343H: Honors AI

Lecture 15: Bayes Nets Independence

3/18/2014

Kristen Grauman

UT Austin

Slides courtesy of Dan Klein, UC Berkeley

Probability recap

- Conditional probability $$P(x|y) = \frac{P(x,y)}{P(y)}$$
- Product rule $$P(x,y) = P(x|y)P(y)$$
- Chain rule $$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i|x_1 \ldots x_{i-1})$$
- X, Y independent if and only if:

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

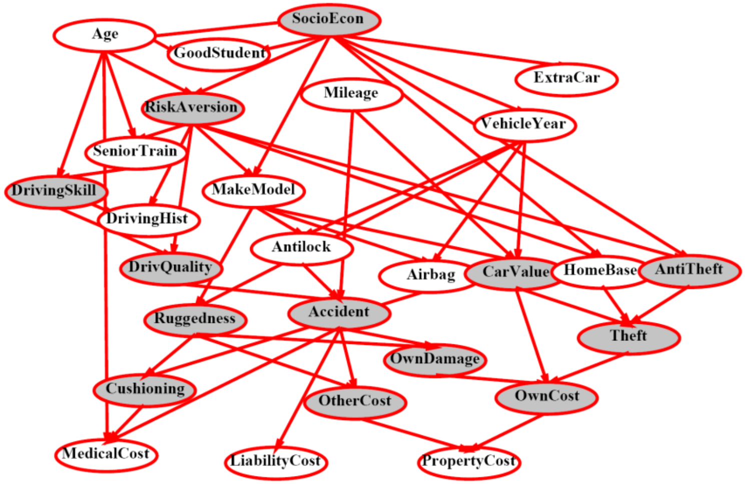
- X and Y are conditionally independent given Z if and only if:

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

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

Bayes' Nets

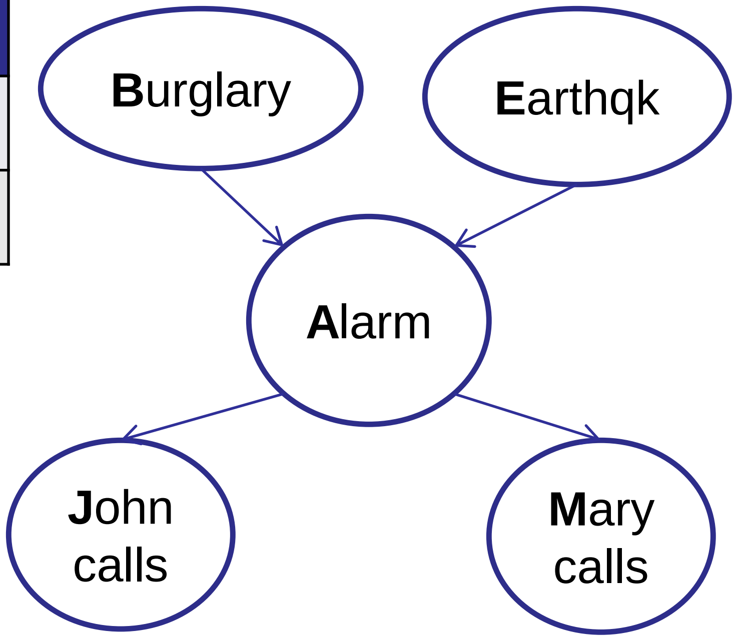
- A Bayes' net is an efficient encoding of a probabilistic model of a domain



- Questions 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?

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

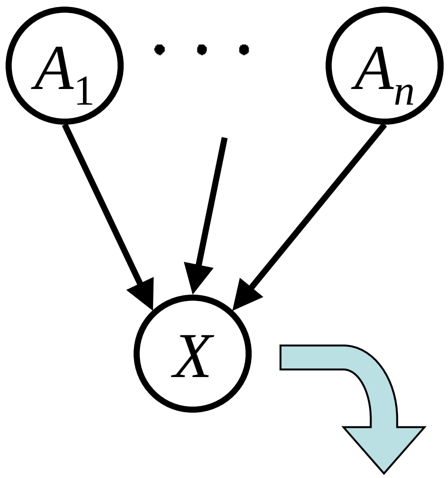
Bayes' Net Semantics

- 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

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

- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions

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



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

Recall: Probabilities in 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 distribution?

