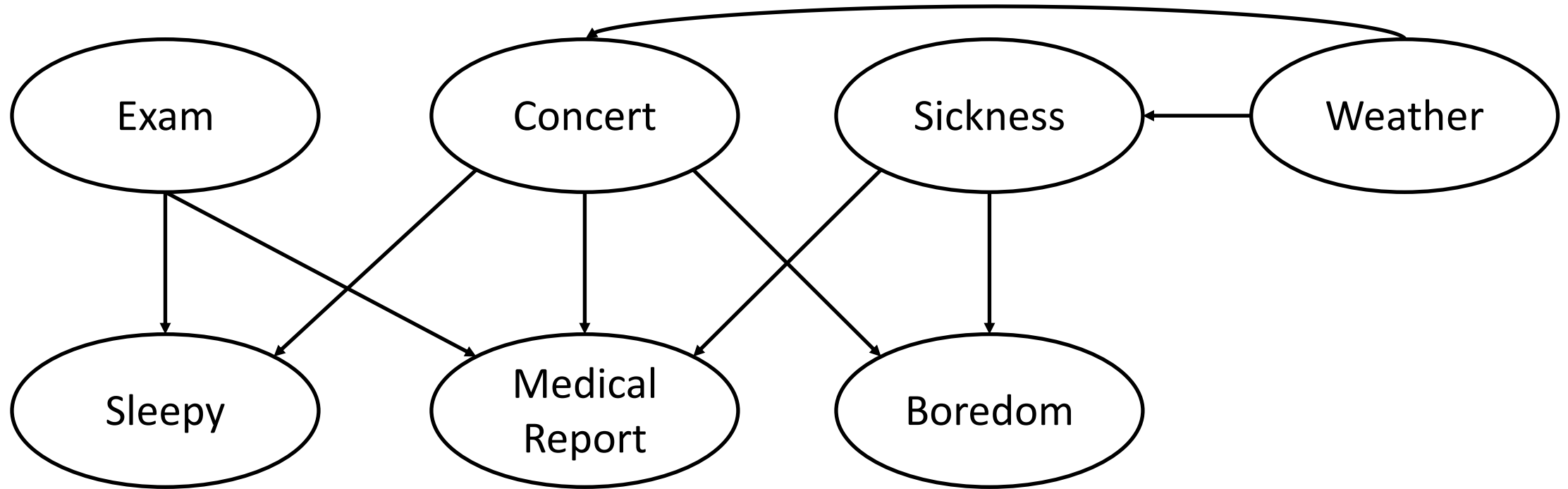


COMP 341 Intro to AI

Bayesian Networks - Representation



Asst. Prof. Barış Akgün

Koç University

Probabilistic Models

- 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 agents do with probabilistic models?
 - Reason about unknown variables, given evidence
 - Example: explanation (diagnostic reasoning)
 - Example: prediction (causal reasoning)
 - Example: value of information



Probabilistic Models

- A probabilistic model is a joint distribution over a set of variables

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

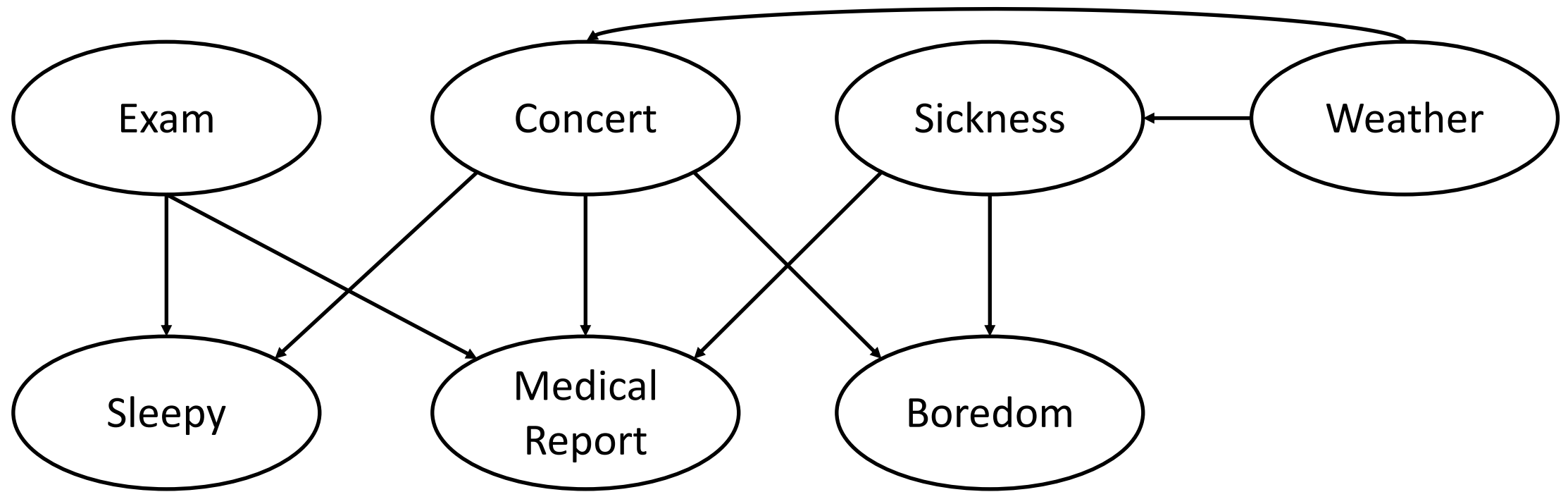
- Posterior probabilities are used to reason about the world, ask queries etc.

