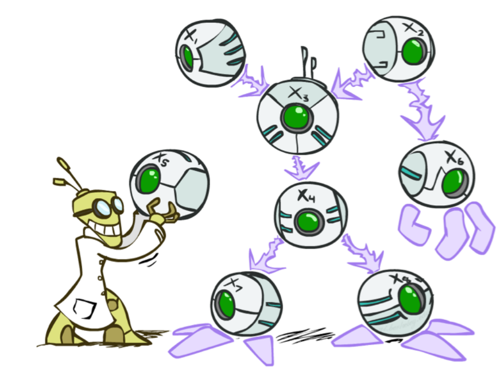
CSE P 573: Artificial Intelligence

Bayes Nets



Daniel Weld

[Most 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.]

Probabilistic Inference

- Probabilistic inference =

"compute a desired probability from other known probabilities (e.g. conditional from joint)"

- We generally compute conditional probabilities
  - P(on time | no reported accidents) = 0.90
  - These represent the agent's *beliefs* given the evidence

- Probabilities change with new evidence:
  - P(on time | no accidents, 5 a.m.) = 0.95
  - P(on time | no accidents, 5 a.m., raining) = 0.80
  - Observing new evidence causes *beliefs to be updated*

Inference by Enumeration

- General case:
  - Evidence variables: $E_1 \dots E_k = e_1 \dots e_k$
  - Query* variable: Q
  - Hidden variables: $H_1 \dots H_r$

$\left.\begin{array}{l} E_1 \dots E_k = e_1 \dots e_k \\ Q \\ H_1 \dots H_r \end{array}\right\} X_1, X_2, \dots X_n$

All variables

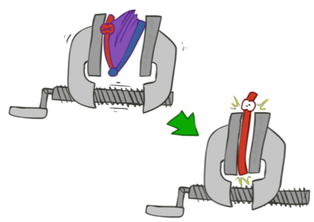
- We want:

** Works fine with multiple query variables, too*

$$P(Q|e_1 \dots e_k)$$

- Step 1: Select the entries consistent with the evidence

- Step 2: Sum out H to get joint distribution of query and evidence



$$P(Q, e_1 \dots e_k) = \sum_{h_1 \dots h_r} \underbrace{P(Q, h_1 \dots h_r, e_1 \dots e_k)}_{X_1, X_2, \dots X_n}$$

- Step 3: Normalize

$$\times \frac{1}{Z}$$

$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(Q|e_1 \cdots e_k) = \frac{1}{Z} P(Q, e_1 \cdots e_k)$$

Example: Inference by Enumeration

P(W=sun | S=winter)?

1. Select data consistent with evidence

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Example: Inference by Enumeration

P(W=sun | S=winter)?

1. Select data consistent with evidence
2. Marginalize away hidden variables (sum out temperature)

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Example: Inference by Enumeration

P(W=sun | S=winter)?

1. Select data consistent with evidence
2. Marginalize away hidden variables (sum out temperature)
3. Normalize

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

S	W	P
winter	sun	0.25
winter	rain	0.25

Example: Inference by Enumeration

P(W=sun | S=winter)?

1. Select data consistent with evidence
2. Marginalize away hidden variables (sum out temperature)
3. Normalize

S	T	W	P
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

S	W	P
winter	sun	0.50
winter	rain	0.50

Inference by Enumeration

- Computational problems?
  - Worst-case time complexity $O(d^n)$
  - Space complexity $O(d^n)$ to store the joint distribution

Don’t be Fooled

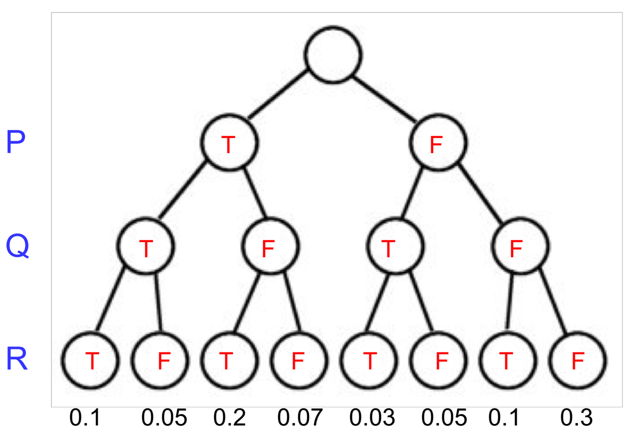
- It may look cute…

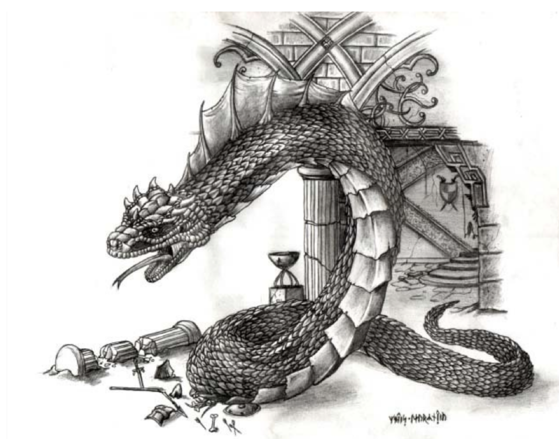
https://fc08.deviantart.net/fs71/i/2010/258/4/4/baby_dragon__charles_by_imsorrybuti-d2yti11.png

15

Don’t be Fooled

- It gets big…





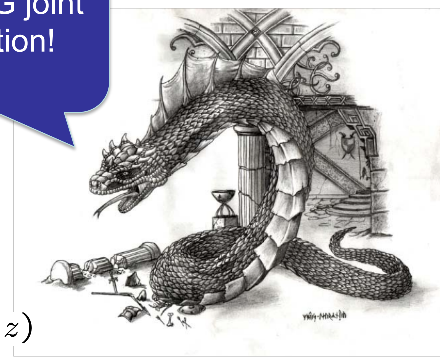
http://img4.wikia.nocookie.net/__cb20090430175407/monster/images/9/92/Basilisk.jpg

16

The Sword of Conditional Independence!

