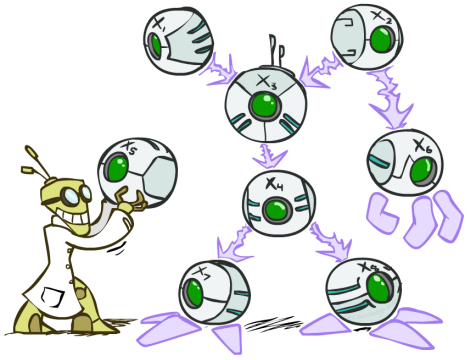
CSE 473: Artificial Intelligence

Bayes' Nets: D-Separation

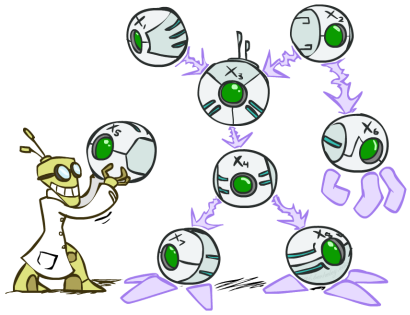


Daniel Weld

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
  - Hard to learn (estimate) anything empirically about more than a few variables at a time

- Bayes' nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly … aka probabilistic graphical model
  - 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

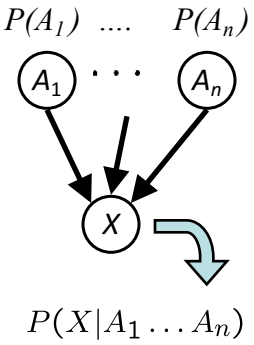


Bayes' Net Semantics

- 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

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

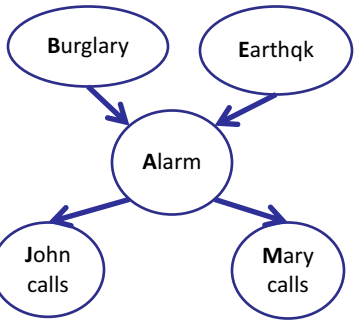
  - CPT: conditional probability table
  - Description of a noisy "causal" process



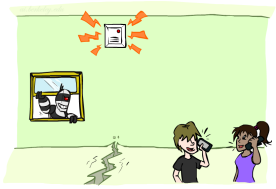
A Bayes net = Topology (graph) + Local Conditional Probabilities

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

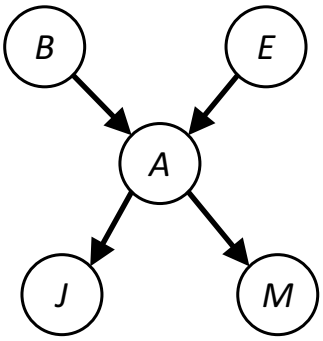
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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

Example: Alarm Network

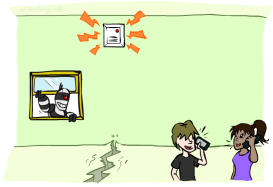
B	P(B)
+b	0.001
-b	0.999

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

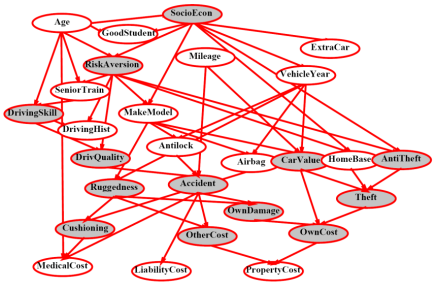
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99



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

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

What's Next with Bayes' Nets



Questions we can ask:

- Definition: P(X = x)
- Representation: given a BN graph, what kinds of distributions can it encode?
- Modeling: what BN is most appropriate for a given domain?
- Inference: given a fixed BN, what is P(X | e)?
- Learning: Given data, what is best BN encoding?

Remember….