$$P(X_q |, x_{e1}, \dots, x_{ek})$$

Bayesian Networks

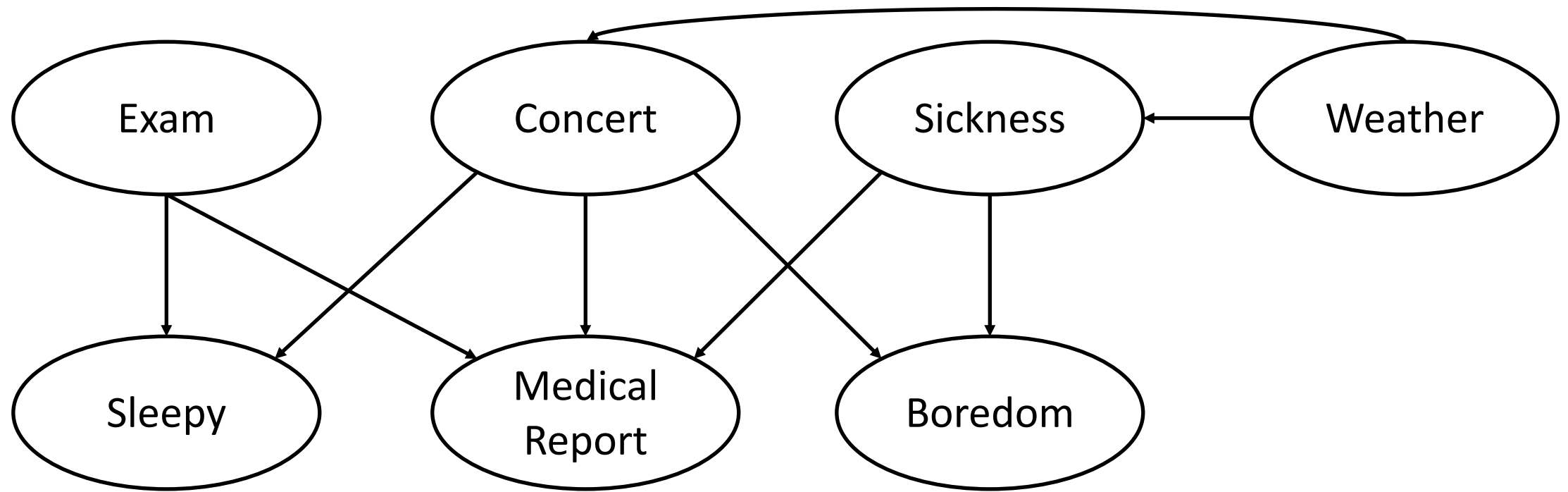
- Two problems with using full joint distribution tables as probabilistic models:
 - Too big to represent($O(d^n)$)
 - Hard to specify
 - Hard to learn anything empirically
- **Bayes' nets:** a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
 - A variety of **graphical models**
 - We describe how variables interact locally
 - Local interactions chain together to give global, indirect interactions

Bayesian Network Example



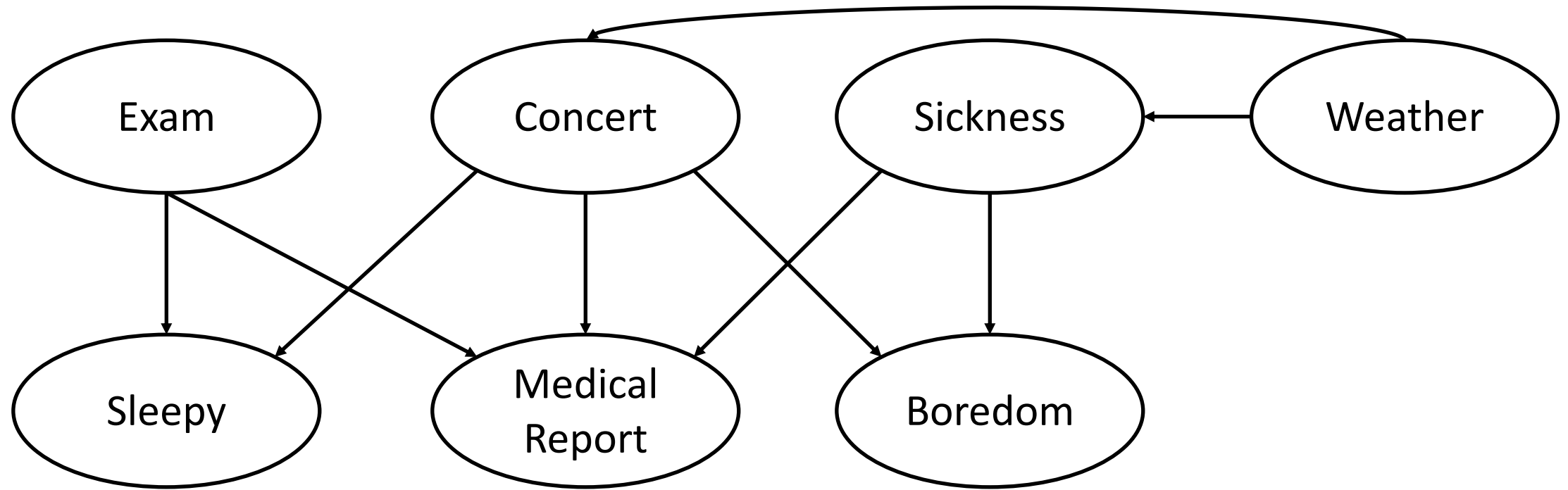
- If Mehmet is sick or there is an exam or there is a concert, he is likely to get a medical report
- If the weather is bad, Mehmet is likely to get sick. The weather also affects the concert
- Mehmet gets bored if he is sick or misses the concert
- Mehmet is sleepy the next day if he studies for an exam or goes to the concert

Bayesian Network Example



- What is the reason that Mehmet had a Medical Report?
- The weather was bad
- There was an exam
- Mehmet was sleepy
- Mehmet was not bored
- TODO: Tell Mehmet he will not be getting a Make Up!