Slay
the
Basilisk!

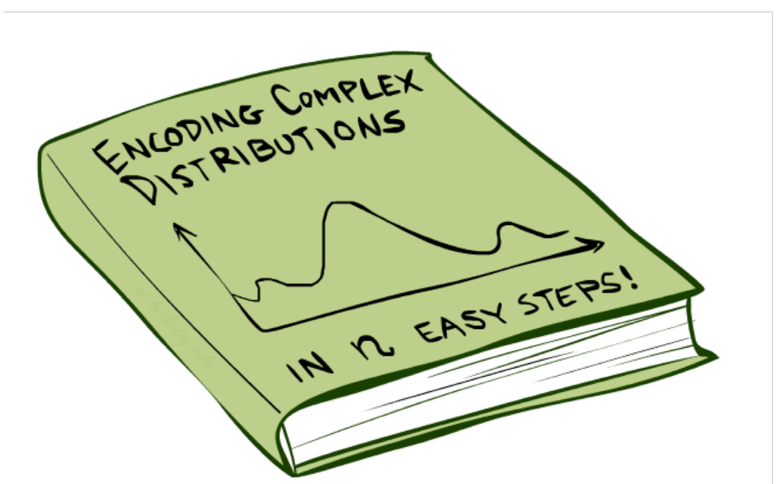
I am a BIG joint distribution!



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

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

Bayes'Nets: Big Picture

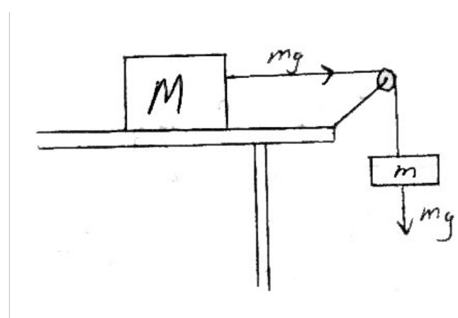


Bayes' Nets

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

Bayes Nets = a Kind of Probabilistic Graphical ***Model***

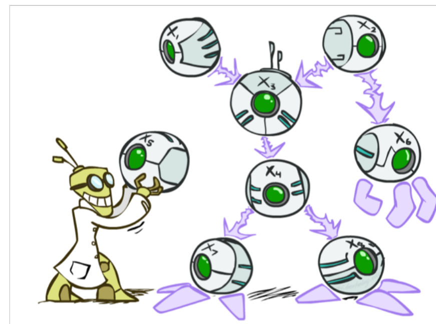
- Models describe how (a portion of) the world works
- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables
  - **"All models are wrong; but some are useful."**
    – George E. P. Box
- What do we do with probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
  - Example: value of information



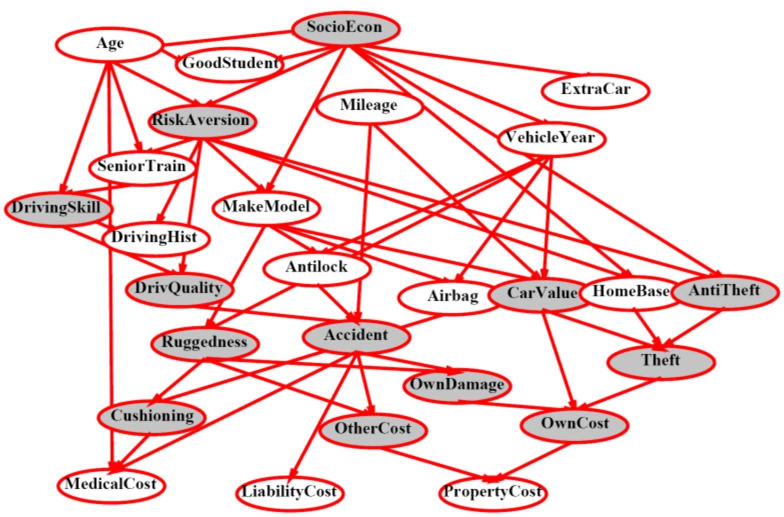
Friction,
Air friction,
Mass of pulley,
Inelastic string, …

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



Example Bayes' Net: Insurance



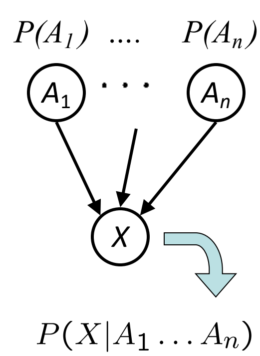
Bayes' Net Semantics

Bayes' Net Semantics

- A set of nodes, one per random variable
- A directed, ***acyclic*** graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents' values

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

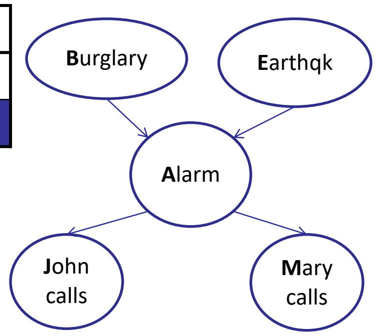
  - Called a "conditional probability table" (CPT)
  - May be thought of as a description of a noisy "causal" process



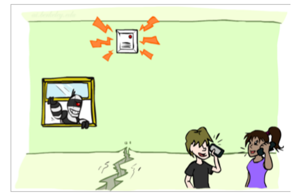
A Bayes net = Topology (graph) + Local Conditional Probabilities

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



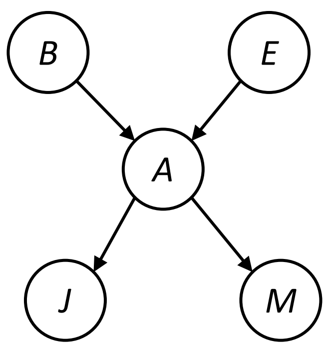
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

