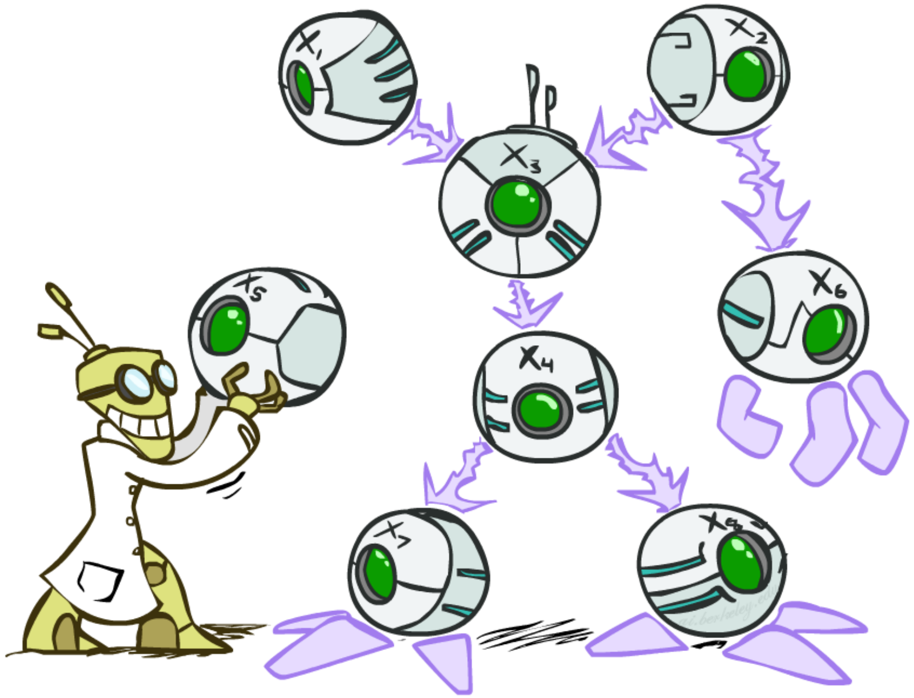
Bayesian Networks



These slides are based on the slides created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley - http://ai.berkeley.edu. The artwork is by Ketrina Yim.

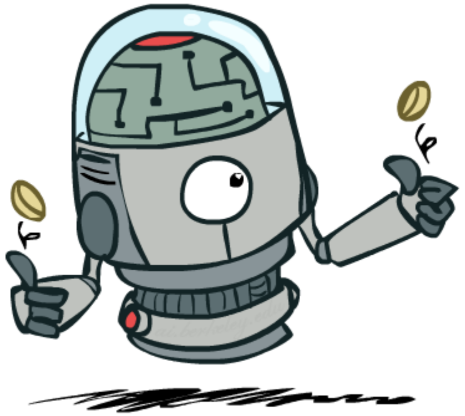
Today

Today:

- Independence between random variables
  - Absolute
  - Conditional
- Bayesian Networks: Representation
- D-Separation

Independence

- Two variables are *independent* in a joint distribution if:



$$P(X,Y) = P(X)P(Y)$$

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

  - the joint distribution *factors* into a product of two simple ones
  - Usually variables aren't independent!

- Can use independence as a *modeling assumption*
  - Independence can be a simplifying assumption
  - *Empirical* joint distributions: at best "close" to independent
  - We can assume that Traffic and Cavity are independent

Example: Independence?

$P(T)$

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

$P(W)$

W	P
sun	0.6
rain	0.4

$P_2(T, W) = P(T)P(W)$

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

No!

Example: Independence

- N fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

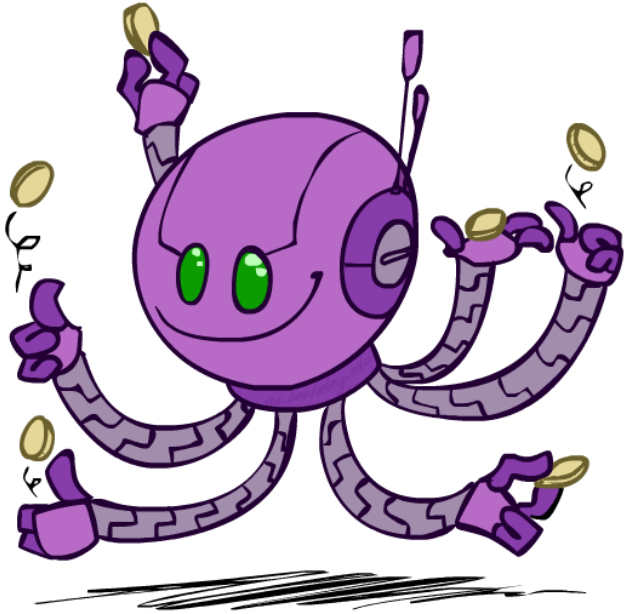
$P(X_2)$

H	0.5
T	0.5

$\cdots$

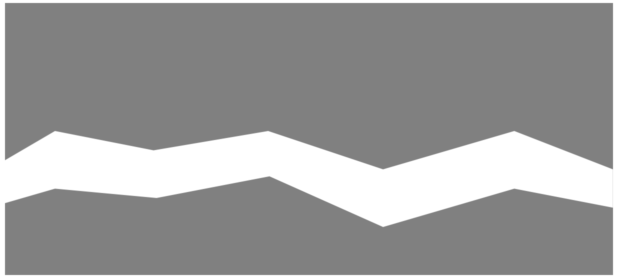
$P(X_n)$

H	0.5
T	0.5



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

2^n



Conditional Independence

- Unconditional (absolute) independence is very rare
- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments
- X is conditionally independent of Y given Z $X \perp\!\!\!\perp Y | Z$

  if and only if:

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

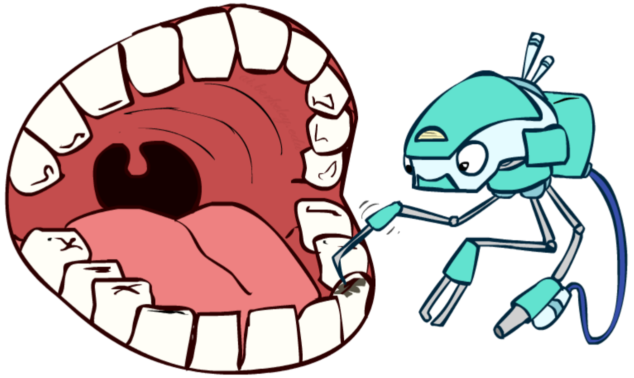
or, equivalently, if and only if

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

Conditional Independence

P(Toothache, Cavity, Catch)

- If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:
  - P(+catch | +cavity, +toothache) = P(+catch | +cavity)
- The same independence holds if I don't have a cavity:
  - P(+catch | -cavity, +toothache) = P(+catch| -cavity)

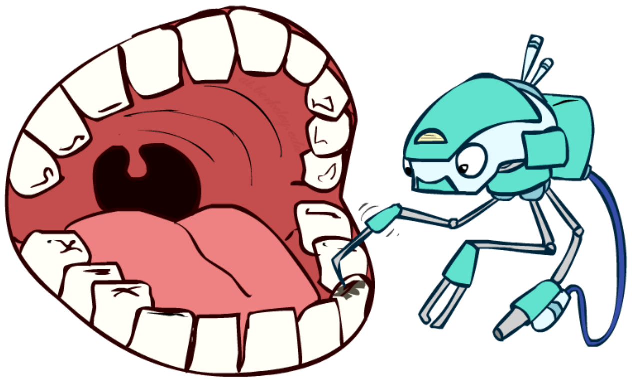


- Catch is *conditionally independent* of Toothache given Cavity:
  - P(Catch | Cavity, Toothache) = P(Catch | Cavity)