Bayesian Network Example

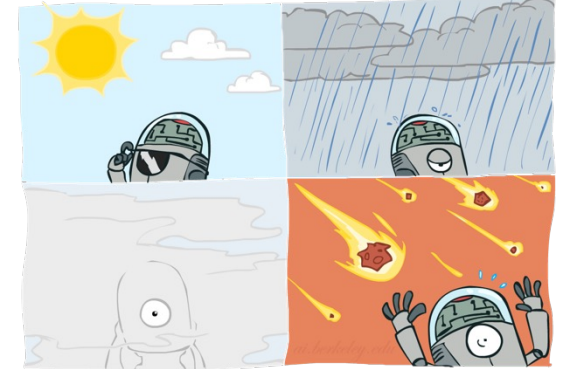


$P(\text{Exam}, \text{Concert}, \text{Sickness}, \text{Weather}, \text{Sleepy}, \text{MedicalReport}, \text{Boredom}) = ?$

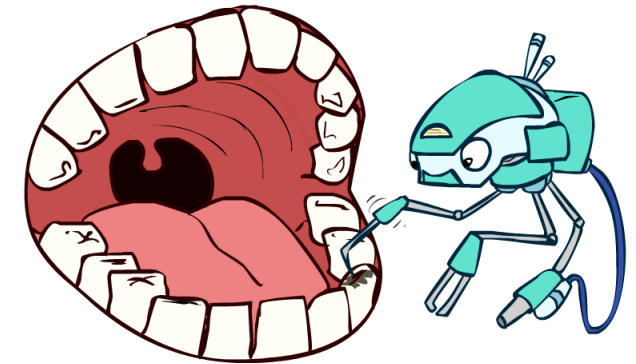
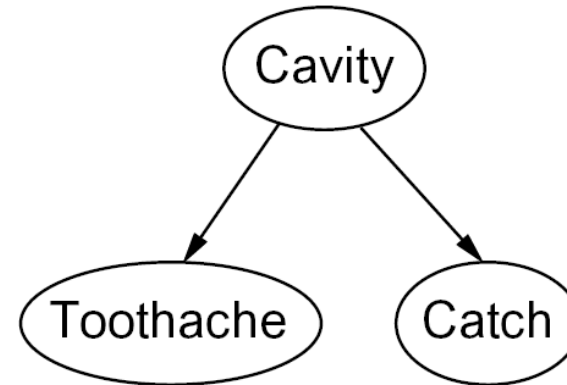
- Answer queries, e.g., about Mehmet getting a medical report
- Use the full joint distribution
- Simplify by using the chain rule and conditional independence!

Bayesian Network Semantics

- Nodes: variables (with domains)
 - Can be assigned (observed) or unassigned (unobserved)



- Arcs: interactions
 - Directional
 - Indicate “direct influence” between variables
 - Formally: encode **conditional independence**

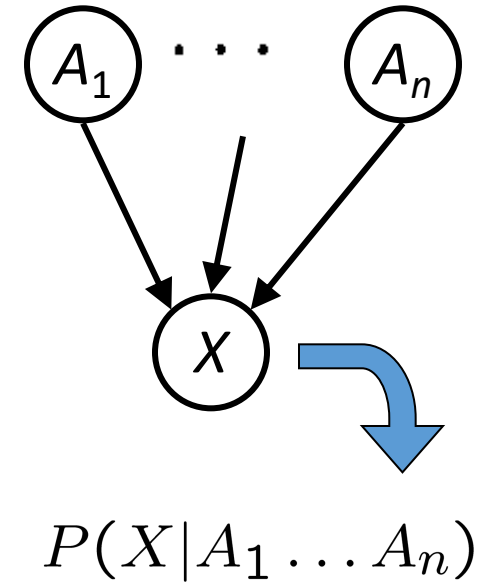


Bayesian Network Semantics

- Represented as a directed acyclic graphs
- A set of nodes, one per variable X
- A conditional distribution for each node
 - A conditional distribution for each node given its parents

$$P(X_i | Parents(X_i))$$

- E.g. as a conditional probability table (CPT)
- Description of a noisy “causal” process



Bayes net = Graph Topology (graph) + CPTs

Probabilities in Bayesian Networks

- Bayesian Networks encode joint distributions **implicitly**
- ... as a product of conditional distributions
- To get the probability of a full assignment, multiply all the conditionals:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Parents}(X_i))$$

- E.g.