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

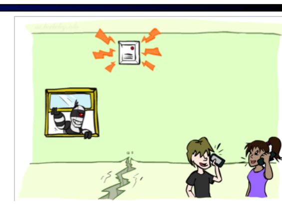
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

Bayes Nets Implicitly Encode Joint Distribution

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

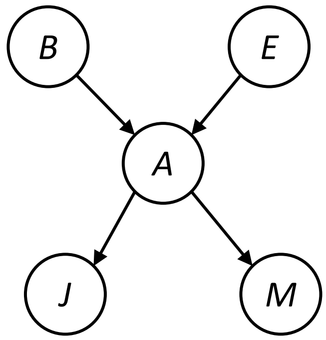
A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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

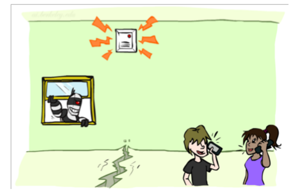
$$P(+b, -e, +a, -j, +m) =$$

Bayes Nets Implicitly Encode Joint Distribution

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

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

Joint Probabilities from BNs

- Why are we guaranteed that setting $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|\mathit{parents}(X_i))$ results in a proper joint distribution?
- Chain rule (valid for all distributions): $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|x_1 \ldots x_{i-1})$
- Assume conditional independences: $P(x_i|x_1, \ldots x_{i-1}) = P(x_i|\mathit{parents}(X_i))$

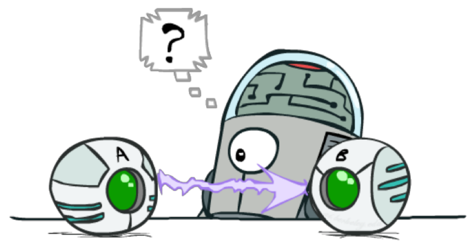
  → Consequence: $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|\mathit{parents}(X_i))$

- Every BN represents a joint distribution, but
- Not every distribution can be represented by a specific BN
  - The topology enforces certain conditional independencies

Causality?

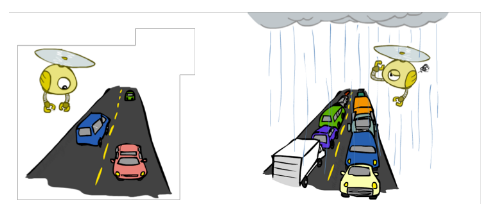
- When Bayes' nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - E.g. 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

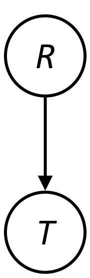
$$P(x_i|x_1, \ldots x_{i-1}) = P(x_i|\mathit{parents}(X_i))$$



What Causes Bad Traffic?

- Causal direction





$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

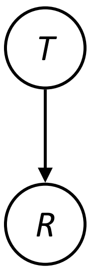
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$P(T,R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

Example: Reverse Traffic

- Reverse causality?



$P(T)$

+t	9/16
-t	7/16

$P(R|T)$

+t	+r	1/3
	-r	2/3
-t	+r	1/7
	-r	6/7

$P(T, R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

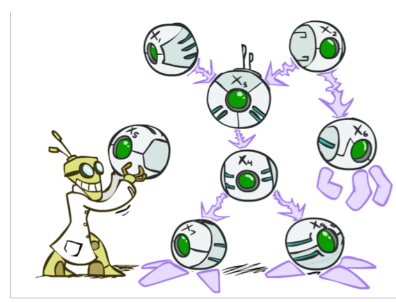
Summary: Bayes' Net Semantics

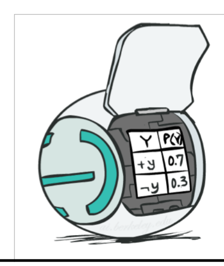
- A directed, acyclic graph, one node per random variable
- A conditional probability table (CPT) for each node
  - A collection of distributions over X, one for each combination of parents' values

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

- Bayes' nets ***compactly*** 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:

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





Size of a Bayes' Net

- How big is a joint distribution over N Boolean variables?

  2^N

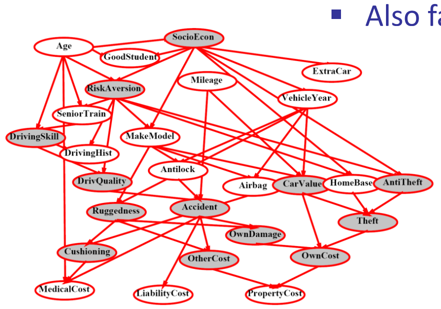
- How big is an N-node Bayes net if nodes have up to k parents?

