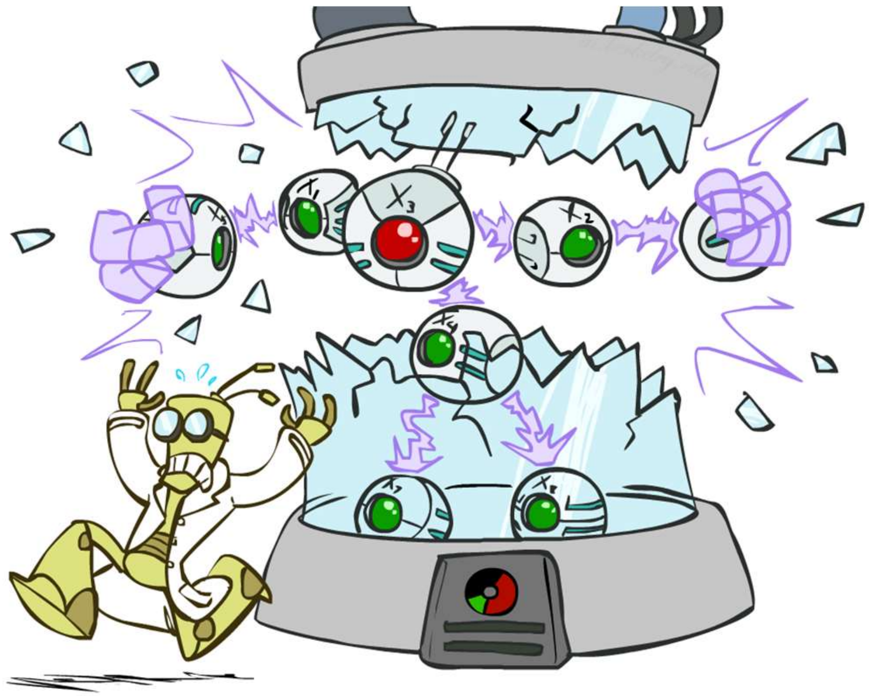
Bayes' Nets: Independence



Instructors: Fatemeh Mansoori--- University of Isfahan

[These slides were created by Dan Klein and Pieter Abbeel for AI Course at UC Berkeley.]

Probability Recap

onditional probability

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

oduct rule

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

ain rule

$$P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1,X_2)\ldots$$
$$= \prod_{i=1}^{n} P(X_i|X_1,\ldots,X_{i-1})$$

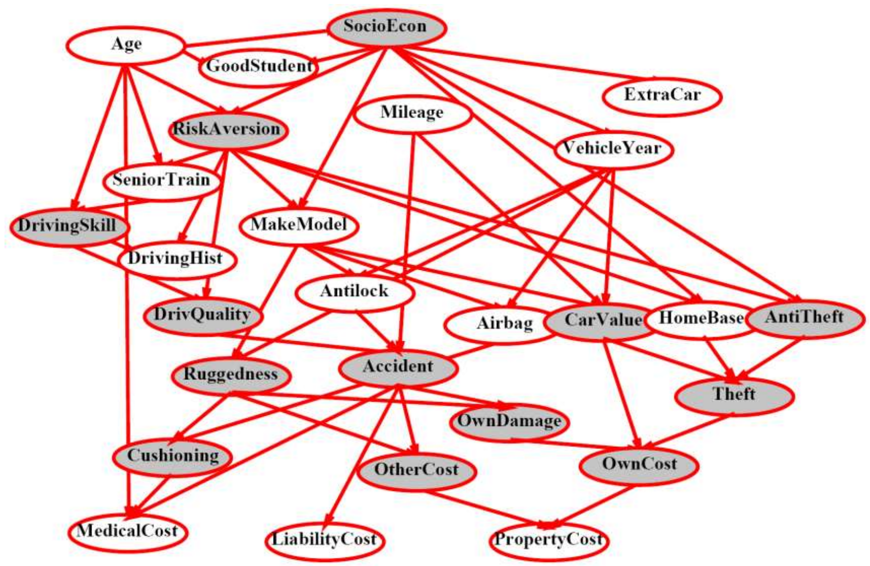
Y independent if and only if: $\forall x,y : P(x,y) = P(x)P(y)$

and Y are conditionally independent given Z if and only if:

$$\forall x,y,z : P(x,y|z) = P(x|z)P(y|z) \qquad X \perp\!\!\!\perp$$

Bayes’ Nets

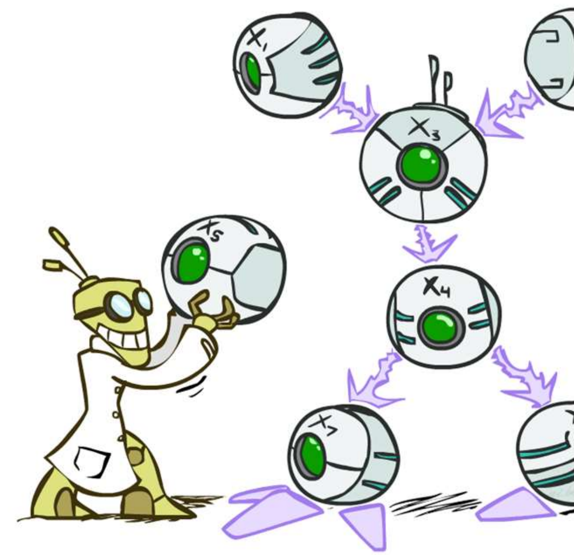
Bayes’ net is an
fficient encoding
f a probabilistic
odel of a domain



uestions 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’ Net Semantics

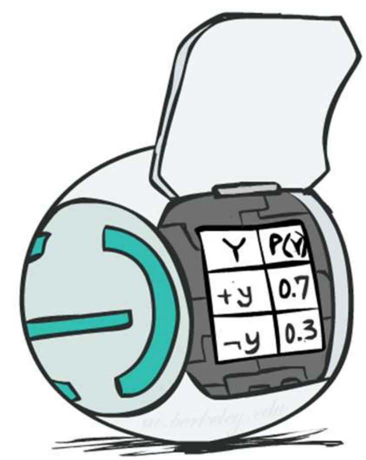


- directed, acyclic graph, one node per random variable
- conditional probability table (CPT) for each node
  - A collection of distributions over X, one for each combination of parents’ values