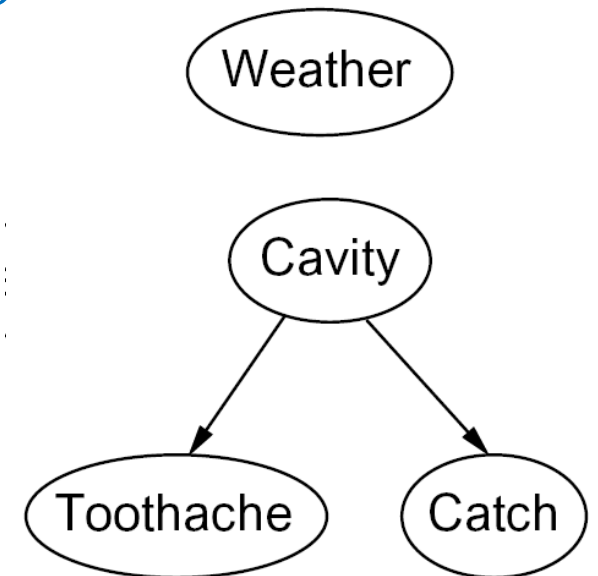
$P(+cavity, +catch, -toothache, sunny)$

$= P(+cavity) \times$

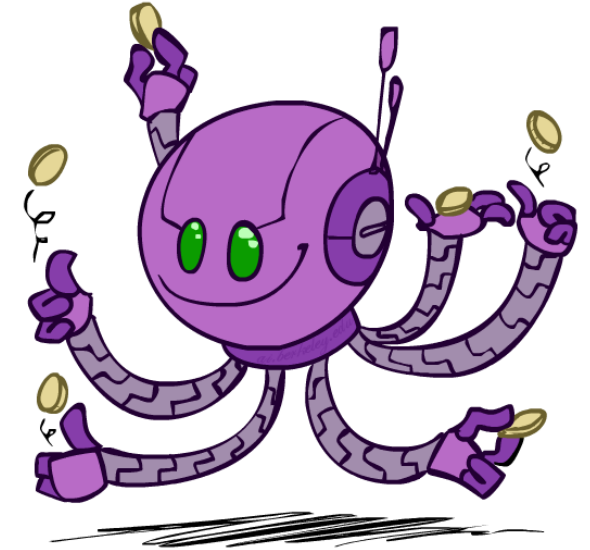
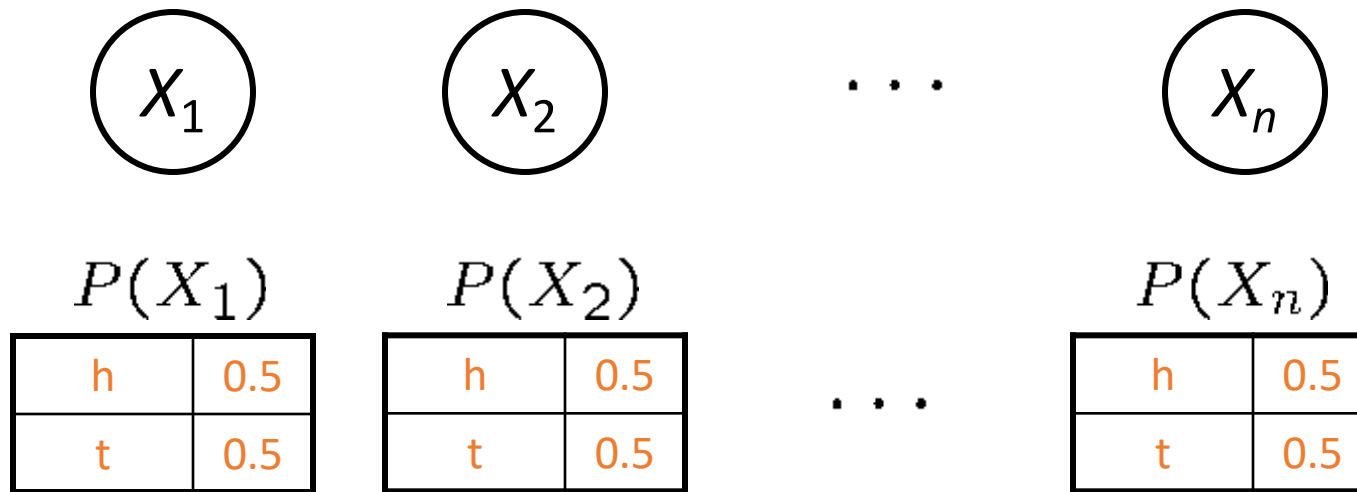
$P(-toothache | +cavity) \times$

$P(+catch | +cavity) \times$

$P(sunny)$



Example: Coin Flips

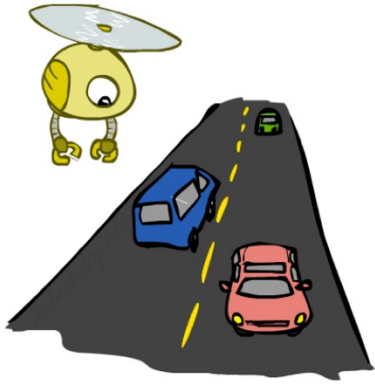
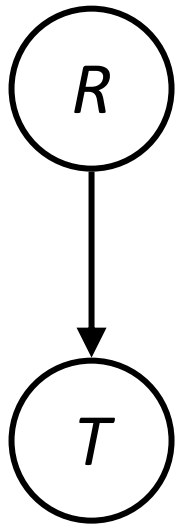


$$P(h,t,t,h,t,h) = ??$$

Only distributions whose variables are absolutely independent can be represented by a Bayes' net with no arcs.

Example: Traffic

- Rain affects traffic



$$P(T, R) = P(T|R) P(R)$$

$P(R)$	
+r	1/4
-r	3/4

$P(T R)$					
+r	<table><tr><td>+t</td><td>3/4</td></tr><tr><td>-t</td><td>1/4</td></tr></table>	+t	3/4	-t	1/4
+t	3/4				
-t	1/4				
-r	<table><tr><td>+t</td><td>1/2</td></tr><tr><td>-t</td><td>1/2</td></tr></table>	+t	1/2	-t	1/2
+t	1/2				
-t	1/2				