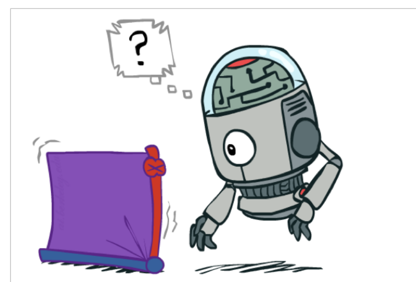
  $O(N * 2^k)$

- Both give you the power to calculate

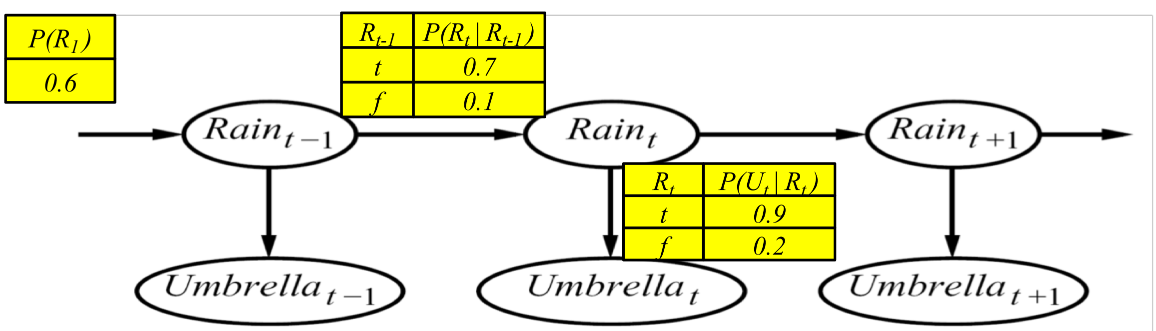
  $$P(X_1, X_2, \ldots X_n)$$

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





Hidden Markov Model: Example

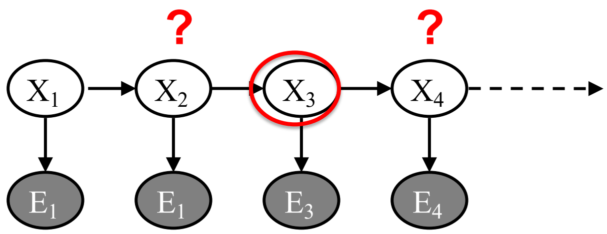


- An HMM is defined by:
  - Initial distribution: $P(X_1)$
  - Transitions: $P(X_t | X_{t-1})$
  - Emissions: $P(E|X)$

Conditional Independence

HMMs have two important independence properties:

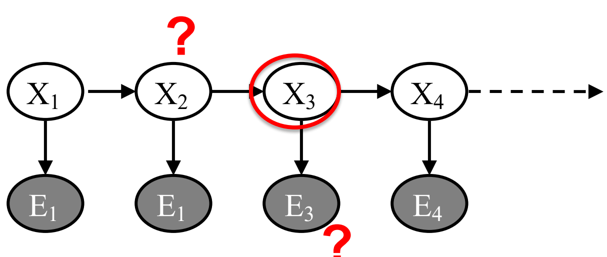
- **Future independent of past given the present**



Conditional Independence

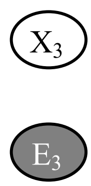
HMMs have two important independence properties:

- Future independent of past given the present
- **Current observation independent of all else given current state**



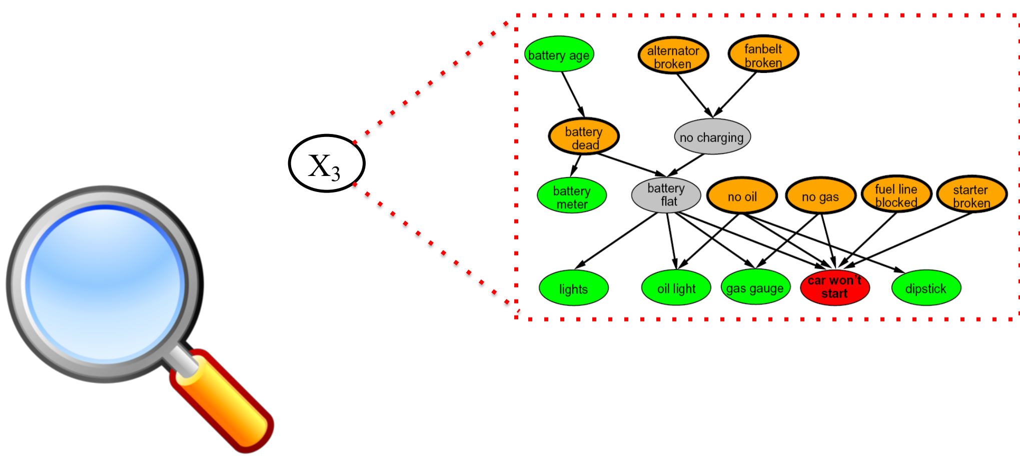
What about Conditional Independence *in Snapshot*

- Can we do something here?
- Factor X into product of (conditionally) independent random vars?

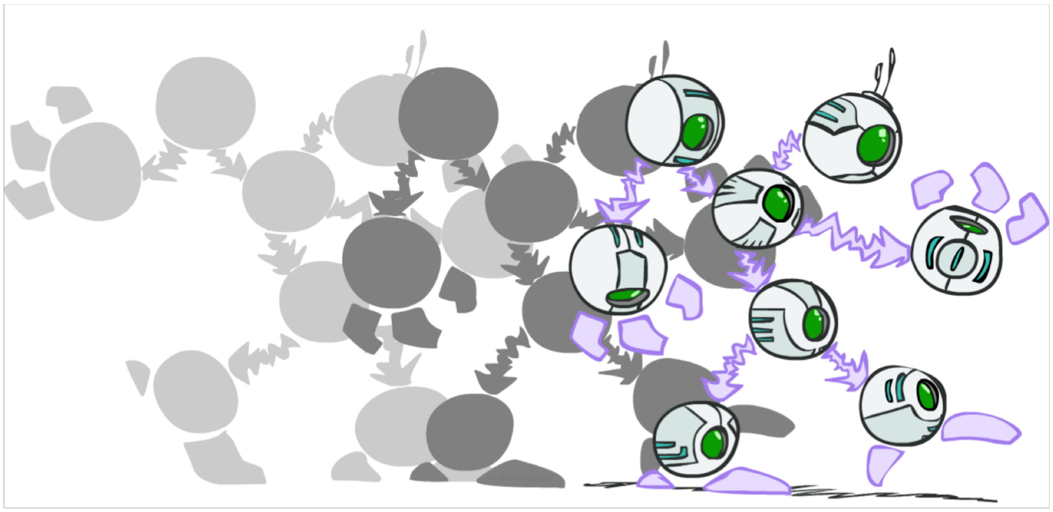


- Maybe also factor E

Yes! with Bayes Nets

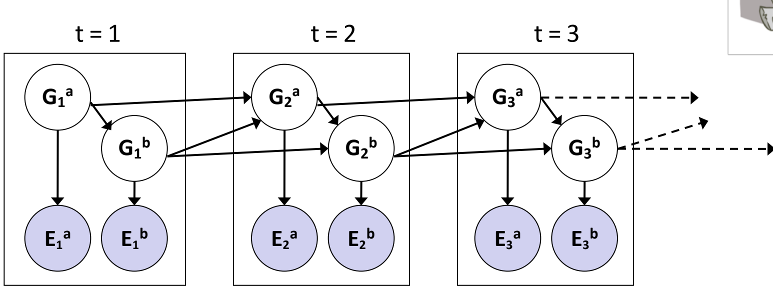


Dynamic Bayes Nets



Dynamic Bayes Nets (DBNs)