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

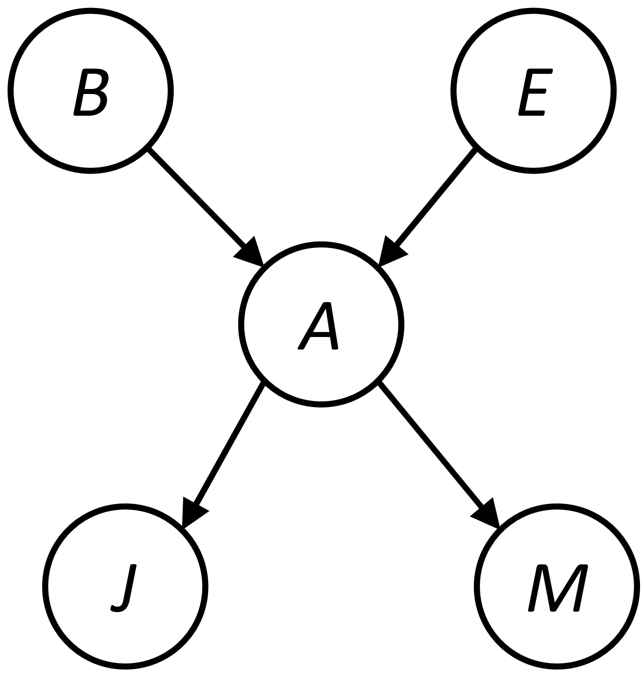
- ayes’ nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

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

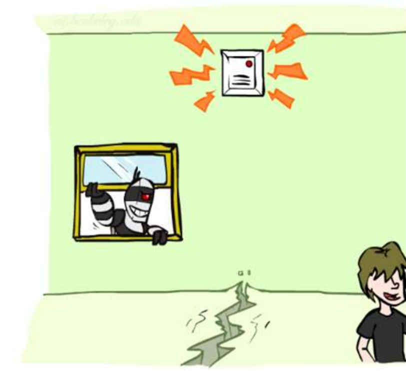


Example: Alarm Network

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

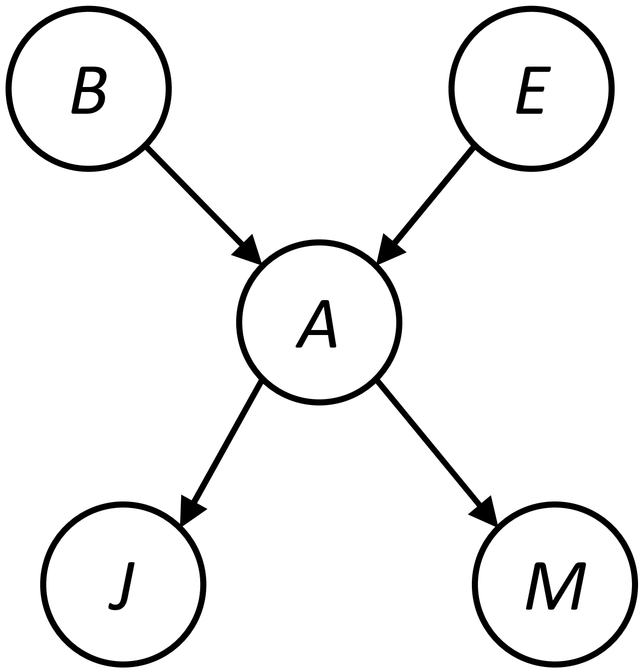
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

B	E	A	P(A|B,
+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

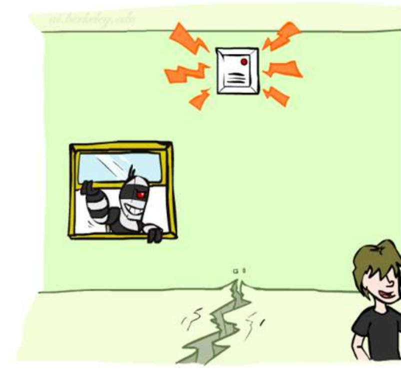
$$+b, -e, +a, -j, +m) =$$

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,
+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

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

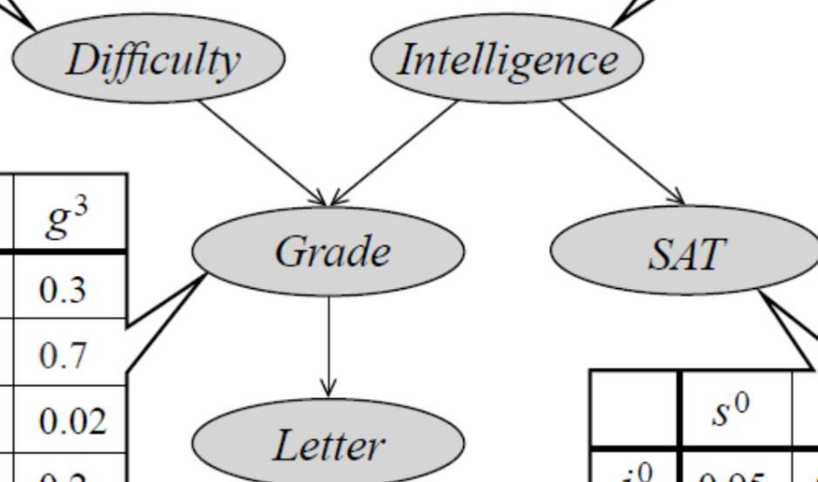
$$b)P(-e)P(+a|+b,-e)P(-j|+a)P(+m|+a) =$$

$$1 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

Example : Student

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3



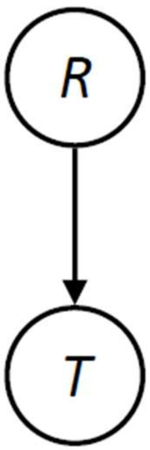
	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