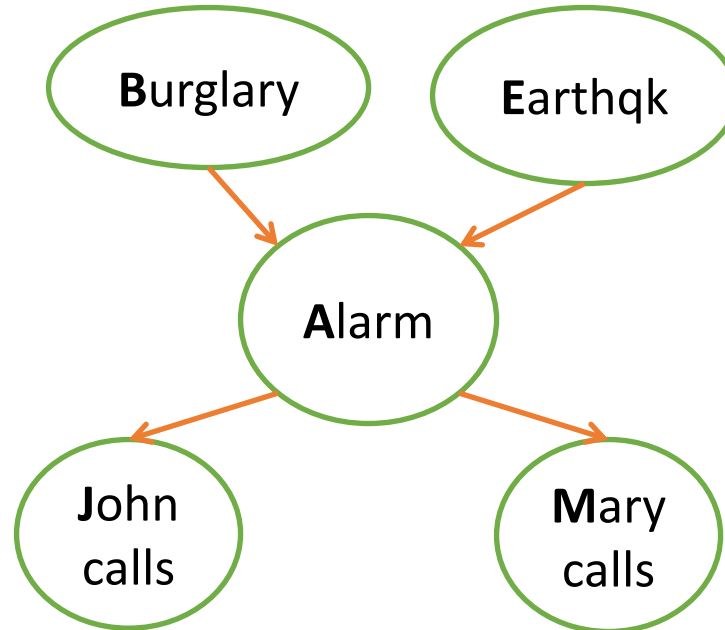
$$P(-t, +r) = ??$$

Example: House Alarm

- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?
- Variables:
 - *Burglar, Earthquake, Alarm, JohnCalls, MaryCalls*
- A burglar can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Mary to call
- The alarm can cause John to call
- What does the Bayes Net look like?

Example: House Alarm

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

Constructing Bayesian networks

- 1. Choose an ordering of variables X_1, X_2, \dots, X_n
- 2. For $i = 1$ to n
 - add X_i to the network
 - select parents from X_1, X_2, \dots, X_{i-1} such that:

$$P(X_i | Parents(X_i)) = P(X_i | X_1, X_2, \dots, X_{i-1})$$

- This choice of parents guarantees:

$$\begin{aligned} P(X_1, X_2, \dots, X_n) &= \prod_{i=1}^n P(X_i | X_1, X_2, \dots, X_{i-1}) \text{ (chain rule)} \\ &= \prod_{i=1}^n P(X_i | Parents(X_i)) \text{ (by construction)} \end{aligned}$$

- Ordering matters!
- Recall that a causal structure is often simpler (rule of thumb!)
- This is just a way to build valid networks and is not too useful in practice (unless you have a good ordering and prune after the construction)

Example

- Suppose we choose the ordering M, J, A, B, E

$P(J \mid M) = P(J)$? **No**

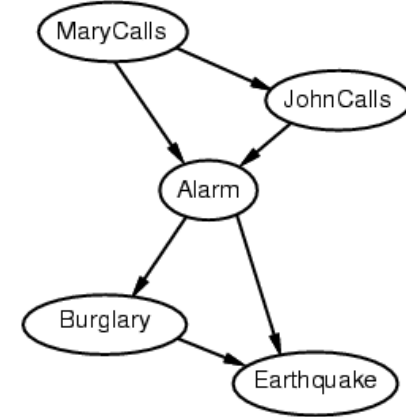
$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$? **No**

$P(B \mid A, J, M) = P(B \mid A)$? **Yes**

$P(B \mid A, J, M) = P(B)$? **No**

$P(E \mid B, A, J, M) = P(E \mid A)$? **No**

$P(E \mid B, A, J, M) = P(E \mid A, B)$? **Yes**

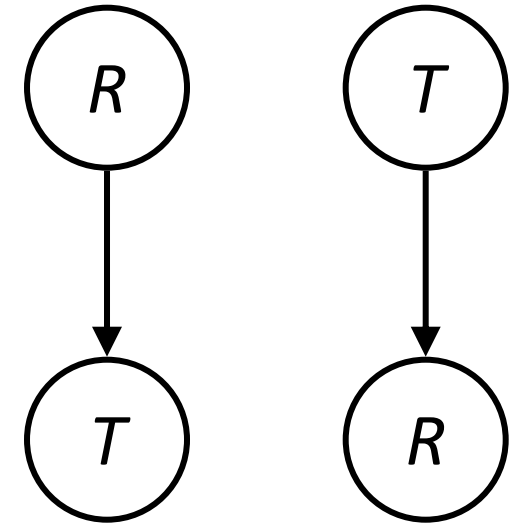


Size of a Bayes Net

- Joint dist. table size for n variables with largest domain size d :
 $O(d^n)$
- Size of a Bayes Net. with n nodes where each variable has no more than k parents:
 $O(n \cdot d^{k+1})$
- If $k \ll n$, then a huge difference! (Exponential to linear)
- For burglary net, $1 + 1 + 4 + 2 + 2 = 10$ numbers (vs. $2^5 - 1 = 31$)

Causality

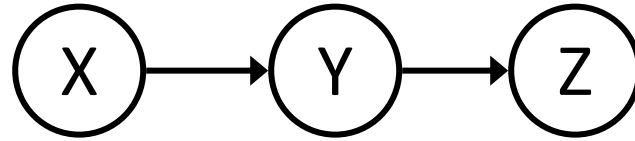
- Do BNs always represent causality?
 - No!
- Both are valid
- Topology encodes *conditional independence*, not causality
- Same joint distribution can be encoded by many different BNs
- When Bayes' nets reflect the true causal patterns:
 - Often simpler (nodes have fewer parents)
 - Often easier to think about
 - Often easier to elicit from experts



Independence in a BN

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

- Are two nodes independent given certain evidence?



- Question: are X and Z necessarily independent?