- X, Y independent if and only if: $X \perp\!\!\!\perp Y|Z$

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

- X and Y are conditionally independent given Z: $X \perp\!\!\!\perp Y|Z$ if and only if:

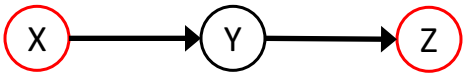
$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

Conditional Independence in a BN

Important question about a BN:

- Are two nodes independent *given certain evidence*?
- If yes, must prove using algebra (tedious in general)
- If no, can prove with a counter example (including CPTs)
- For example: P(Y|X) = 1 P(Z|Y) = .5
  P(Y|#X) = 1 P(Z|#Y) = .5

P(X) = .9

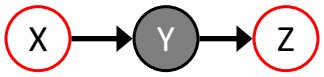


Note: one can change the CPTs so they ***are*** independent, but that's not enough for proof

- Question 1: are X and Z ***necessarily*** independent?
  - Answer: NO. Example: low pressure causes rain, which causes traffic.
  - Information about X may change our belief in Z (via Y)
  - Similarly, info about Z may change our belief in X

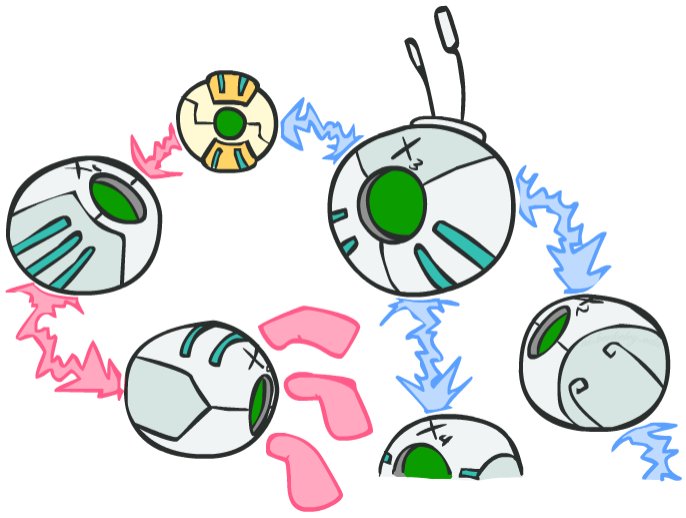
Conditional Independence in a BN

- 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
  - For example:



  - Question 1: are X and Z ***necessarily*** independent?
    - Answer: no.
  - Question 2: are X and Z ***conditionally independent***, given Y?
    - Answer: yes, as we'll see

D-separation



D-separation: Outline

- Study independence properties for subgraphs
  (connected triples)
- Analyze complex cases in terms of triples along paths between vars
- ***D-separation:*** a condition / algorithm for answering such queries

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?

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

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

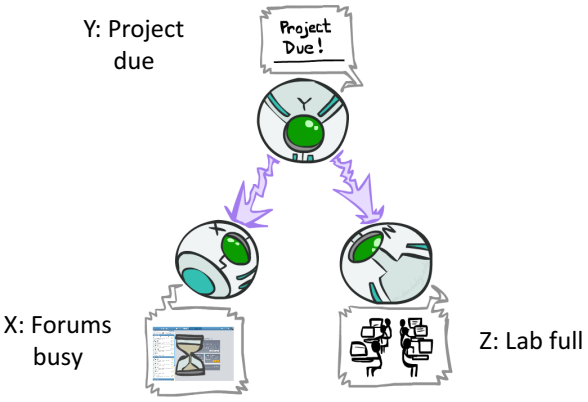
$$= P(z|y)$$

Yes!

- Evidence along the chain "blocks" the influence (makes "inactive")

Common Cause

- This configuration is a "common cause"



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

- Guaranteed X and Z independent given Y?

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

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

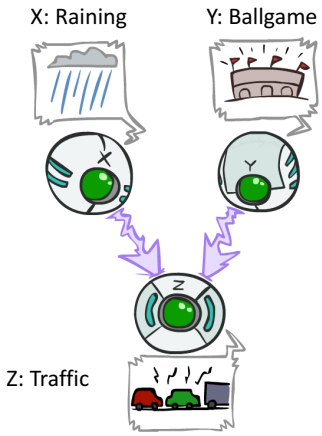
$$= P(z|y)$$

Yes!