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

ausal direction

$P(R)$

+r	1/4
-r	3/4

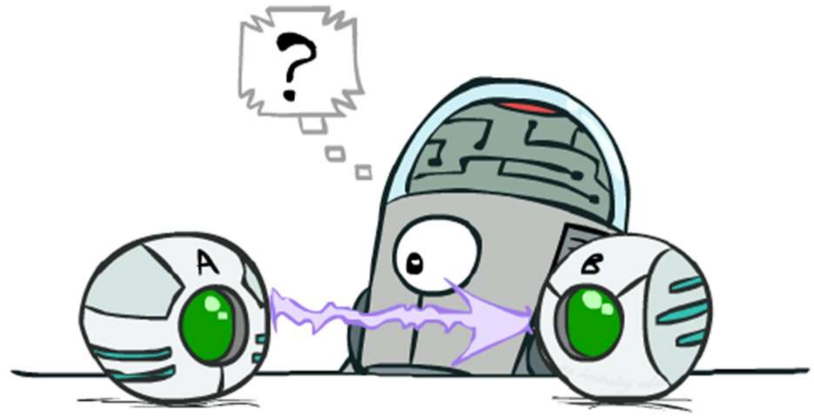
$P(T|R)$

+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

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

- v big is a joint distribution over N
  lean variables?

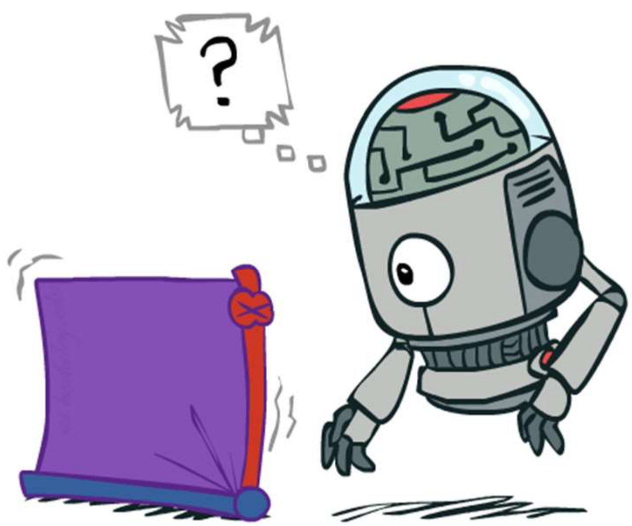
- v big is an N-node net if nodes
  e up to k parents?

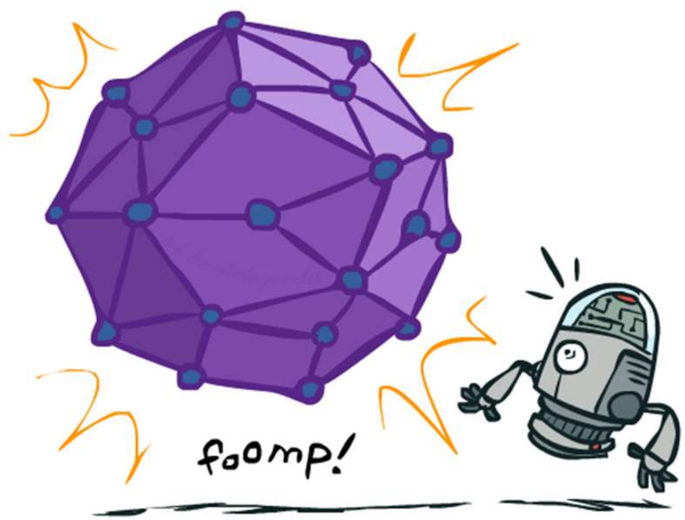
  $(N * 2^{k+1})$

- 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)





Bayes' Nets

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

Conditional Independence

and Y are independent if

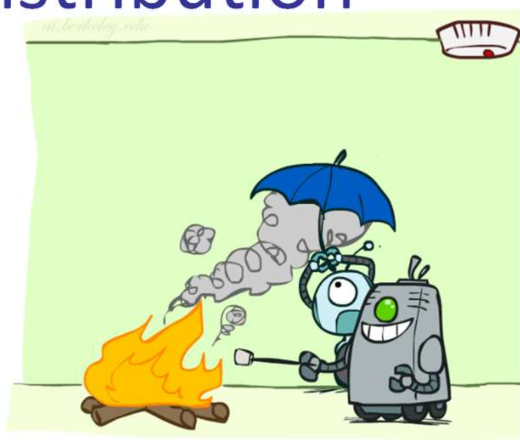
$$\forall x, y \;\; P(x, y) = P(x)P(y) \dashrightarrow X \perp\!\!\!\perp Y$$

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

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



Bayes Nets: Assumptions

umptions we are required to make to define the
yes net when given the graph:

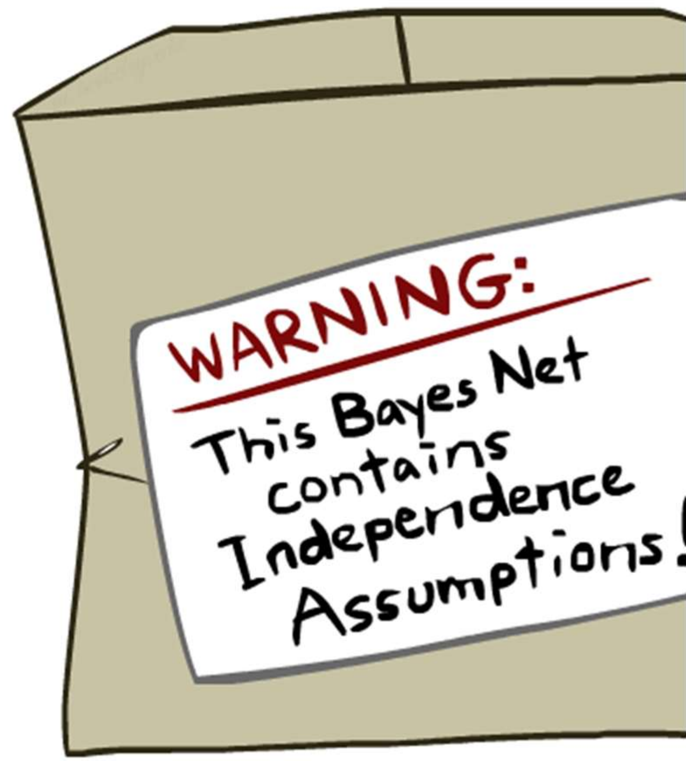
$$P(x_i|x_1 \cdots x_{i-1}) = P(x_i|parents(X_i))$$

yond above "chain rule → Bayes net" conditional
ependence assumptions

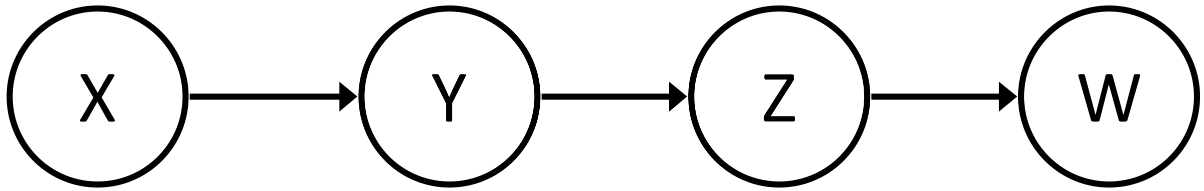
Often additional conditional independences

They can be read off the graph

portant for modeling: understand assumptions made
en choosing a Bayes net graph



Example

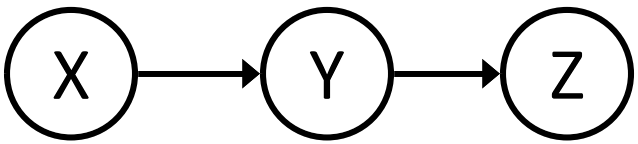


onditional independence assumptions directly from simplifications in chain rule:

dditional implied conditional independence assumptions?