- We want to track multiple variables over time, using multiple sources of evidence
- Idea: Repeat a fixed Bayes net structure at each time
- Variables from time *t* can condition on those from *t-1*

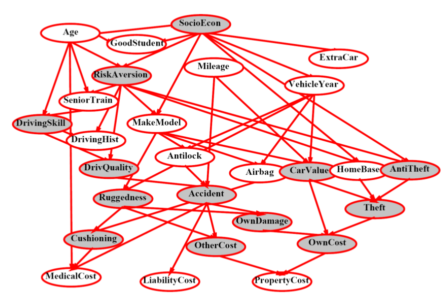


- Dynamic Bayes nets are a generalization of HMMs

DBN Particle Filters

- A particle is a complete sample for a time step
- **Initialize**: Generate prior samples for the t=1 Bayes net
  - Example particle: $\mathbf{G_1^a} = (3,3)$ $\mathbf{G_1^b} = (5,3)$
- **Elapse time**: Sample a successor for each particle
  - Example successor: $\mathbf{G_2^a} = (2,3)$ $\mathbf{G_2^b} = (6,3)$
- **Observe**: Weight each *entire* sample by the likelihood of the evidence conditioned on the sample
  - Likelihood: $P(\mathbf{E_1^a} \mid \mathbf{G_1^a}) * P(\mathbf{E_1^b} \mid \mathbf{G_1^b})$
- **Resample:** Select prior samples (tuples of values) in proportion to their likelihood

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 $P(X \mid e)$?
- Learning: Given data, what is best BN encoding?

Remember....

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

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

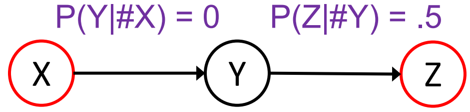
- X and Y are conditionally independent given Z: $X \perp\!\!\!\perp Y|Z$
  if and only if:

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

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) = .9
  P(Y|#X) = 0 P(Z|#Y) = .5



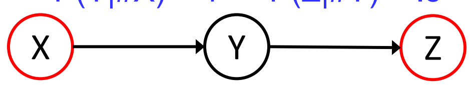
- Question 1: are X and Z ***necessarily*** independent?
  - Answer: NO. Example: low pressure causes rain, which causes traffic.
  - Information about X may change our belief in Z (via Y)
  - Similarly, info about Z may change our belief in X

Conditional Independence in a BN

Important question about a BN:

- Are two nodes independent *given certain evidence*?
- If yes, 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

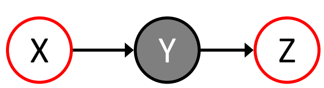


Note: one can change the CPTs so they ***are*** independent, but that's not enough for proof

- 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

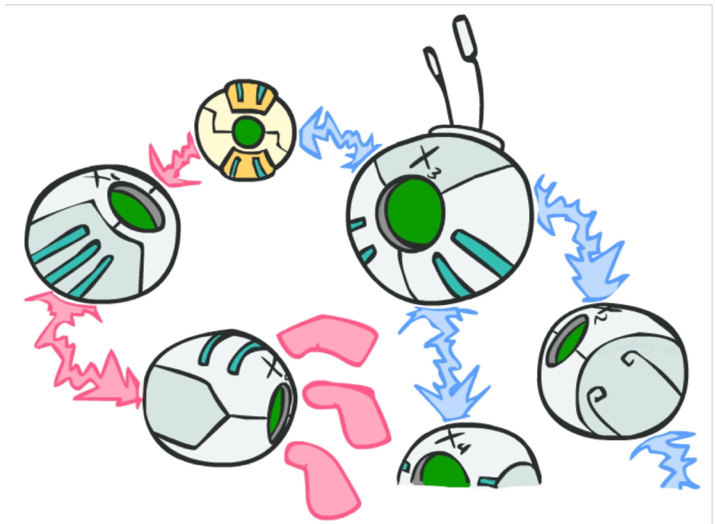
Conditional Independence in a BN

- 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 I asserted with HMMs) … we'll see why shortly

D-separation

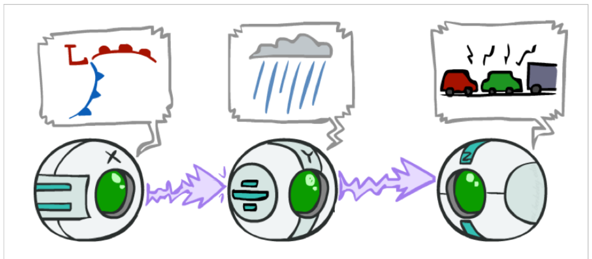


D-separation: Outline

- 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?

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

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

$$= P(z|y)$$

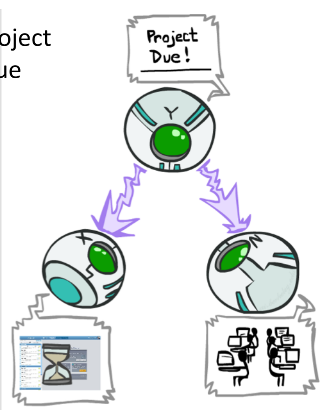
Yes!

