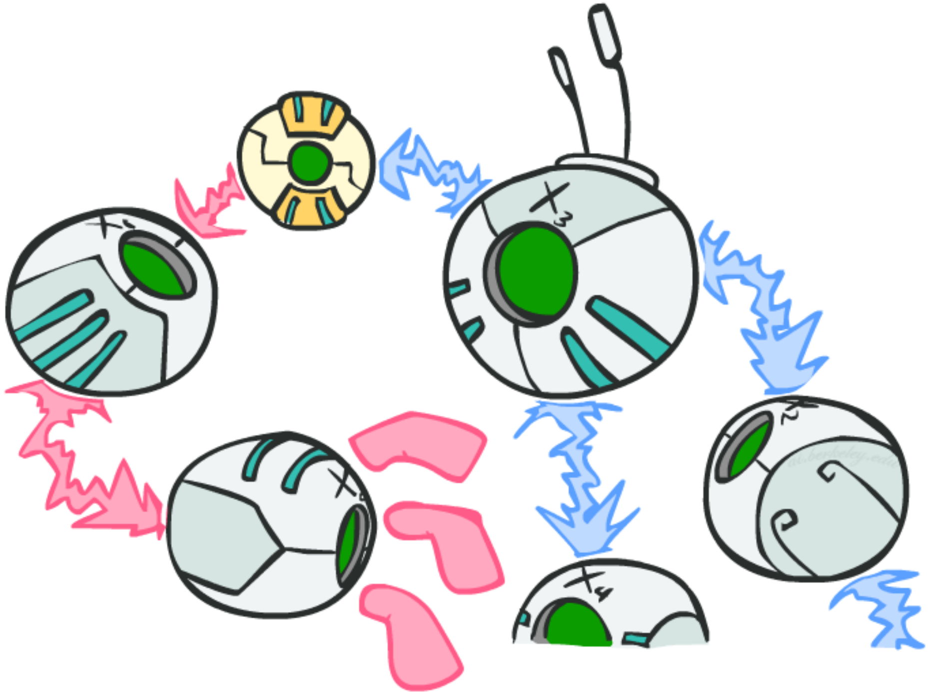
- Query: $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$?

- 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

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



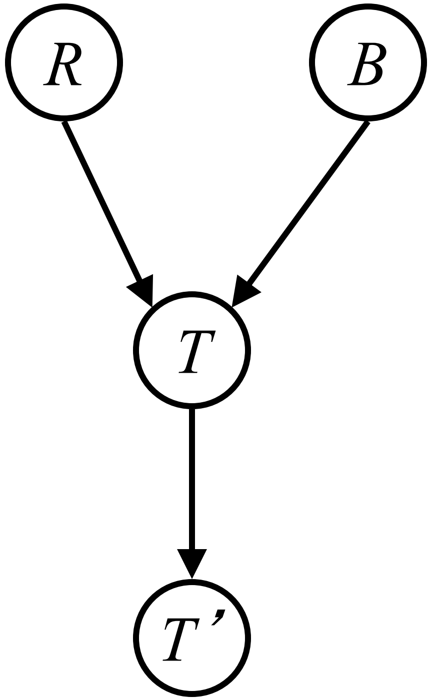
```
def d-separated(first, second):
    for path in paths(first, second):
        path_active = True
        for triple in path:
            if not active(triple):
                path_active = False
                break
        if path_active:
            return False
    return True
```

Example: which assumptions apply?

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

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

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



Example: which assumptions apply?

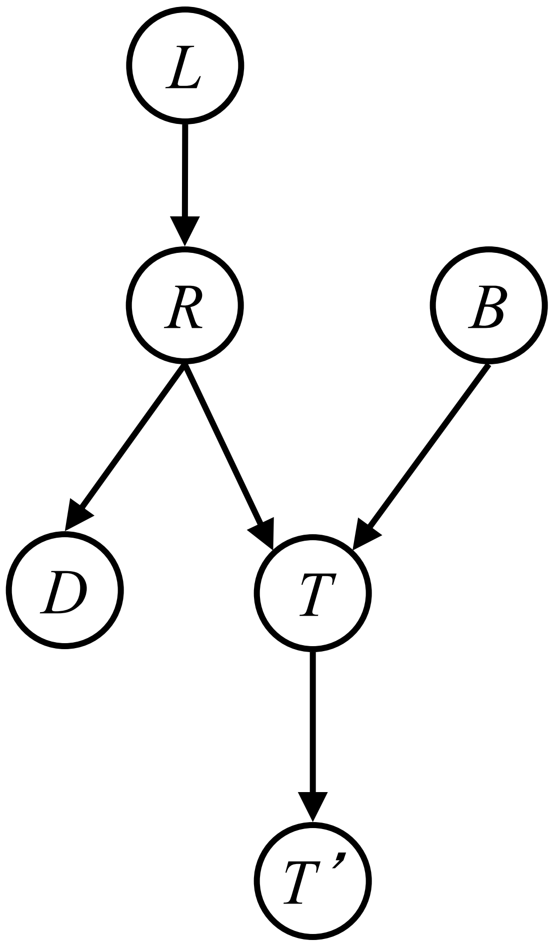
$L \perp\!\!\!\perp T' | T$ *Yes*

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

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

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

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



Example: which assumptions apply?

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

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

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

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