Conditional Independence

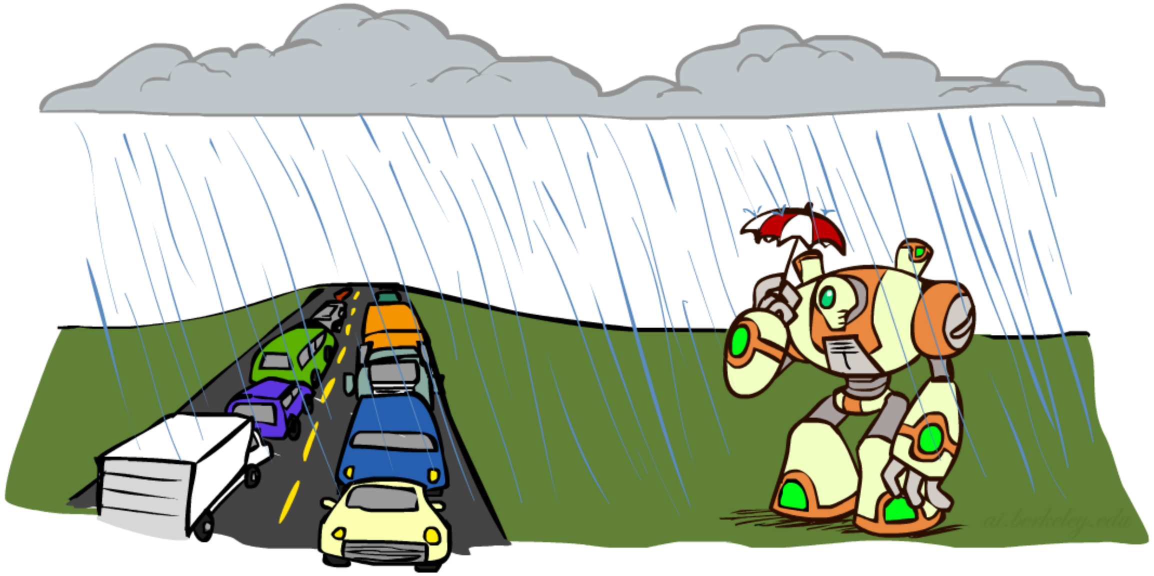
Catch is *conditionally independent* of Toothache given Cavity:

- Equivalent statements:
  - P(Toothache | Catch , Cavity) = P(Toothache | Cavity)
  - P(Toothache, Catch | Cavity) = P(Toothache | Cavity) P(Catch | Cavity)



Conditional Independence

- What about this domain?
  - Traffic
  - Umbrella
  - Rain



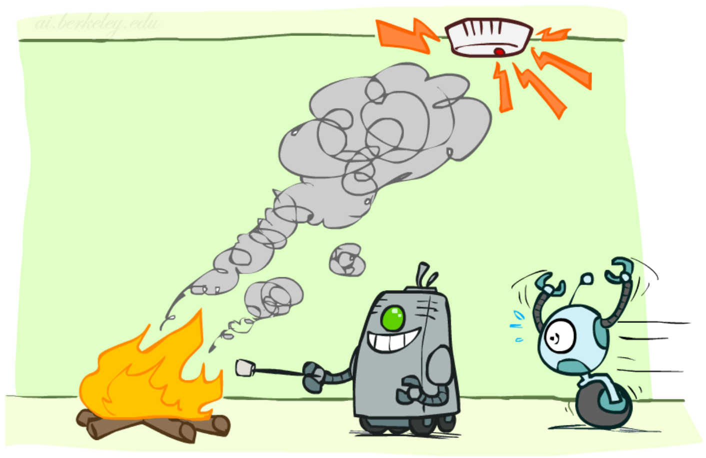
Traffic $\perp\!\!\!\perp$ Umbrella | Rain

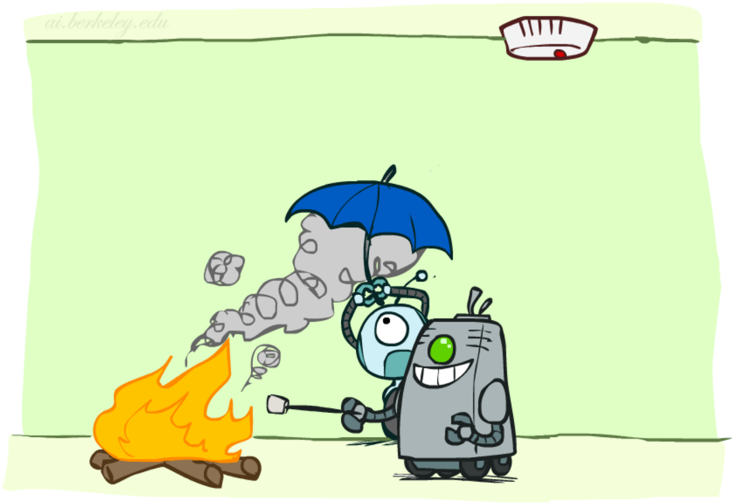
Conditional Independence

- What about this domain:

  - Fire
  - Smoke
  - Alarm

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





Conditional Independence and the Chain Rule

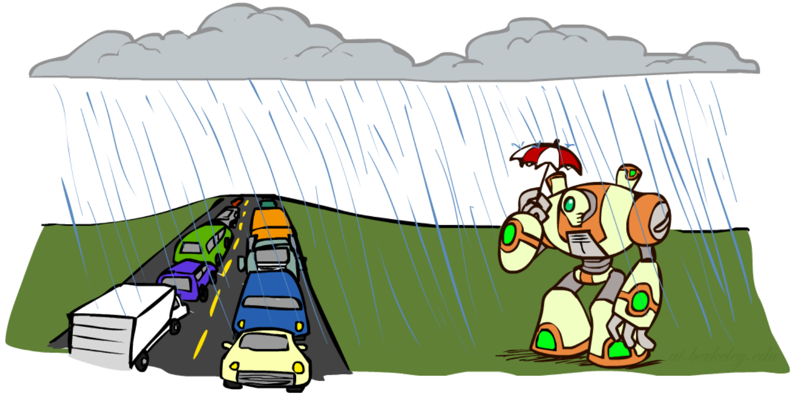
- Chain rule: $P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$

- Trivial decomposition:

  P(Rain, Traffic, Umbrella) = P(Rain) P(Traffic|Rain) P(Umbrella|Rain, Traffic)

- With assumption of conditional independence:

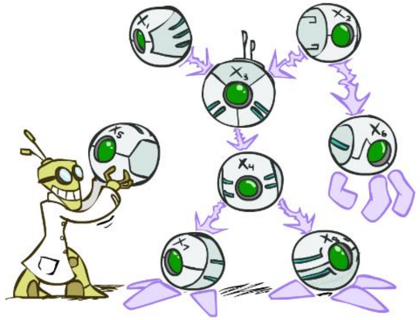
  P(Rain, Traffic, Umbrella) = P(Rain) P(Traffic|Rain) P(Umbrella|Rain)