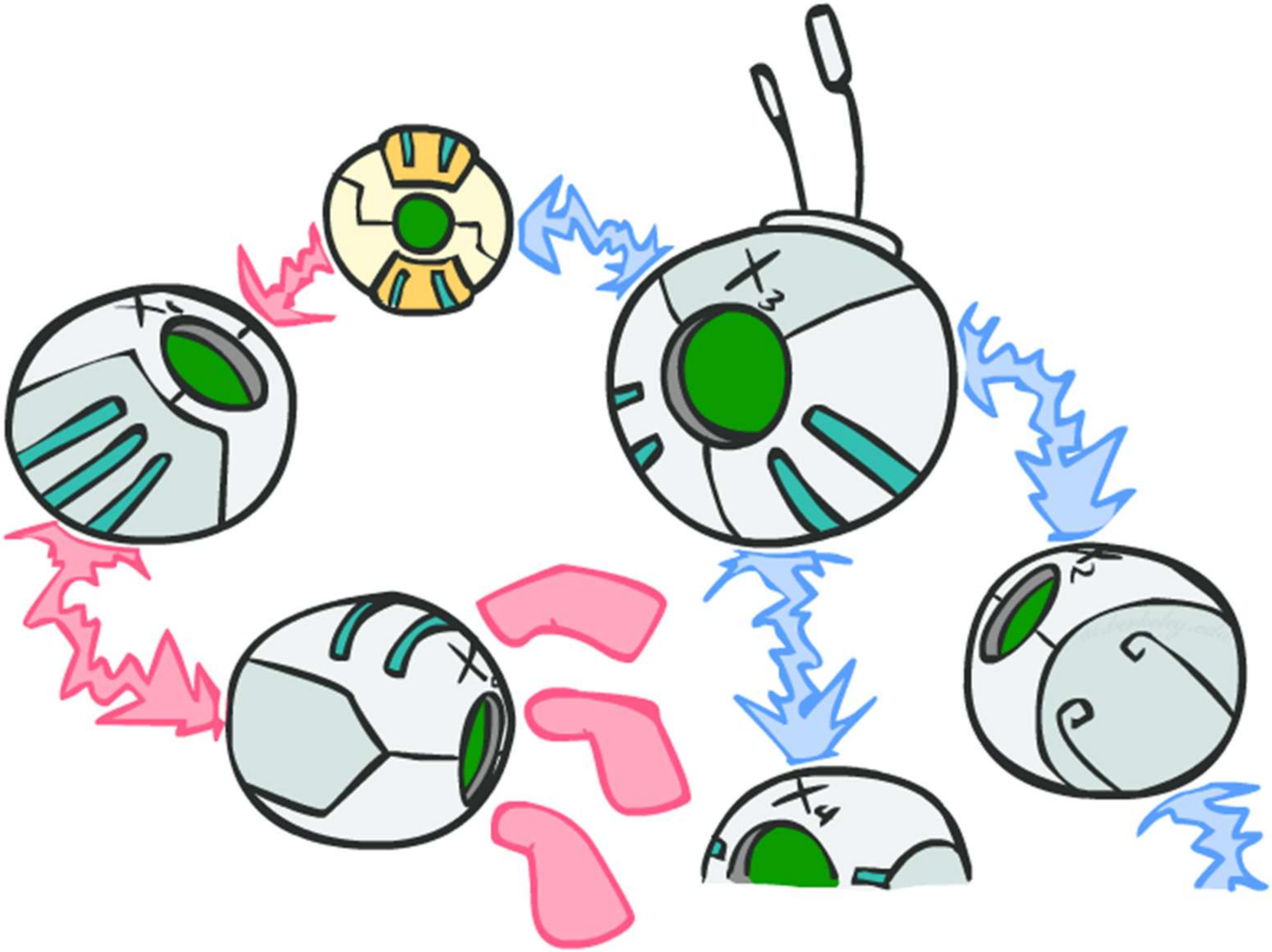
Independence in a BN

- 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

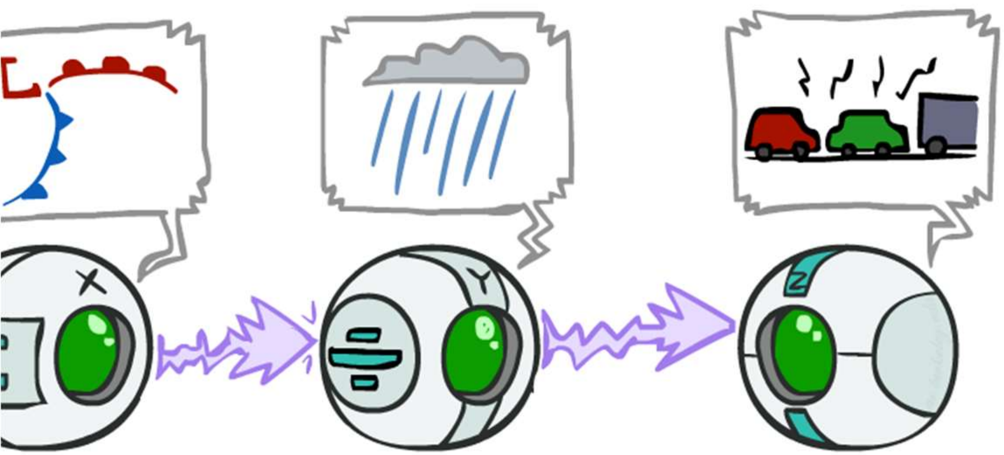


D-separation: Outline

- Study independence properties for triples
- Analyze complex cases in terms of member triples
- D-separation: a condition / algorithm for answering such queries

Causal Chains

- is configuration is a "causal chain"



Low pressure Y: Rain Z: Traffic

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

- Guaranteed X independent of Z ? *N*
  - One example set of CPTs for which X is n independent of Z is sufficient to show th independence is not guaranteed.
  - Example:
    - Low pressure causes rain causes traf high pressure causes no rain causes traffic
    - In numbers:
      - P(+x) = ½ P(-x) = ½
      - P(+y | +x) = 1, P(-y | - x) = 1,
      - P(+z | +y) = 1, P(-z | -y) = 1

Causal Chains

- is configuration is a "causal chain"

Low pressure　　Y: Rain　　Z: Traffic

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

- Guaranteed X independent of Z giver

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

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

$$= P(z|y)$$

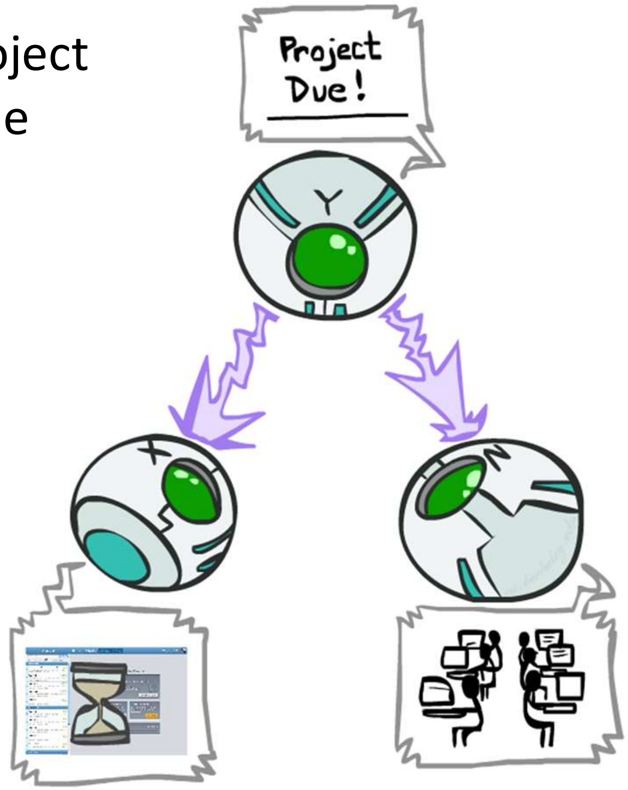
Yes!

- Evidence along the chain "blocks" influence

Common Cause

- is configuration is a “common cause”

: Project due



ns

Z: Lab full

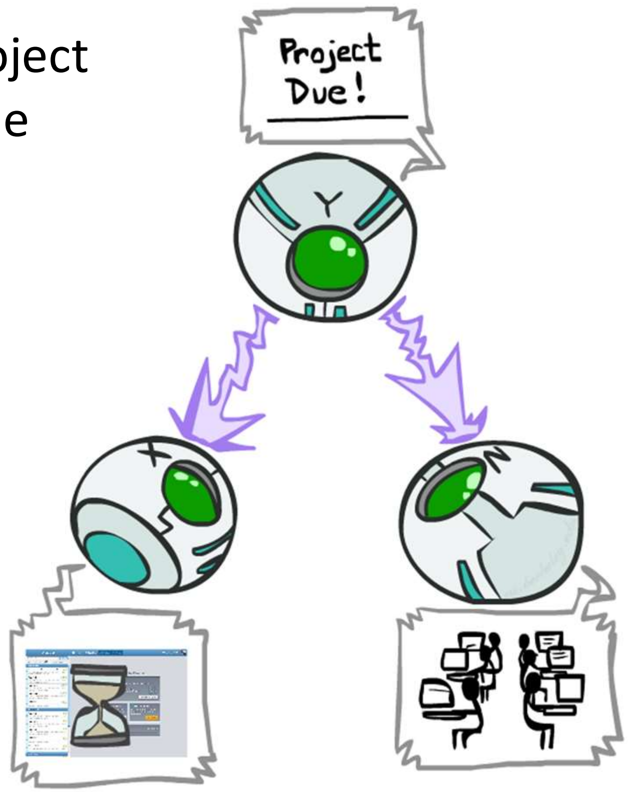
$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