- **No!**
- X can influence Z, Z can influence X (via Y)
- E.g. low pressure → rain → traffic
- They *could* be independent: how?

Counter Example:

$$P(+y \mid +x) = 1, P(-y \mid -x) = 1$$

$$P(+z \mid +y) = 1, P(-z \mid -y) = 1$$

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

$$P(X, Z) = P(X) \neq P(X)P(Z) \text{ (in general)}$$

$$P(X, Y, Z) = P(X)P(Y|X)P(Z|Y) \text{ (from the BN)}$$

$$P(X, Z) = \sum_i P(X)P(Y = y_i|X)P(Z|Y = y_i) \text{ (marginalize)}$$

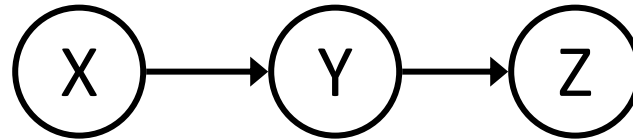
$$P(X, Z) = P(X) \sum_i P(Y = y_i|X)P(Z|Y = y_i) \stackrel{?}{=} P(X)P(Z) \text{ (defn. independence)}$$

$$P(Z) \stackrel{?}{=} \sum_i P(Y = y_i|X)P(Z|Y = y_i) \rightarrow \text{Not in general!}$$

Causal Chains

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

- This configuration is a *causal chain*



- Is X independent of Z given Y?
 - Yes!

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

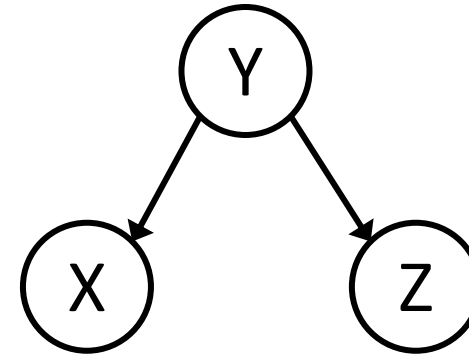
$$P(Z|X, Y) = \frac{P(X, Y, Z)}{P(X, Y)} = \frac{P(X)P(Y|X)P(Z|Y)}{P(X)P(Y|X)} = P(Z|Y)$$

- Evidence along the chain “blocks” influence

Common Cause

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

- Are X and Z independent?
 - No!
- Are X and Z independent given Y?
 - Yes!



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

- Observing the cause “blocks” influence between effects

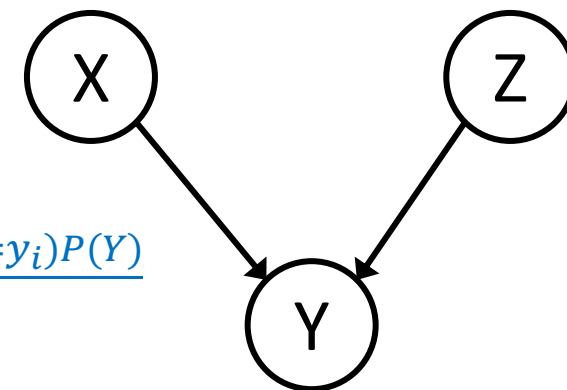
Common Effect

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

- Are X and Z independent?

- Yes!

$$\begin{aligned} P(X, Z) &= \sum_i P(X, Y = y_i, Z) = \sum_i P(X)P(Z)P(Y = y_i | X, Z) = \sum_i \frac{P(X)P(Z)P(X, Z | Y = y_i)P(Y)}{P(X, Z)} \\ &= \frac{P(X)P(Z)}{P(X, Z)} \sum_i P(X, Z | Y = y_i)P(Y) = \frac{P(X)P(Z)}{P(X, Z)} P(X, Z) = P(X)P(Z) \end{aligned}$$



- Are X and Z independent given Y?

- No!

- Seeing Y puts X and Z in competition as the explanation

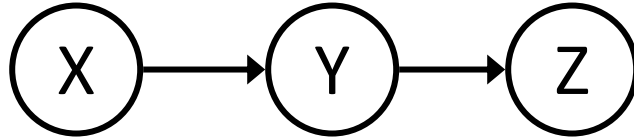
$$P(Z | X, Y) = \frac{P(X, Y, Z)}{P(X, Y)} = \frac{P(X)P(Z)P(Y | Z, X)}{P(X)P(Y | X)} = \frac{P(Z)P(Y | Z, X)}{P(Y | X)} \neq P(Z | Y) \text{ (inequality holds in general)}$$