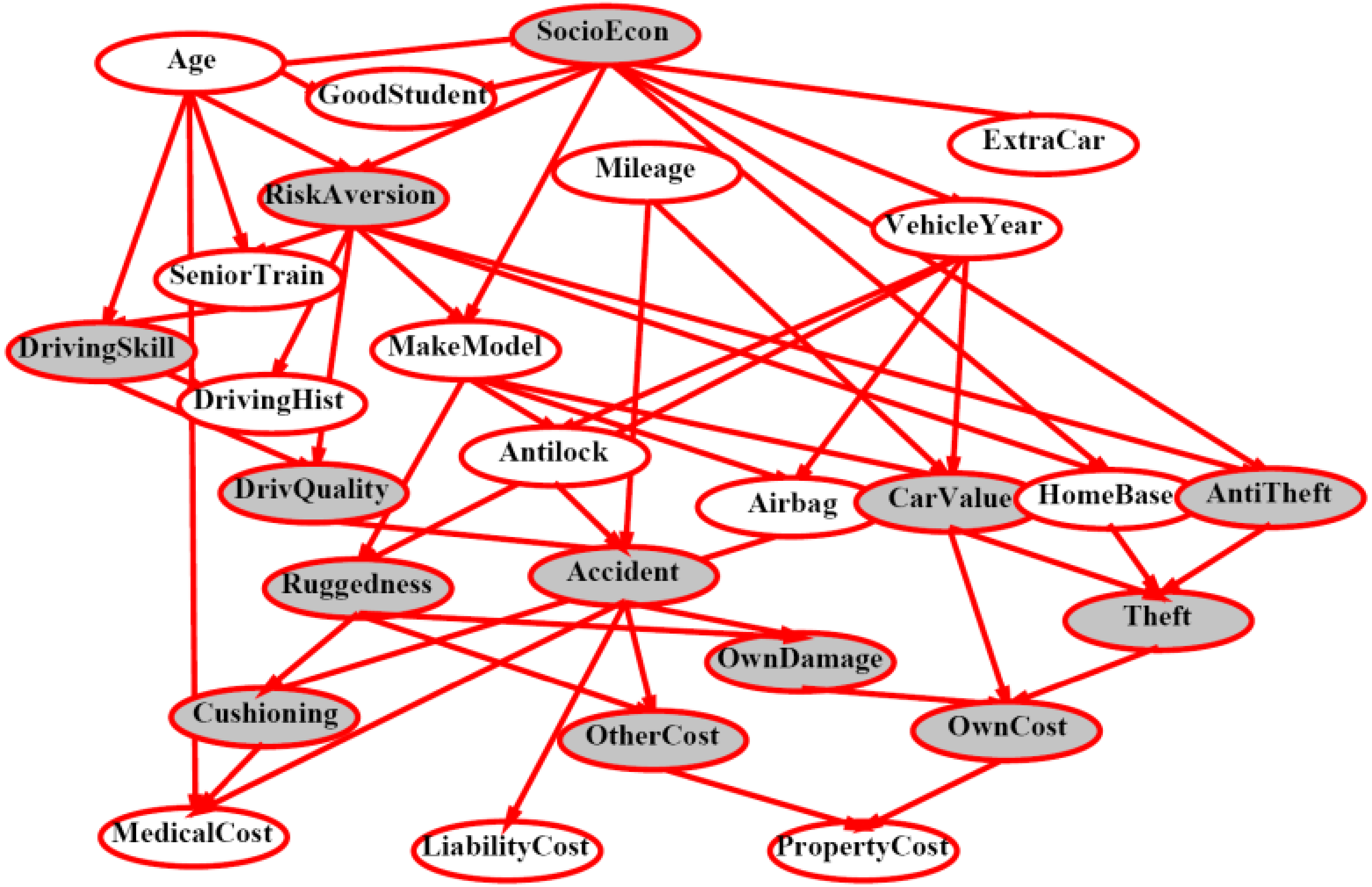
Bayes' Nets: Big Picture

Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)

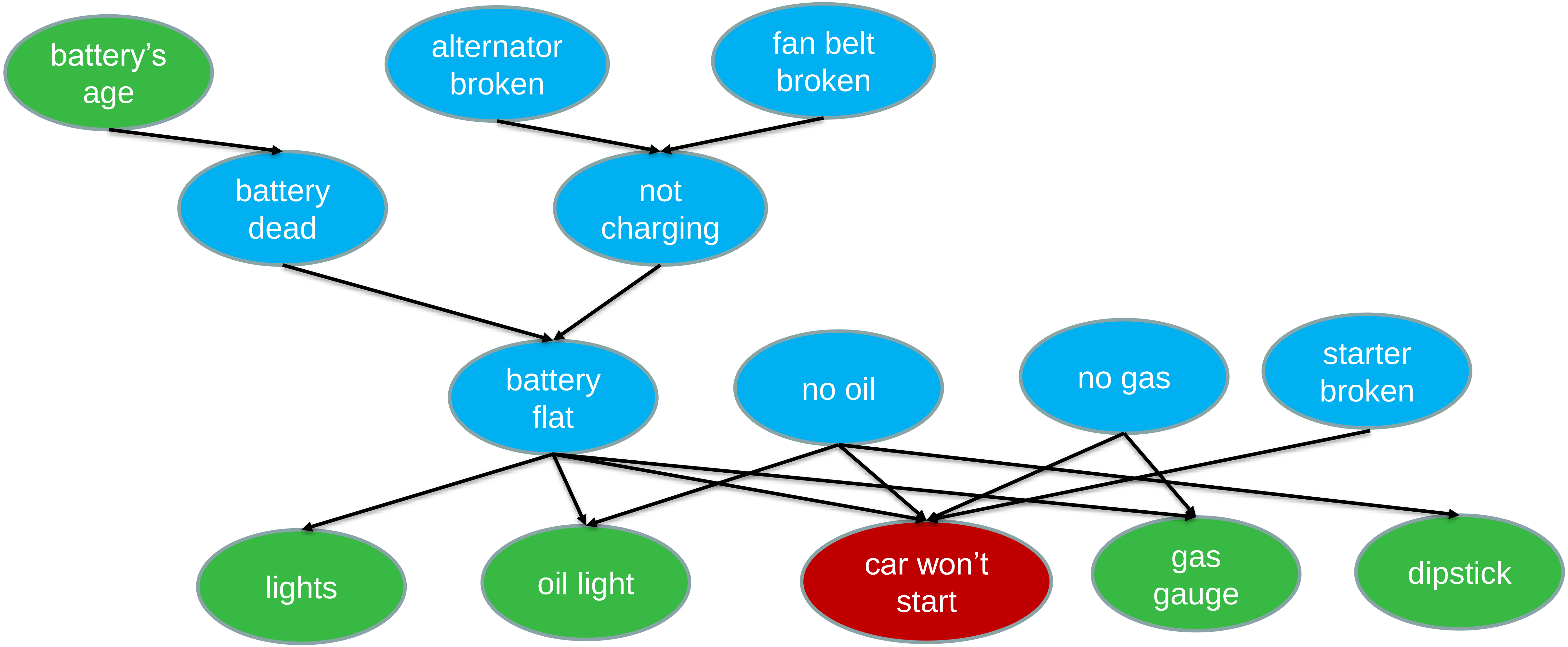
- More properly called graphical models
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions



Example Bayes' Net: Insurance

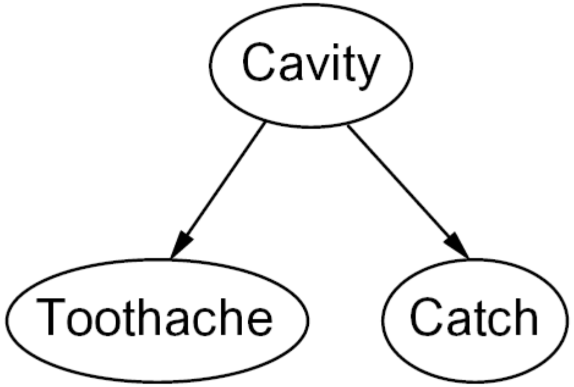


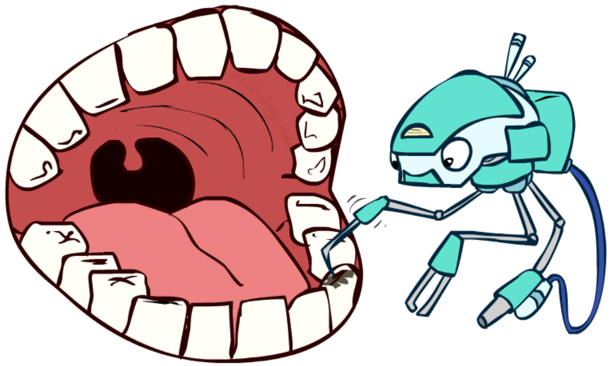
Example Bayes' Net: Car



Graphical Model Notation

- Nodes: variables (with domains)
  - Can be assigned (observed) or unassigned (unobserved)
- Arcs: interactions
  - Similar to CSP constraints
  - Indicate “direct influence” between variables
  - Formally: encode conditional independence (more later)