- Observing the cause blocks influence between effects. (makes inactive)

Common Effect

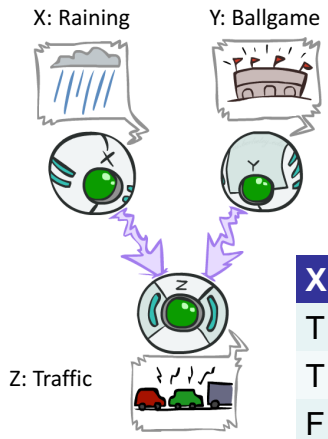
- Last configuration: two causes of one effect (v-structures)



- 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.
- This is ***backwards*** from the other cases
  - Observing an effect activates influence between possible causes. (makes active!)

Common Effect

P(X) = 0.8 P(Y) = 0.1

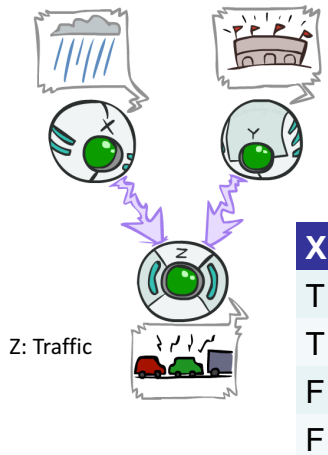


X	Y	P(Z)
T	T	0.95
T	F	0.8
F	T	0.9
F	F	0.5

X	Y	Z	P
T	T	T	0.076
T	T	F	0.004
T	F	T	0.576
T	F	F	0.144
F	T	T	0.162
F	T	F	0.002
F	F	T	0.090
F	F	F	0.009

Common Effect

P(X) = 0.8 P(Y) = 0.1

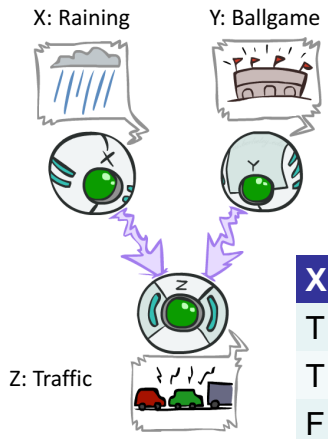


X	Y	P(Z)
T	T	0.95
T	F	0.8
F	T	0.9
F	F	0.5

X	Y	Z	P
T	T	T	0.076
T	T	F	0.004
T	F	T	0.576
T	F	F	0.144
F	T	T	0.162
F	T	F	0.002
F	F	T	0.090
F	F	F	0.009

Common Effect

P(X) = 0.8 P(Y) = 0.1



X	Y	P(Z)
T	T	0.95
T	F	0.8
F	T	0.9
F	F	0.5

X	Y	Z	P
T	T	T	0.076
T	T	F	0.004
T	F	T	0.576
T	F	F	0.144
F	T	T	0.018
F	T	F	0.002
F	F	T	0.090
F	F	F	0.009

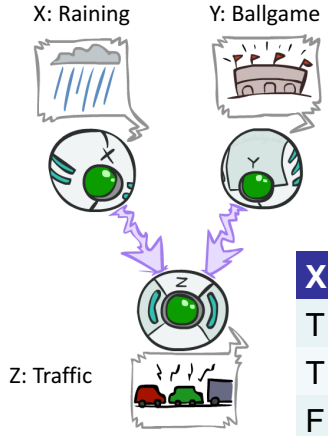
$$P(X|Y) = \frac{0.076+0.004}{0.076+0.004+0.018+0.002}$$

$$= 0.08 / 0.1$$

$$= 0.8$$

But Suppose Also Know Z=T

P(X) = 0.8 P(Y) = 0.1



$$P(X|Z) = \frac{.076+.576}{.076+.576+.018+.090}$$

$$= 0.652/0.76$$

$$= 0.858$$

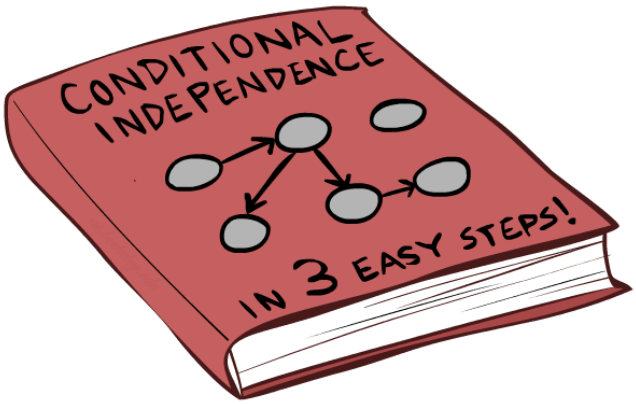
X	Y	P(Z)
T	T	0.95
T	F	0.8
F	T	0.9
F	F	0.5