- Seeing evidence “activates” influence

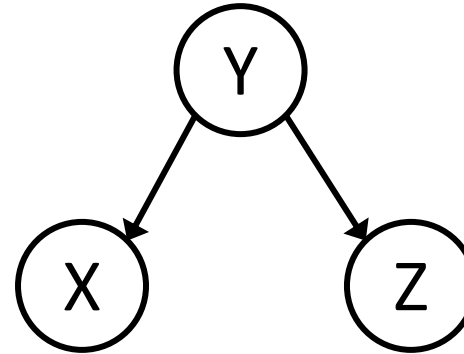
3-Node Cond. Indep. Recap

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

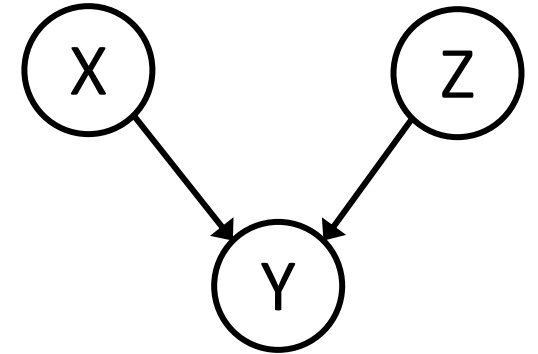
Causal Chain



Common Cause



Common Effect



$X \perp\!\!\!\perp Z?$

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

General Case

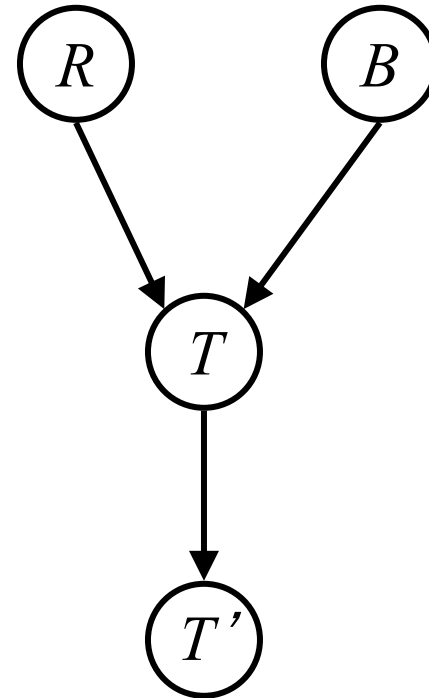
- Any complex example can be analyzed using these three canonical cases
- General question: In a given BN, are two variables independent (given evidence)?
- Solution: Analyze the graph