- Guaranteed X independent of Z ? *N*
  - One example set of CPTs for which X is n independent of Z is sufficient to show th independence is not guaranteed.
  - Example:
    - Project due causes both forums bus and lab full
    - In numbers:

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

Common Cause

- is configuration is a "common cause"



: Project due

ns

Z: Lab full

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

- Guaranteed X and Z independent giv

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

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

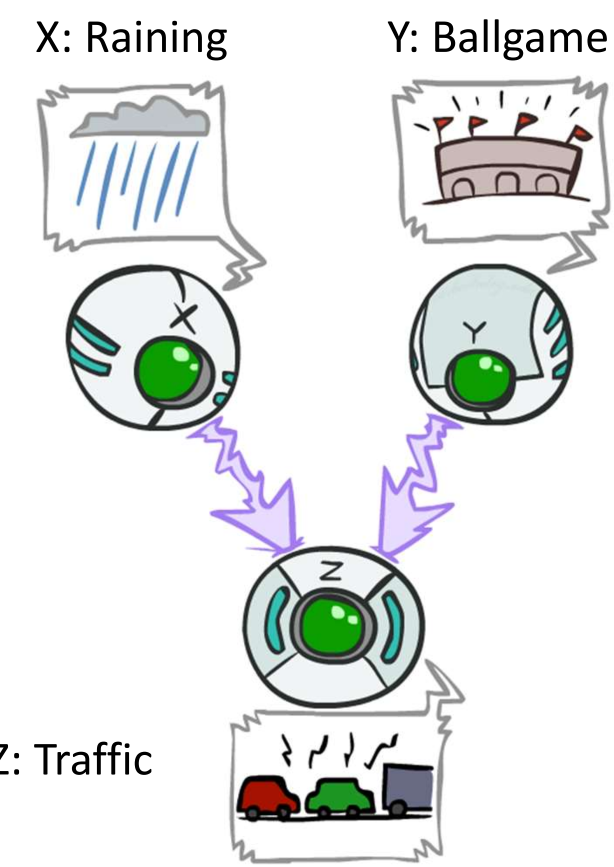
$$= P(z|y)$$

Yes!

- Observing the cause blocks influe between effects.

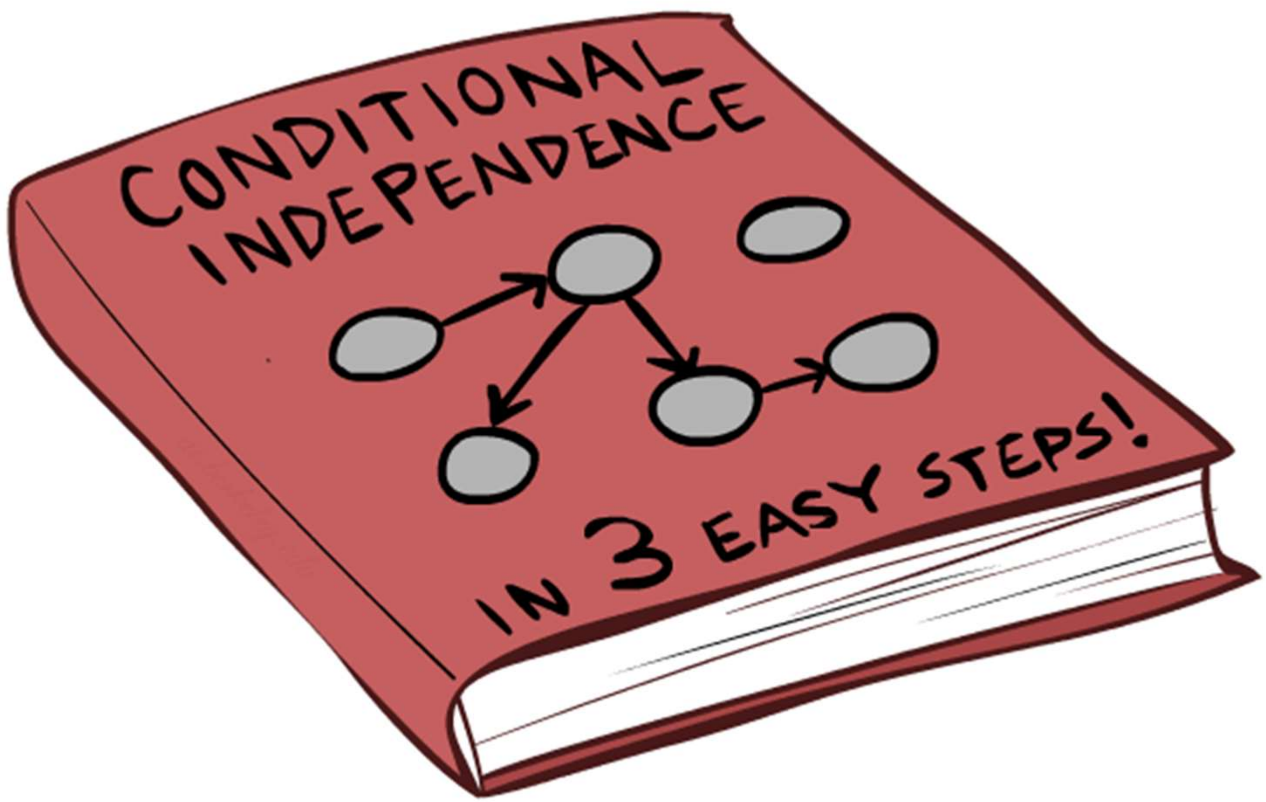
Common Effect

- t configuration: two causes of one ect (v-structures)