X	Y	Z	P
T	T	T	0.076
T	T	F	0.004
T	F	T	0.576
T	F	F	0.144
F	T	T	0.018
F	T	F	0.002
F	F	T	0.090
F	F	F	0.009

$$P(X|Y,Z) = \frac{0.076}{0.076+ 0.018}$$

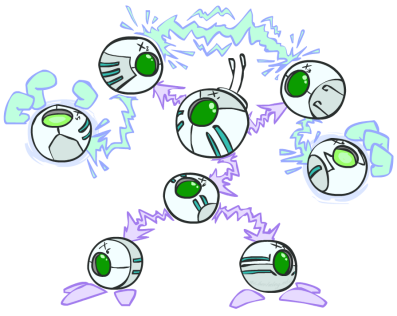
$$= 0.8085$$

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



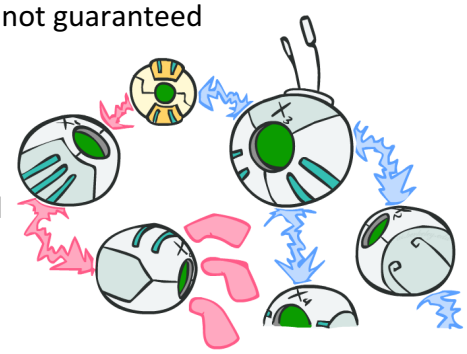
D-Separation

- 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 paths is active, then independence not guaranteed

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

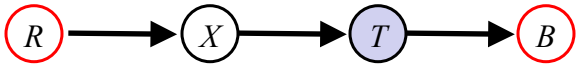
  - Otherwise (i.e. if ***all paths*** are inactive), then "D-separated" = independence ***is*** guaranteed

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



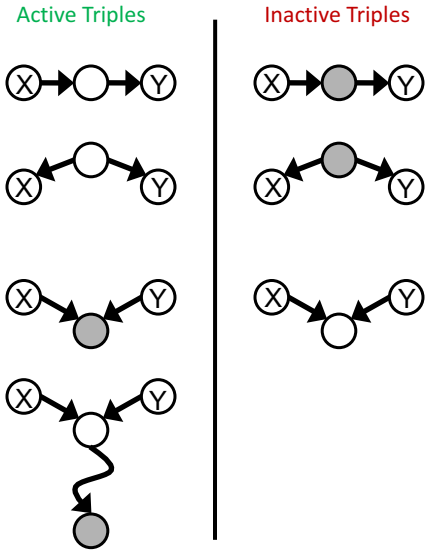
Example

- Query: $R \perp\!\!\!\perp B | T$



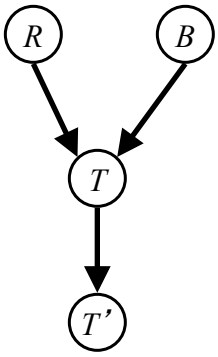
Active / Inactive Paths

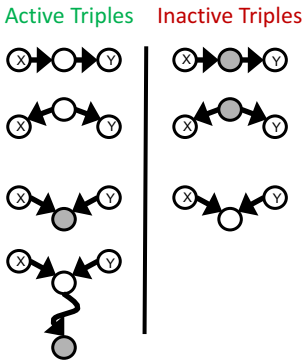
- Question: Are X and Y conditionally independent given evidence variables {Z}?
  - Yes, if X and Y "d-separated" by Z
  - Consider all (undirected) paths from X to Y
  - If no path is active → independence!
- A ***path*** is active if every triple in path is active:
  - Causal chain A → B → C where 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
  - (But ***all*** paths must be inactive)