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

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?

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

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

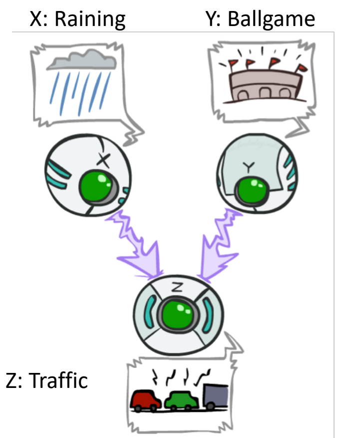
$$= P(z|y)$$

Yes!

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

Common Effect

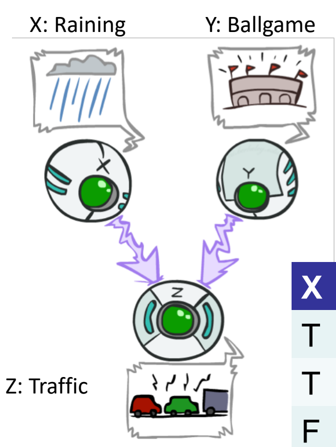
- Last configuration: two causes of one effect (v-structures)



- 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

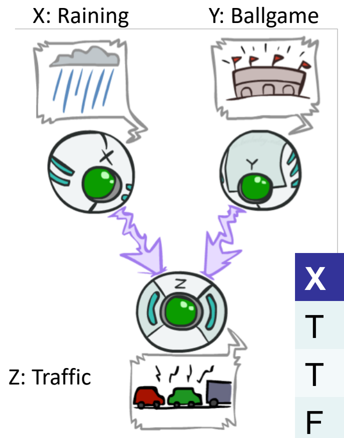


X	Y	P(Z)
T	T	0.95
T	F	0.8
F	T	0.9
F	F	0.5

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

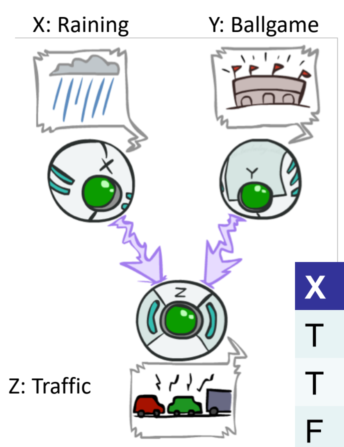


X	Y	P(Z)
T	T	0.95
T	F	0.8
F	T	0.9
F	F	0.5

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

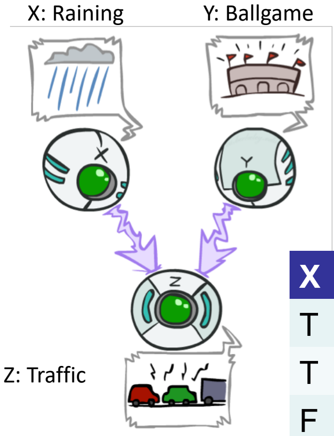
$$P(X|Y) = \frac{0.076+0.004}{0.076+0.004+0.018+0.002}$$

$$= 0.08 / 0.1$$

$$= 0.8$$

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

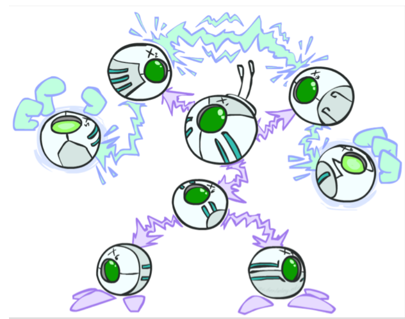
X	Y	Z	P
T	T	T	0.076
T	T	F	0.004
T	F	T	0.576
T	F	F	0.144
F	T	T	0.018
F	T	F	0.002
F	F	T	0.090
F	F	F	0.009

$$P(X|Y,Z) = \frac{0.076}{0.076+0.018}$$
$= 0.8085$
$\neq P(X|Y) = 0.8$

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



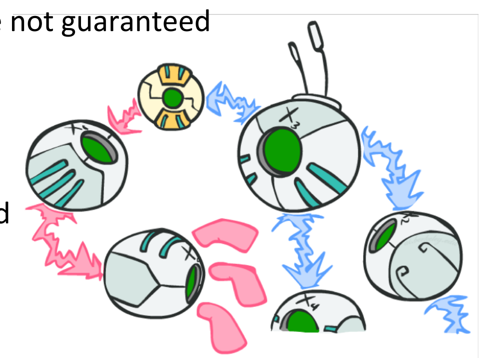
D-Separation

- 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

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

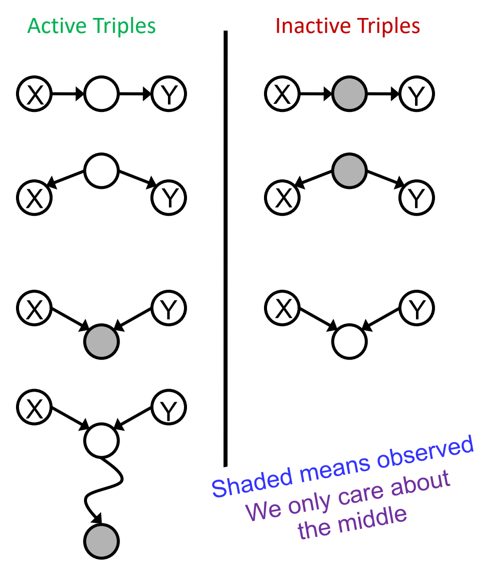
  - Otherwise (i.e. if ***all paths*** are inactive), then “D-separated” = independence ***is*** guaranteed