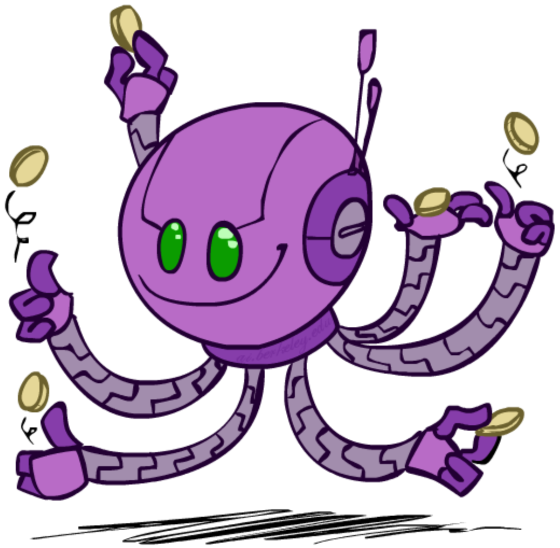


- For now: imagine that arrows mean direct causation

Example: Coin Flips

- N independent coin flips

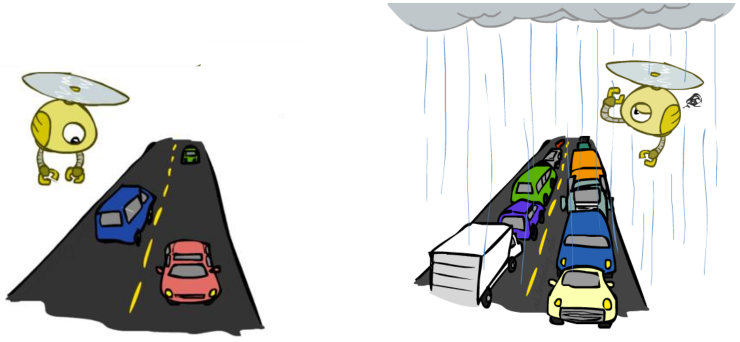
X_1 $\quad$ X_2 $\quad \cdots \quad$ X_n



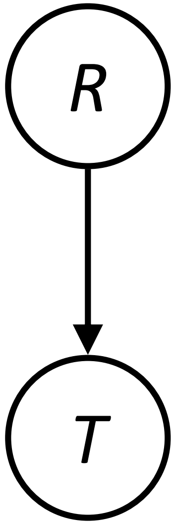
- No interactions between variables: absolute independence

Example: Traffic

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

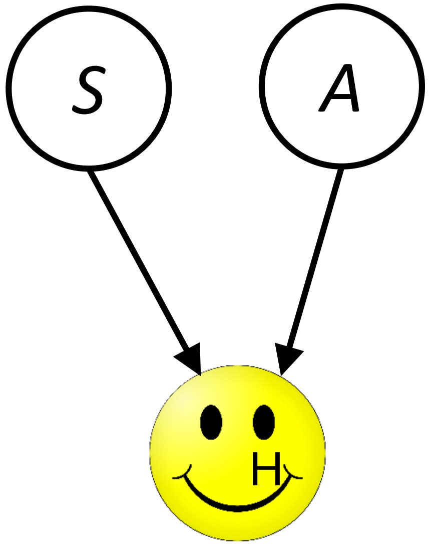


- Model: rain causes traffic



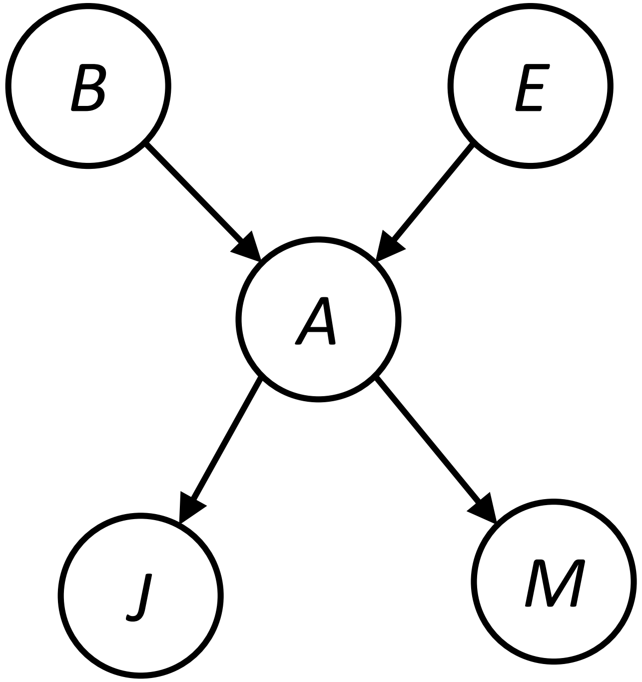
Example: Happiness

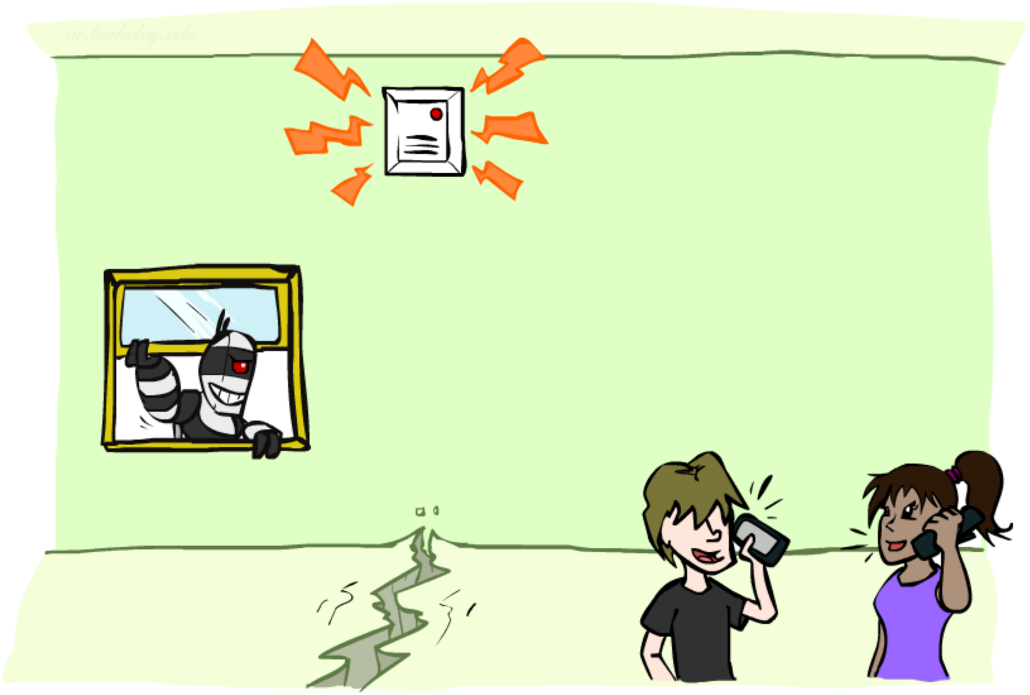
- Variables:
  - A: Anna gets an A
  - S: It is sunny
  - H: Anna is happy



Example: Alarm Network

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

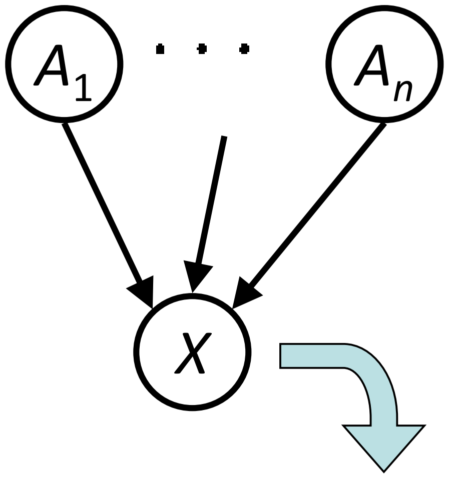




Bayes' Net Semantics

- A directed, acyclic graph, one node per random variable
- A conditional probability table (CPT) for each node
  - A collection of distributions over X, one for each combination of parents' values

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



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

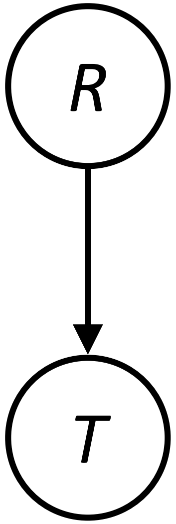
A Bayes net = Topology (graph) + Local Conditional Probabilities