Example

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

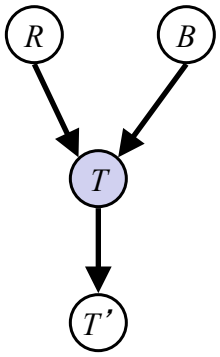


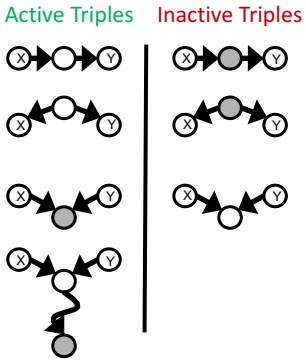


Example

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

$R \perp\!\!\!\perp B|T$ *No*



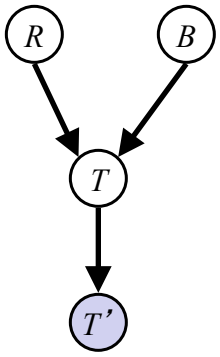


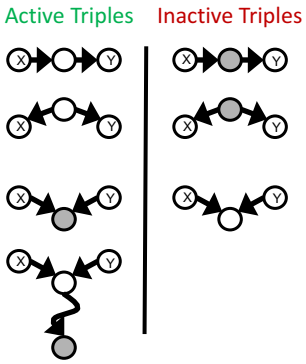
Example

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

$R \perp\!\!\!\perp B|T$ *No*

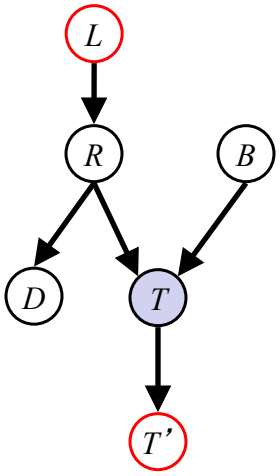
$R \perp\!\!\!\perp B|T'$ *No*

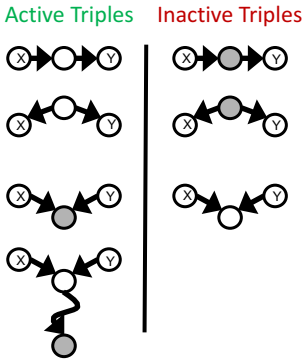




Example

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

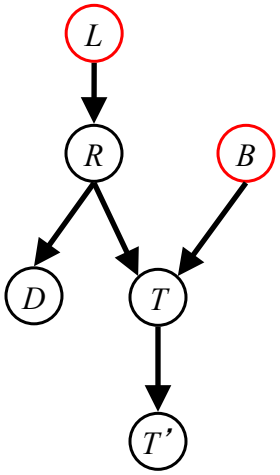


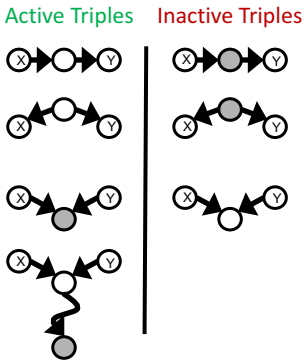


Example

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

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



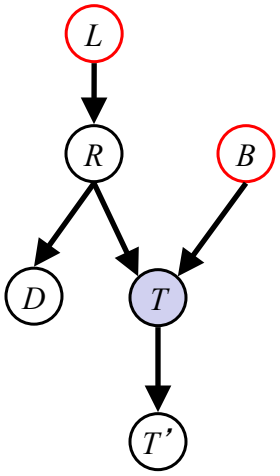


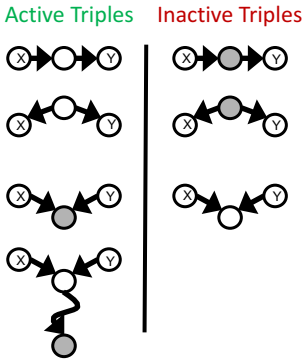
Example

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

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

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





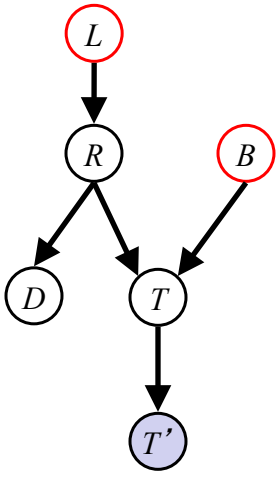
Example

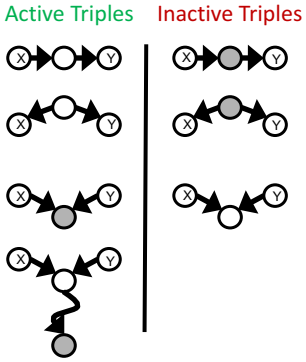
$L \perp\!\!\!\perp T' | T$ *Yes, Independent*