Example

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

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

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



Example

R: Rain
T: Traffic
D: Roof drip
S: Sadness

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

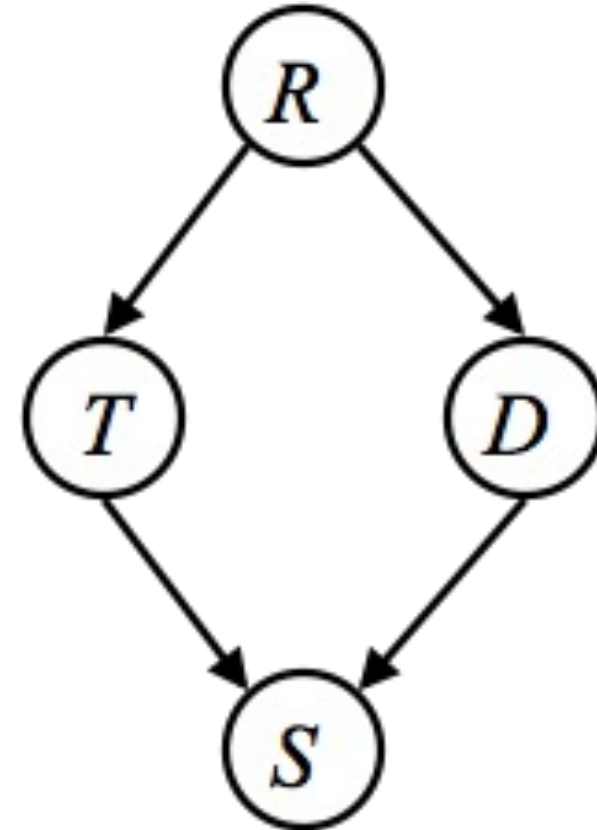
No, common cause

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

Yes

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

No, common effect



Example

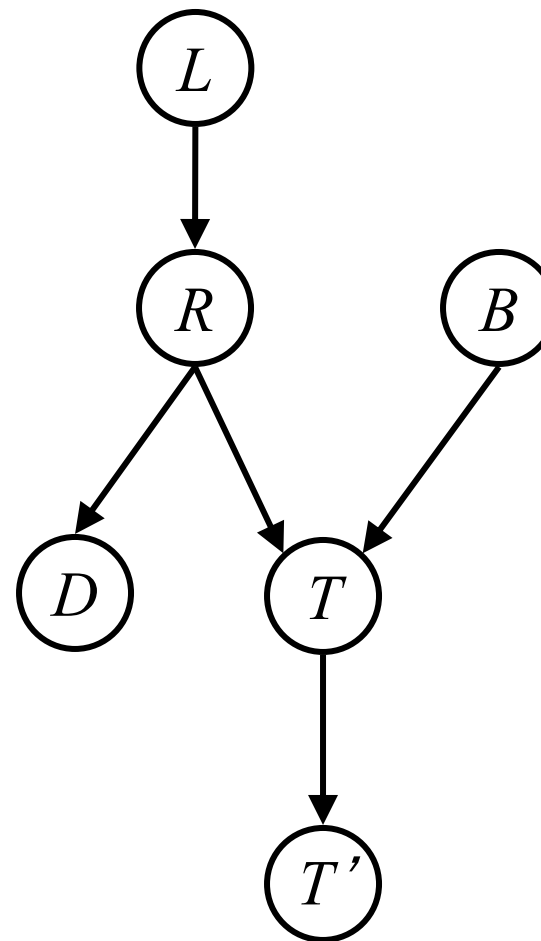
$L \perp\!\!\!\perp T' | T$ *Yes*

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

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

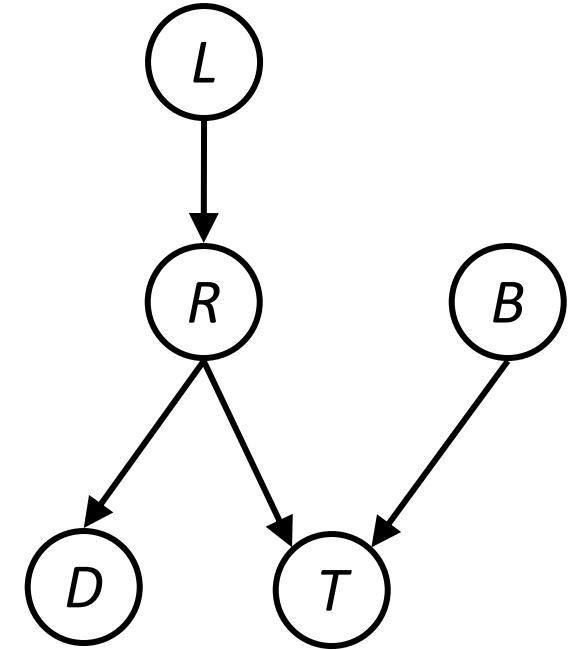
$L \perp\!\!\!\perp B | T'$ *No*

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



Reachability

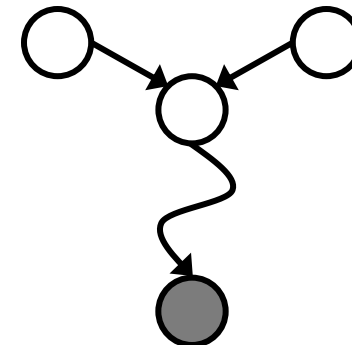
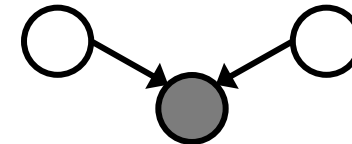
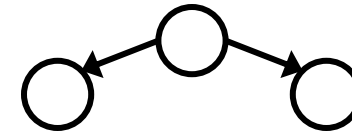
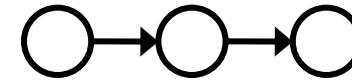
- Recipe: shade evidence nodes, look for paths in the resulting graph
- Attempt 1: if two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent
- Almost works, but not quite
 - Where does it break?
 - Answer: the v-structure at T doesn't count as a link in a path unless “active”



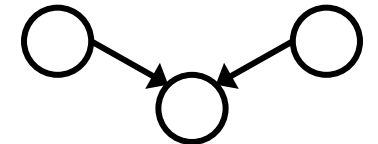
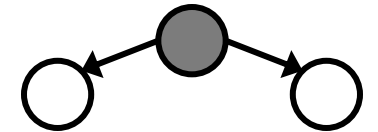
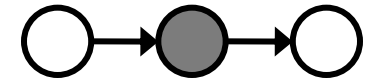
Active / Inactive Paths

- Question: Are X and Y conditionally independent given evidence variables {Z}?
 - Yes, if X and Y “d-separated” by Z
 - Consider all (undirected) paths from X to Y
 - No active paths = independence!
- A path is active if each triple is active:
 - Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
 - Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
 - Common effect (aka v-structure)
 $A \rightarrow B \leftarrow C$ where B or one of its descendants is observed
- All it takes to block a path is a single inactive segment

Active Triples



Inactive Triples



D-Separation

- Query: $X_i \perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\} \text{ ?}$
- Check all (undirected!) paths between X_i and X_j
 - If one or more active, then independence not guaranteed

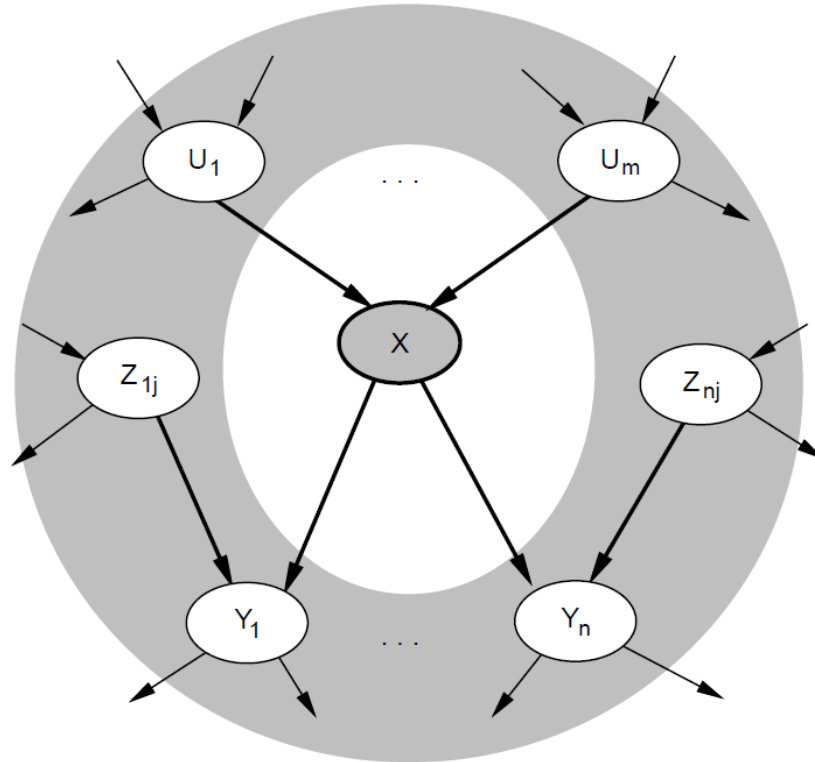
$$X_i \not\perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$$

- Otherwise (i.e. if all paths are inactive),
then independence is guaranteed

$$X_i \perp\!\!\!\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$$

Markov Blanket

- Each node is conditionally independent of all others given its *Markov Blanket*
- Markov Blanket: parents + children + children's parents



Summary

- Bayesian networks provide a natural representation for conditional independence
- Topology + CPTs = compact representation of joint distribution
- Guaranteed independencies of distributions can be deduced from BN graph structure
- Generally easy for domain experts to construct