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: Traffic



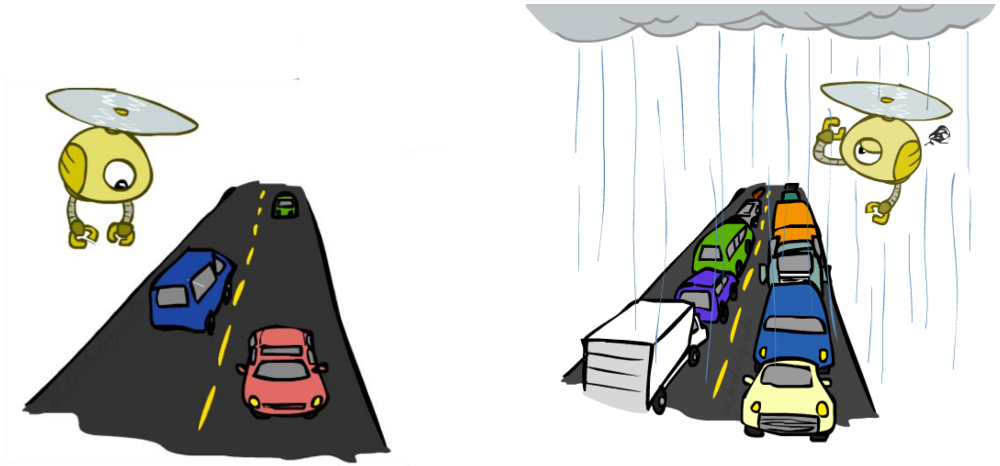
$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

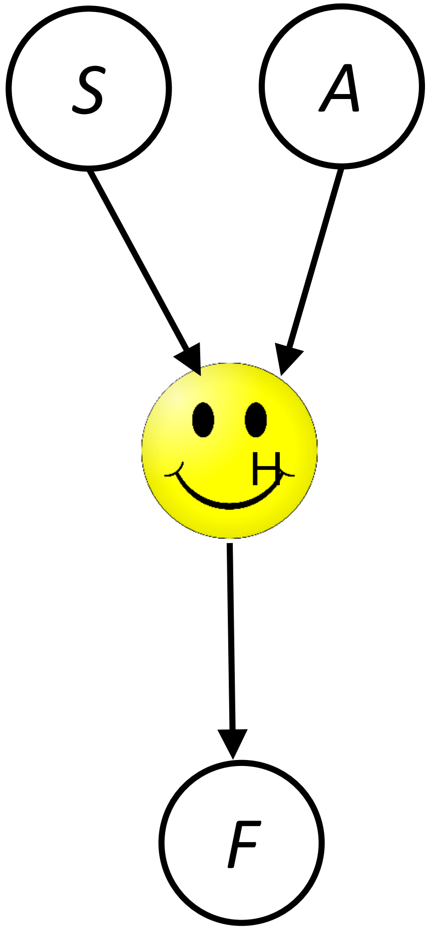
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$$P(+r, -t) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$$



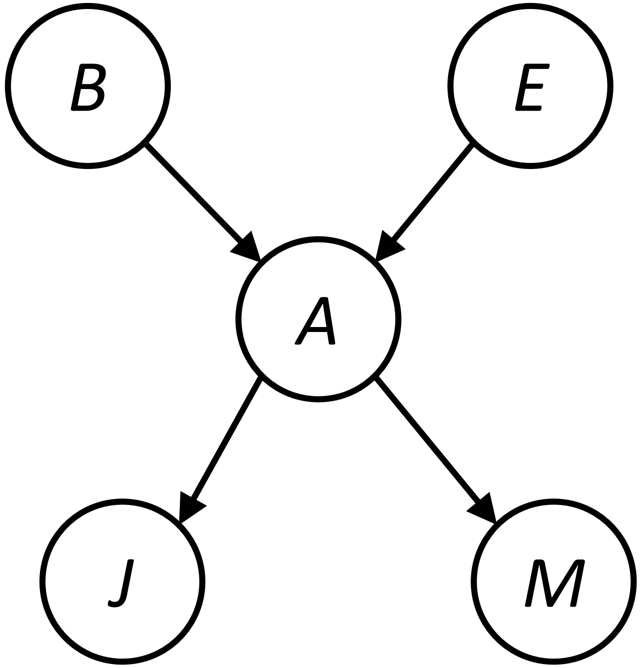
Example: Happiness in the Applet

- Variables:
  - A: Anna gets an A
  - S: It is sunny
  - H: Anna is happy
  - F: Anna buys flowers

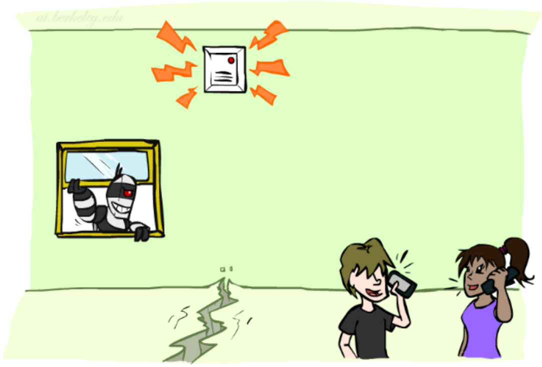


Example: Alarm Network

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

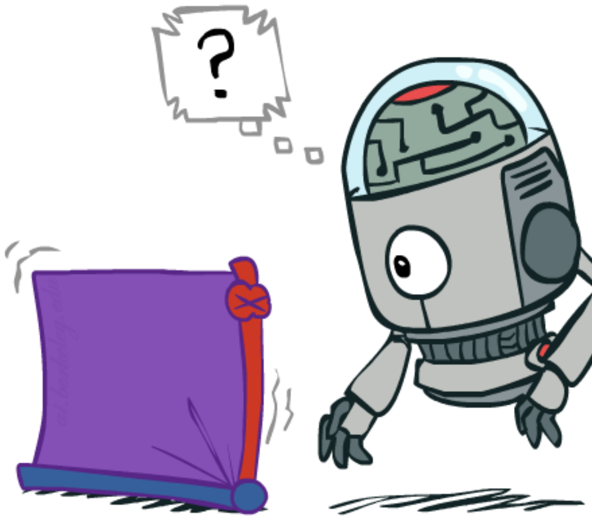
- How big is an N-node net if nodes have up to k parents?

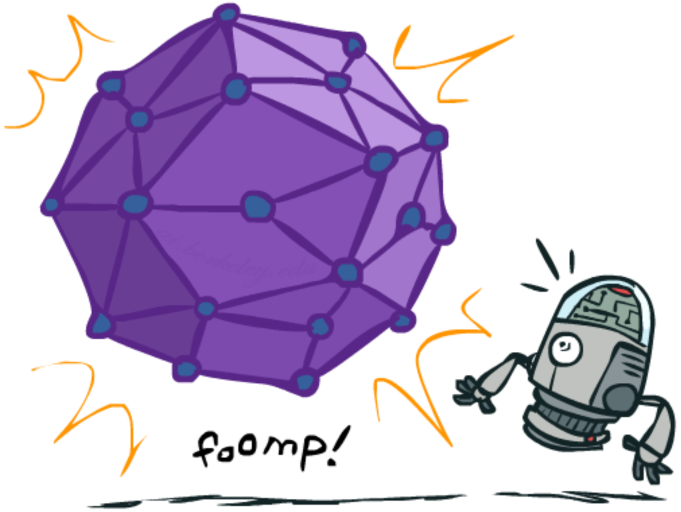
  $O(N * 2^{k+1})$

- 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)





Conditional Independence

- X and Y are independent if

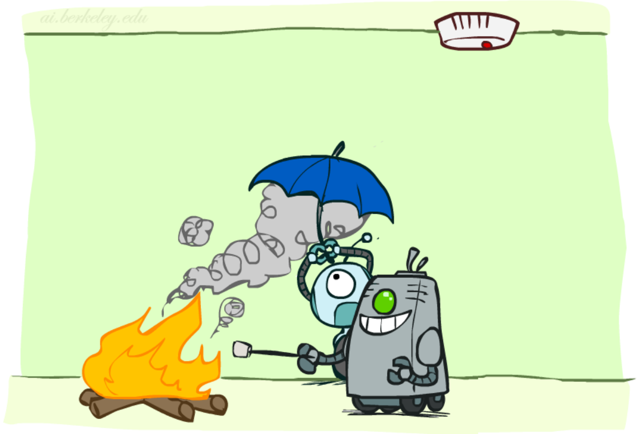
$$\forall x, y \ P(x,y) = P(x)P(y) \dashrightarrow X \perp\!\!\!\perp Y$$