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

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

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





Example

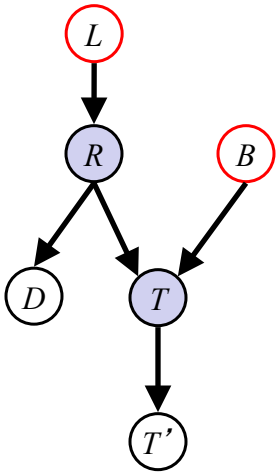
$L \perp\!\!\!\perp T' | T$ *Yes, Independent*

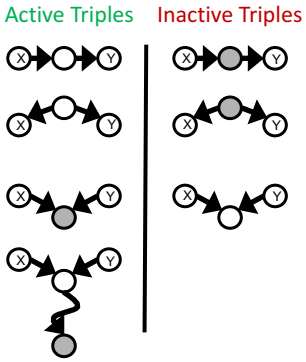
$L \perp\!\!\!\perp B$ *Yes, Independent*

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

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

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





Example

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

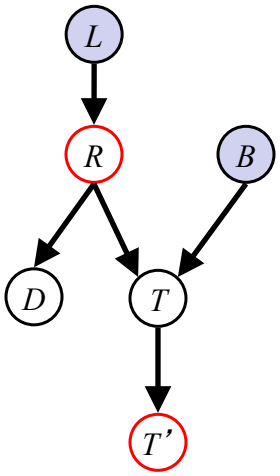
$L \perp\!\!\!\perp B$ *Yes, Independent*

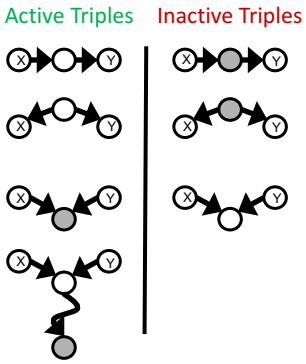
$L \perp\!\!\!\perp B | T$ *No*

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

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

$R \perp\!\!\!\perp T' | L, B$ *No*

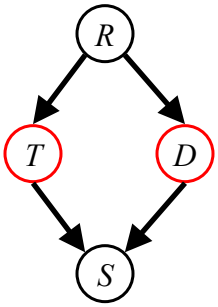


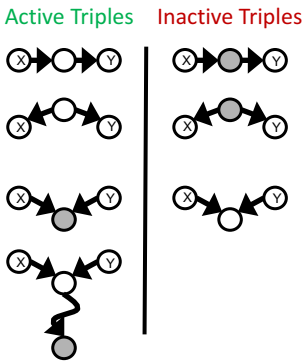


Example

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

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



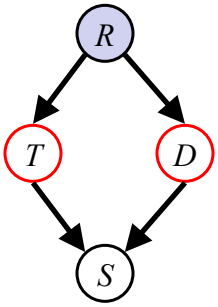


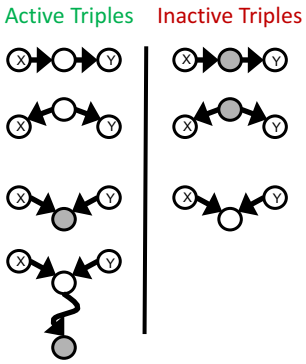
Example

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

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

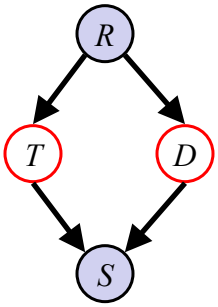
$T \perp\!\!\!\perp D|R$ *Yes, Independent*