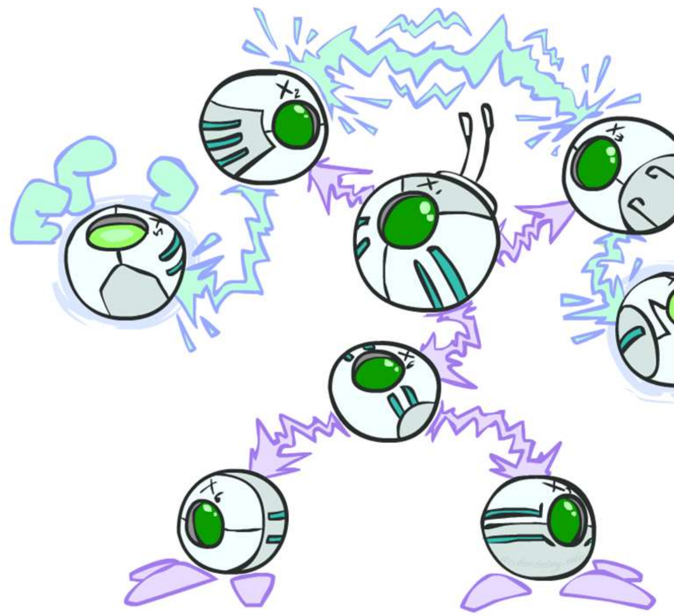
- Are X and Y independent?
  - *Yes*: the ballgame and the rain cause traffic, b they are not correlated
  - Still need to prove they must be (try it!)
- Are X and Y independent given Z?
  - *No*: seeing traffic puts the rain and the ballga competition as explanation.
- This is backwards from the other cases
  - Observing an effect activates influence betwe possible causes.

The General Case



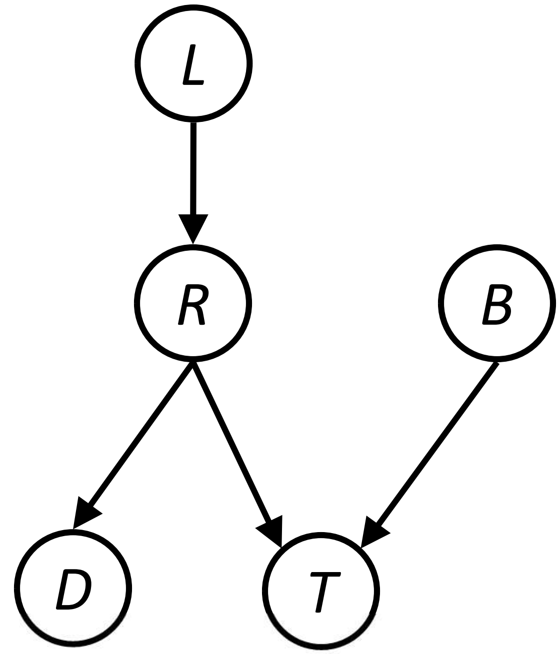
The General Case

- 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

- ecipe: shade evidence nodes, look or paths in the resulting graph
- ttempt 1: if two nodes are connected y an undirected path not blocked by shaded node, they are conditionally ndependent
- lmost 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

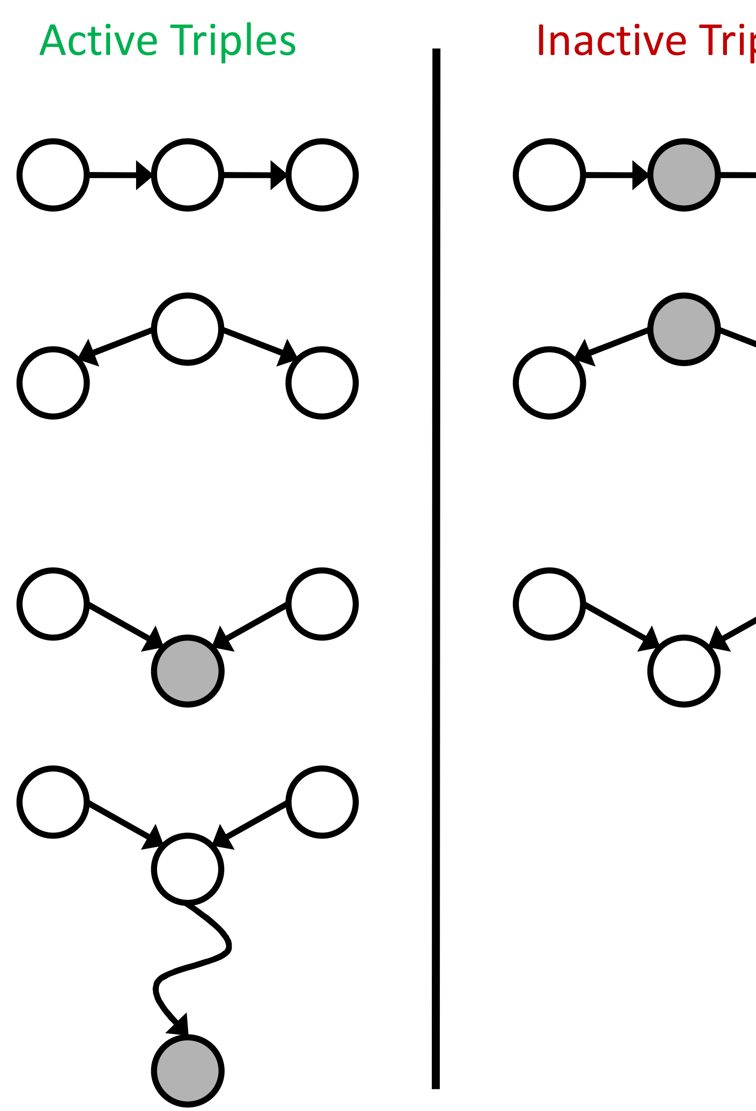
estion: Are X and Y conditionally independent given lence variables {Z}?