- X and Y are conditionally independent given Z

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

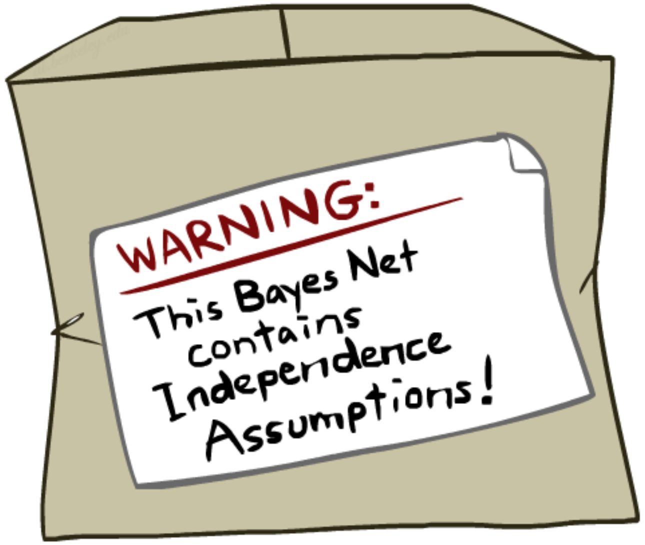
- Example:

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

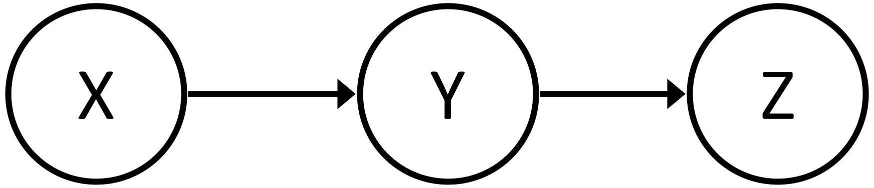


Bayes Nets: Assumptions

- Some “chain rule → Bayes net” conditional independence assumptions
- Often additional conditional independences
  - They can be read off the graph



Bayes Nets: Assumptions



- Conditional independence assumptions directly from chain rule:

  $P(X, Y, Z) = P(X)\ P(Y|X)\ P(Z|X, Y)$

  This is a Bayes' net, so:

  $P(X, Y, Z) = P(X)\ P(Y|X)\ P(Z|Y)$

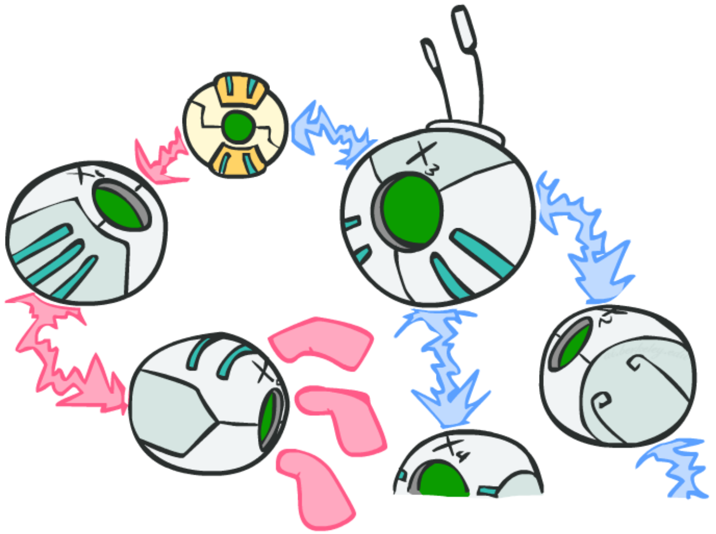
  So $P(Z|X, Y) = P(Z|Y)$

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

- Additional implied conditional independence assumptions?

D-separation: Outline

- Study independence properties for triples
- Analyze complex cases in terms of member triples
- D-separation: a condition / algorithm for determining whether two variables are independent



Causal Chains

This configuration is a "causal chain"

X: Low pressure Y: Rain Z: Traffic

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

- X independent of Z ?
  No!

Causal Chains

This configuration is a "causal chain"

X: Low pressure   Y: Rain   Z: Traffic

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

- X independent of Z given Y?

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

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

$$= P(z|y)$$

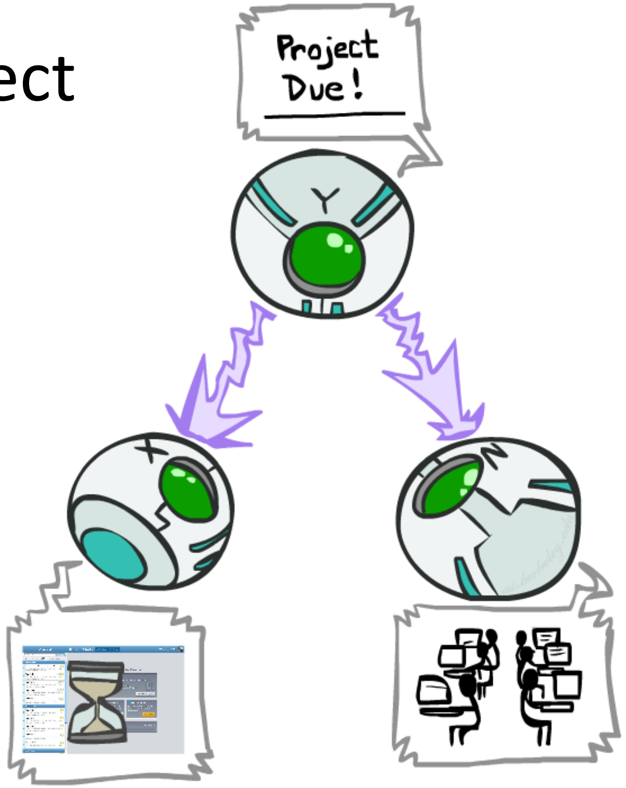
Yes!

Evidence along the chain "blocks" the influence

Common Cause

This configuration is a "common cause"

Y: Project due



X: Forums busy

Z: Lab full

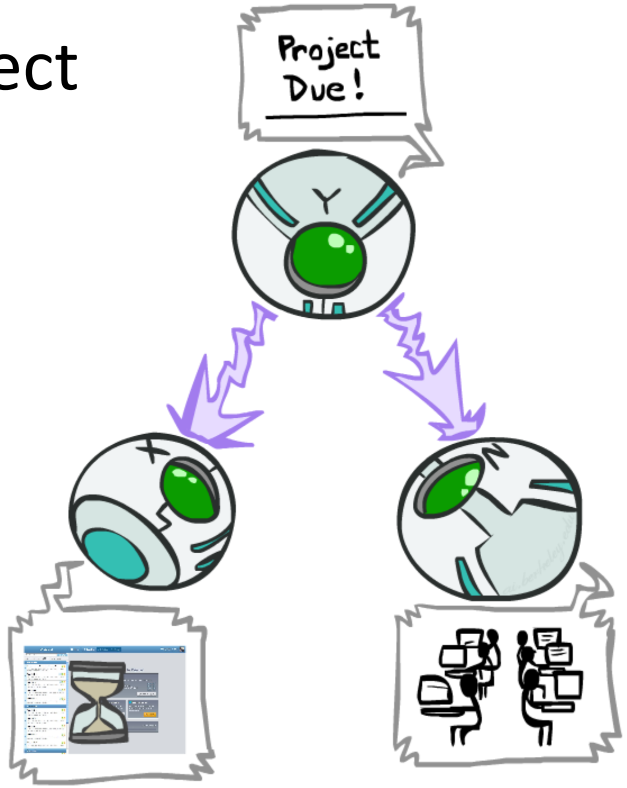
$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

- X independent of Z ?
  No!

Common Cause

This configuration is a "common cause"

Y: Project due



X: Forums busy

Z: Lab full

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

- X and Z independent given Y?