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



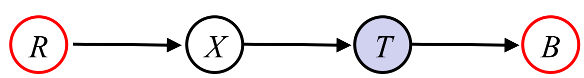
Active / Inactive Paths

- 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 $A \to B \to C$ where B is unobserved (either direction)
  - Common cause $A \leftarrow B \to C$ where B is unobserved
  - Common effect (aka v-structure)
    $A \to B \leftarrow C$ 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)



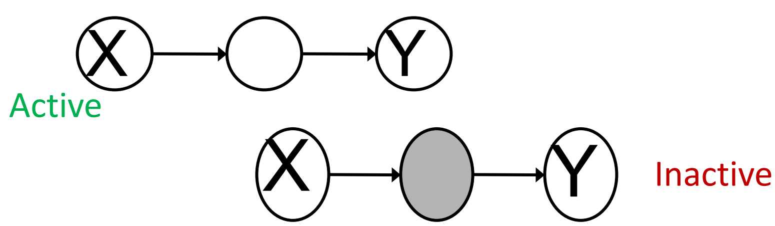
Example

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



- How many paths?
- How many triples?

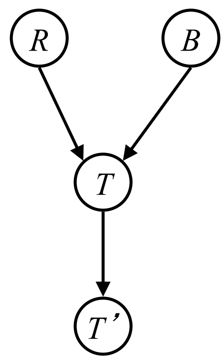
Are they both active?