- 'es, if X and Y "d-separated" by Z
- Consider all (undirected) paths from X to Y
- No active paths = independence!

ath 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)
  $\rightarrow B \leftarrow C$ where B *or one of its descendents* is observed

t takes to block a path is a single inactive segment



D-Separation

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

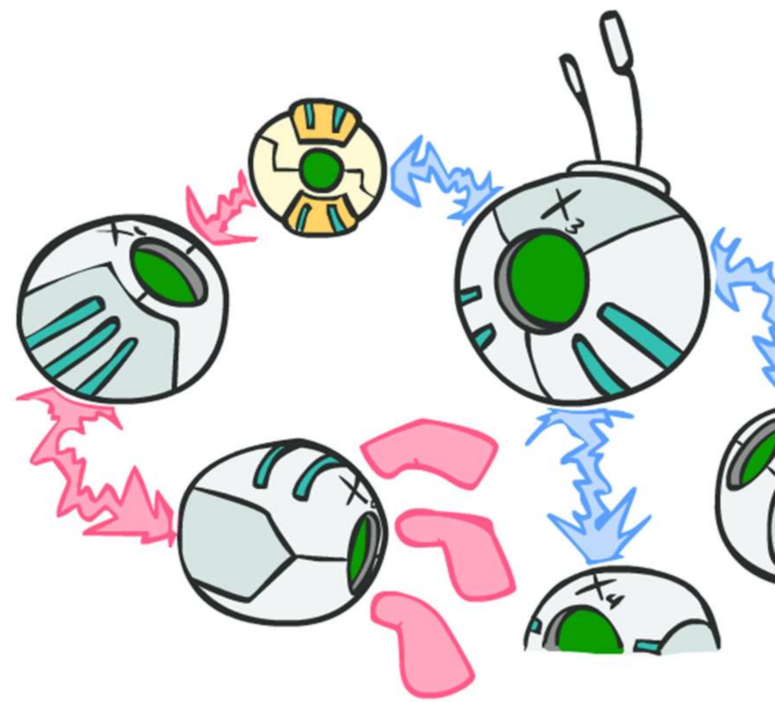
Check all (undirected!) paths between X_i and X_j

- 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

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

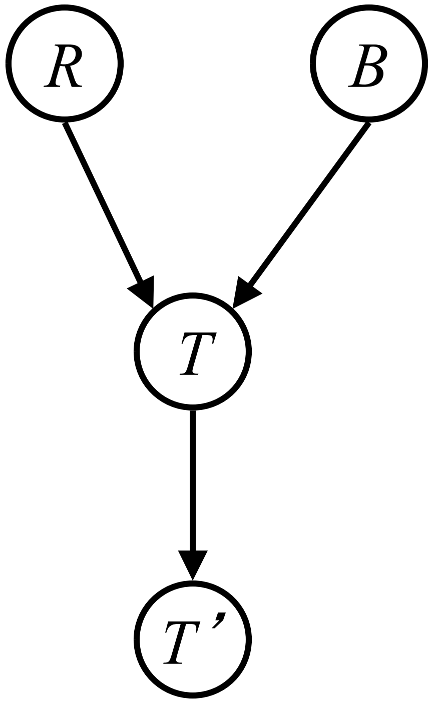


Example

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

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

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



Example

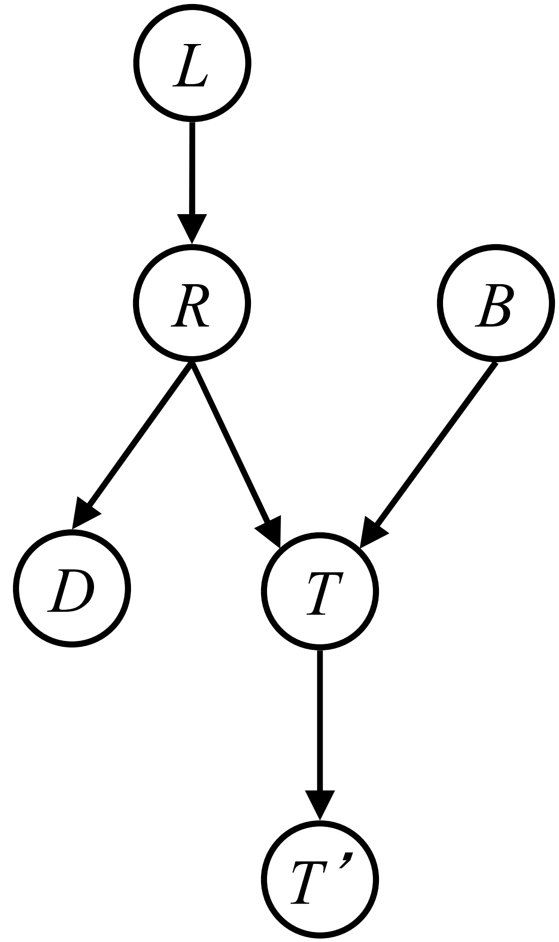
$L \perp\!\!\!\perp T' | T$ Yes

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

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

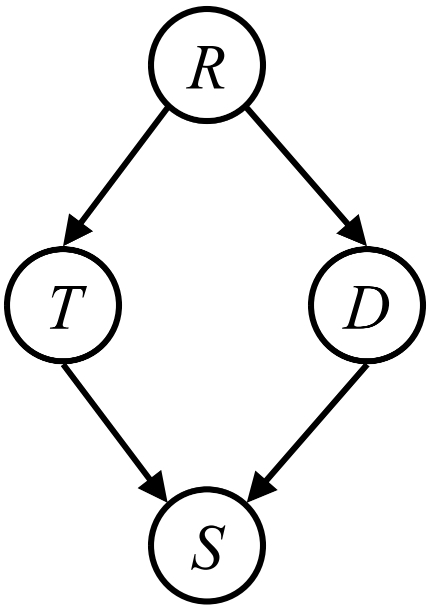
$L \perp\!\!\!\perp B | T'$

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

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

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