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

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

$$= P(z|y)$$

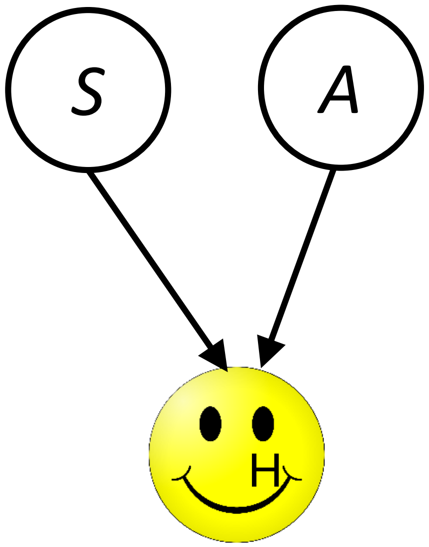
Yes!

Observing the cause blocks influence between effects

Common Effect

Common effect (v-structure)

X: It is sunny   Y: Anna gets A



Z: Happiness

- Are X and Y independent?
  - *Yes*: they both make Anna happy, but they are not correlated
- Are X and Y independent given Z?
  - *No*: Anna's happiness puts the grade and the weather in causal competition.

Observing an effect activates influence between possible causes.

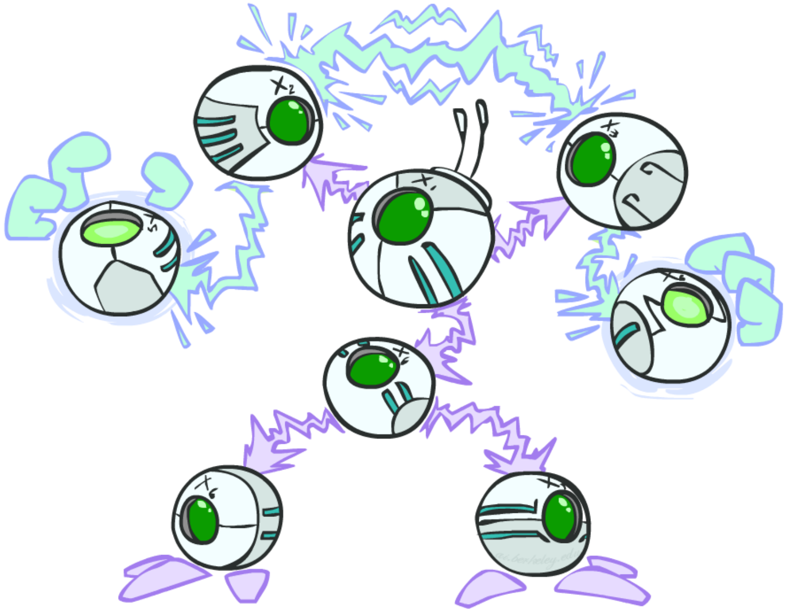
The General Case

The General Case

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

- Solution: analyze the graph

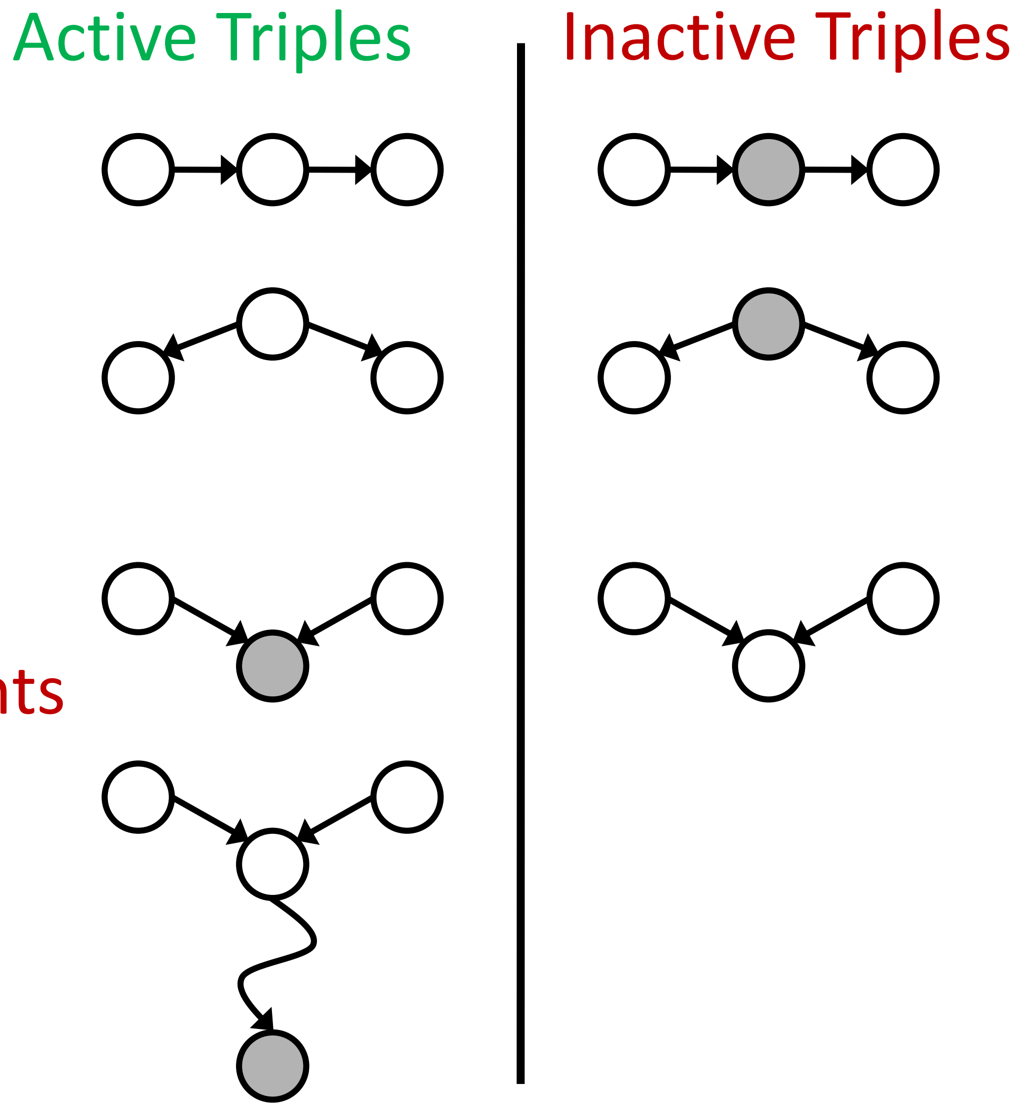
- Any complex example can be broken into repetitions of the three canonical cases