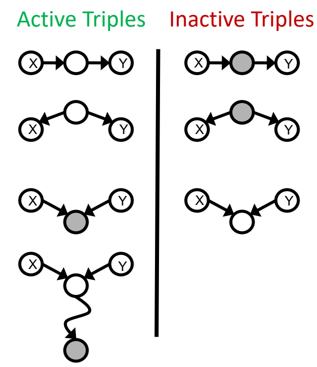


- Answer: Independent. T makes the whole path inactive

Example

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

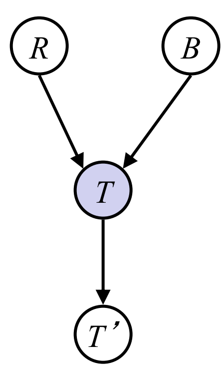


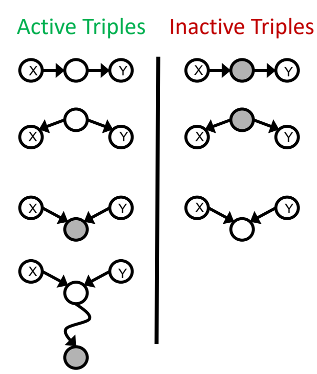


Example

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

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



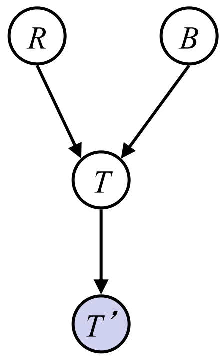


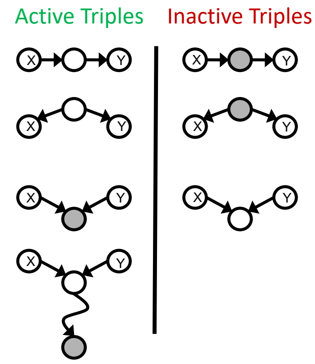
Example

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

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

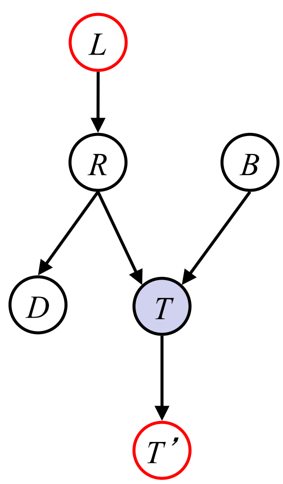
$R \perp\!\!\!\perp B | T'$ *No*

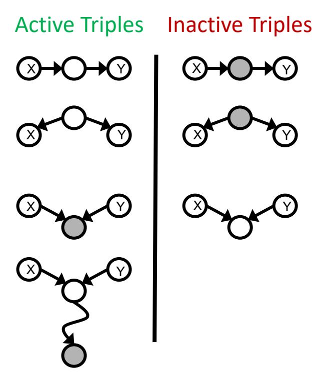




Example

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

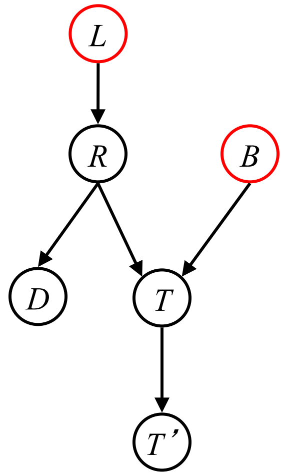


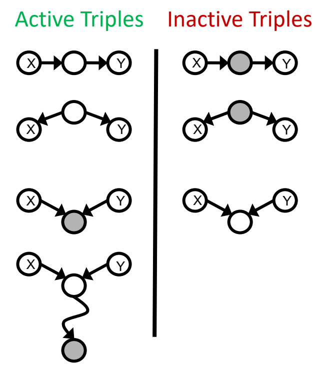


Example

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

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



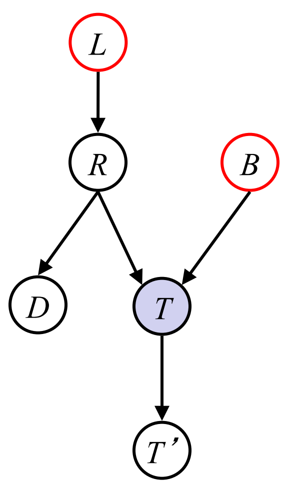


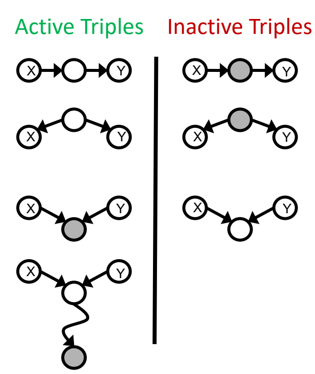
Example

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

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

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





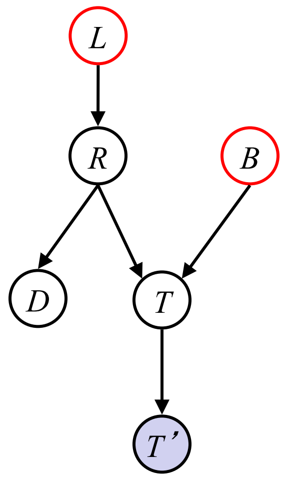
Example

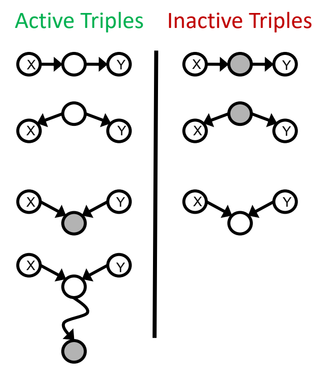
$L \perp\!\!\!\perp T' | T$ *Yes, Independent*

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

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

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





Example

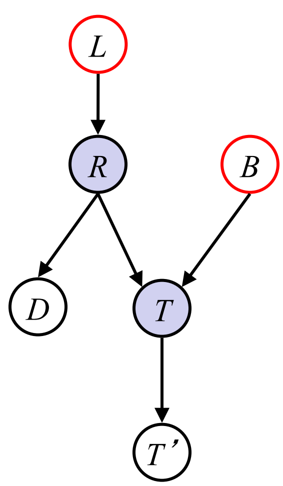
$L \perp\!\!\!\perp T' | T$ *Yes, Independent*

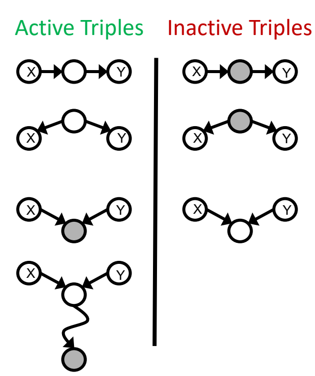
$L \perp\!\!\!\perp B$ *Yes, Independent*

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

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

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





Example

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

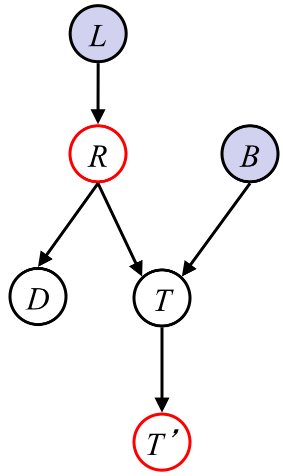
$L \perp\!\!\!\perp B$ — *Yes, Independent*

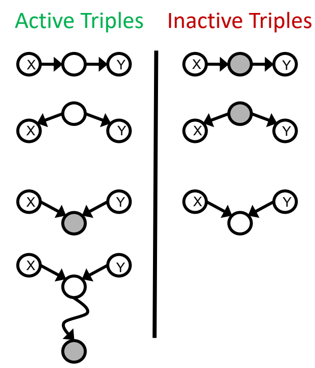
$L \perp\!\!\!\perp B | T$ — *No*

$L \perp\!\!\!\perp B | T'$ — *No*

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

$R \perp\!\!\!\perp T' | L, B$ — *No*

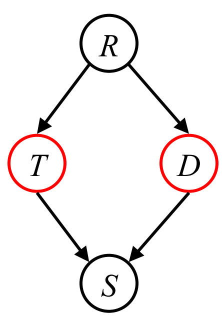


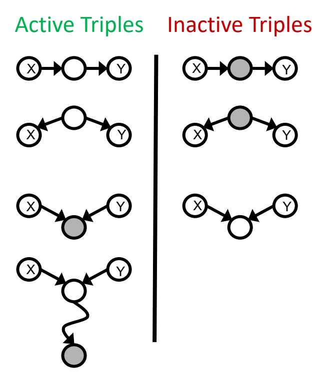


Example

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

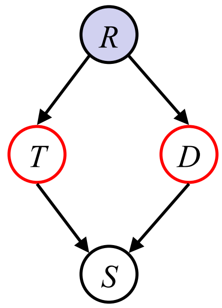
$T \perp\!\!\!\perp D$ — *No*

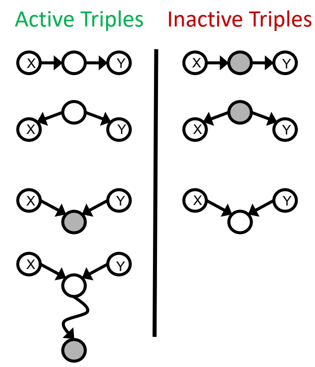




Example

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



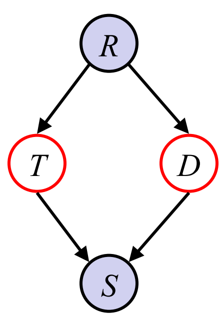


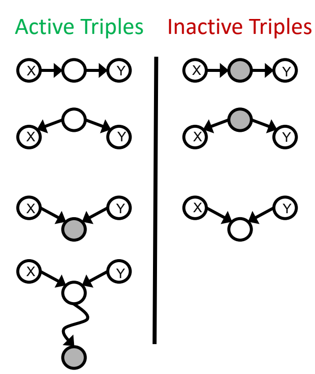
$T \perp\!\!\!\perp D$ *No*

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

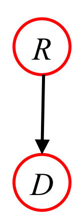
$T \perp\!\!\!\perp D | R, S$ *No*

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

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