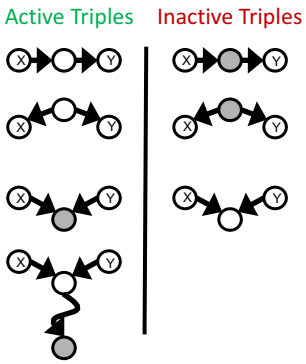




Example

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





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

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

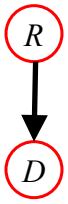
$T \perp\!\!\!\perp D | R, S$ *No*

Two Degenerate Cases

Unconnected → Always Independent
There ***are no*** paths between R and D, so definitely no active paths)

Two Degenerate Cases

Directly connected → ***De***pendent (for some values of the CPT)

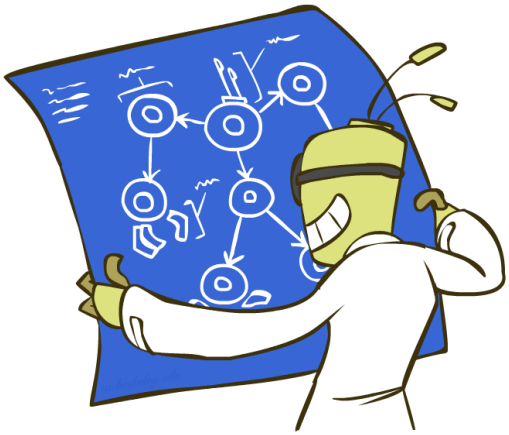


Structure Implications

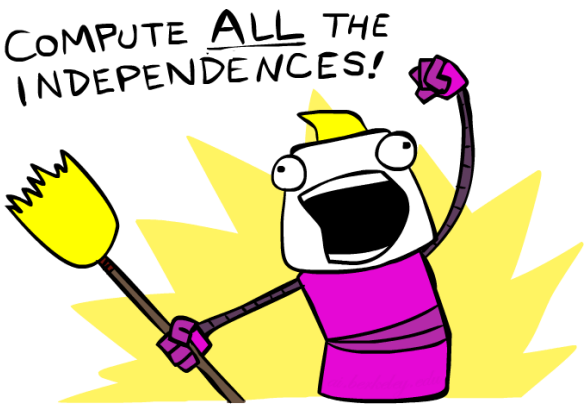
- Given a Bayes net structure, can run d-separation algorithm to build a complete list of conditional independences that are necessarily true of the form

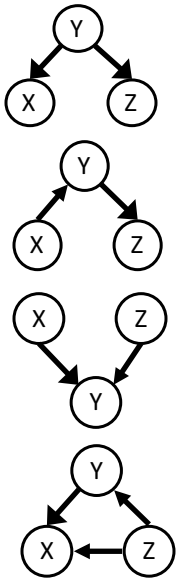
$$X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$$

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



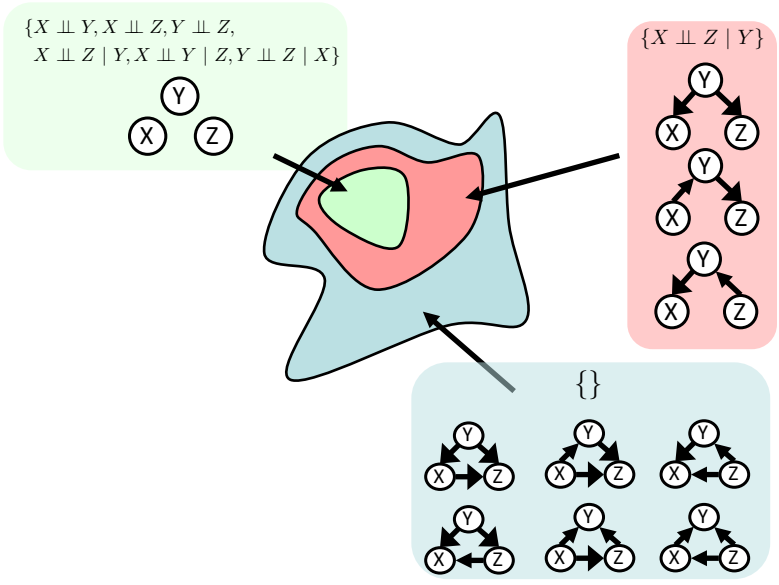
Computing All Independences





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