Active / Inactive Paths

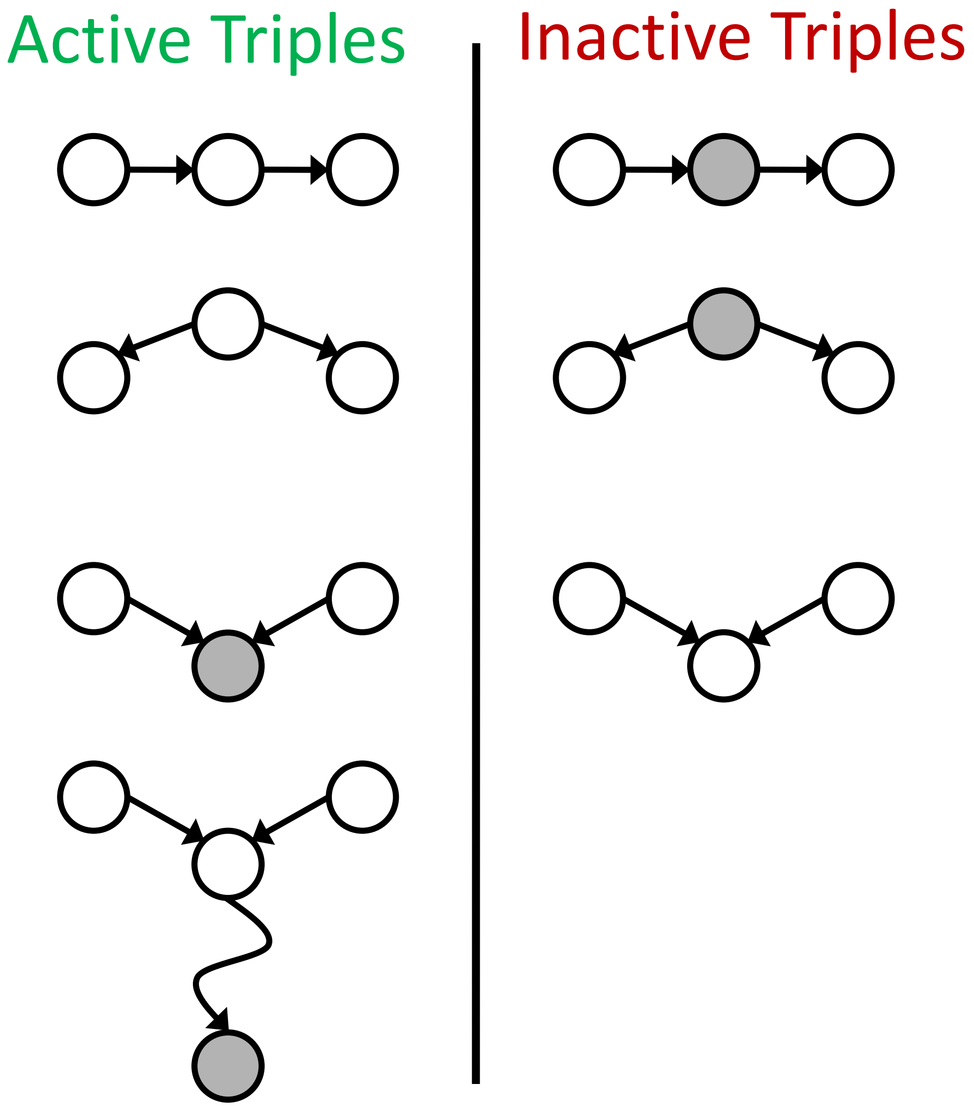
- A triple is active:
  - Causal chain $X \rightarrow Y \rightarrow Z$ where Y is unobserved (either direction)
  - Common cause $X \leftarrow Y \rightarrow Z$ where Y is unobserved
  - Common effect (aka v-structure)

    $X \rightarrow Y \leftarrow Z$ where Y or one of its descendants is observed



Active / Inactive Paths

- Are X and Y conditionally independent given evidence variables Z?
  - Consider all (undirected) paths from X to Y
  - No active paths = independence!
  - X and Y "d-separated" by Z
- A path is active if each triple is active
- All it takes to block a path is a single inactive segment



D-Separation

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

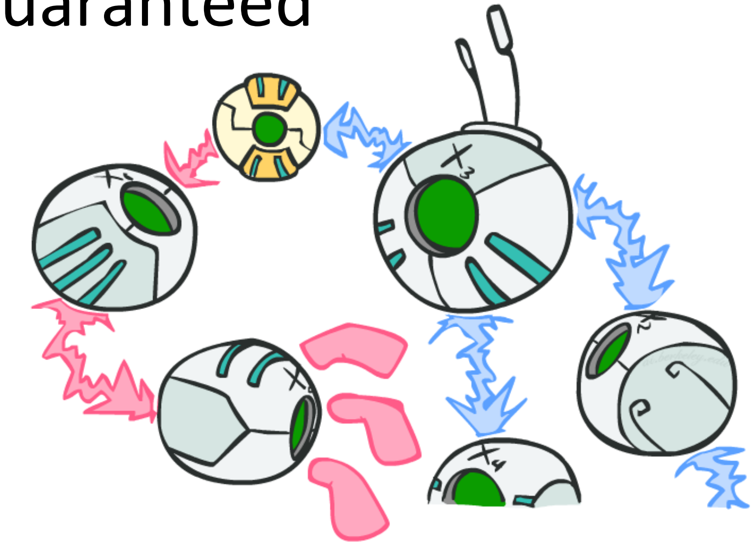
- Check all (undirected!) paths between X_i and X_j

  - 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

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

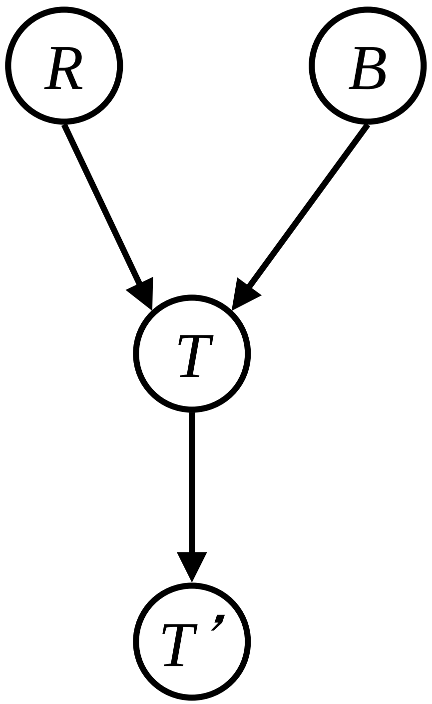


Example

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

$R \perp\!\!\!\perp B | T$ *No (RTB active)*

$R \perp\!\!\!\perp B | T'$ *No (RTB active)*



Example

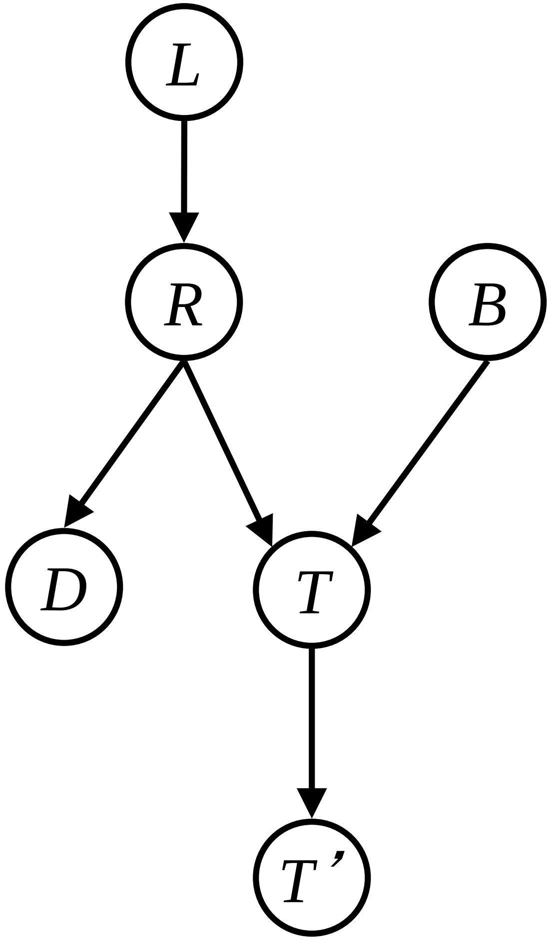
$L \perp\!\!\!\perp T' | T$ *Yes*

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

$L \perp\!\!\!\perp B | T$ *No (LRTB active)*

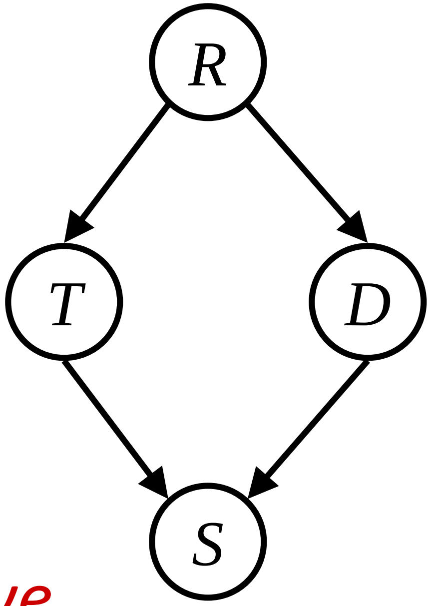
$L \perp\!\!\!\perp B | T'$ *No (LRTB active)*

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



Example

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



$T \perp\!\!\!\perp D$ *No TRD active*

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

$T \perp\!\!\!\perp D | R, S$ *No TSD ACTIVE*

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