- Chain rule (valid for all distributions):

$$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i|x_1 \ldots x_{i-1})$$

- Due to <u>assumed</u> conditional independences:

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

- Consequence:

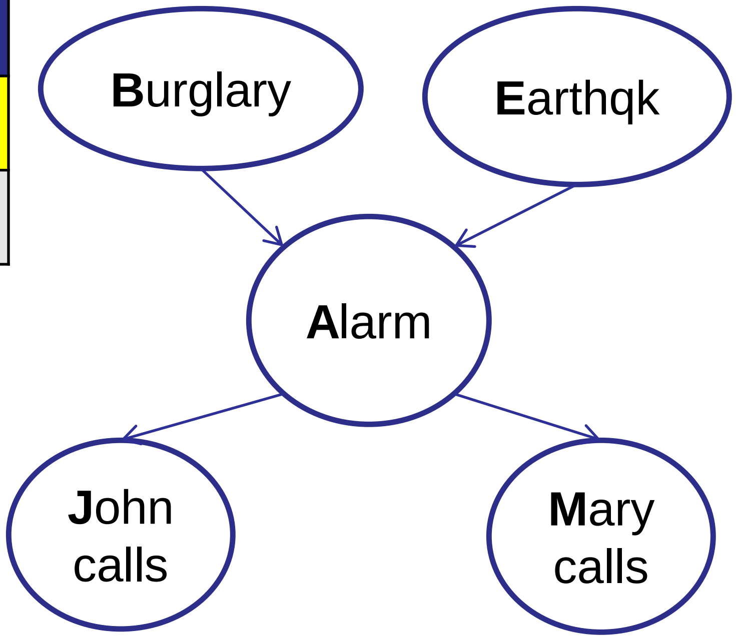
$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|\mathit{parents}(X_i))$$

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

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

0.001 x 0.998 x 0.94 x 0.1 x 0.7

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

Size of a Bayes' Net

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

  2^N

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

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

- Both give you the power to calculate $P(X_1, X_2, \ldots X_n)$
- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also turns out to be faster to answer queries (coming)

Bayes’ Net

- Representation
- Conditional independences
- Probabilistic inference
- Learning Bayes’ Nets from data

Conditional Independence

- X and Y are independent if

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

- X and Y are conditionally independent given Z

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

- (Conditional) independence is a property of a distribution

- Example: $Alarm \perp\!\!\!\perp Fire|Smoke$

Bayes Nets: Assumptions

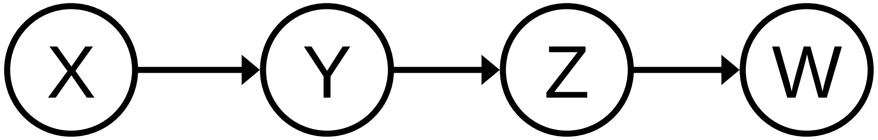
- Assumptions we are required to make to define the Bayes net when given the graph:

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

- Beyond the above ("chain-rule→Bayes net") conditional independence assumptions
  - Often have many more conditional independences
  - They can be read off the graph

- Important for modeling: understand assumptions made when choosing a Bayes net graph

Example

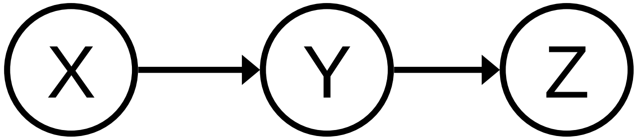


- Conditional independence assumptions directly from simplifications in chain rule:

- Additional 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:



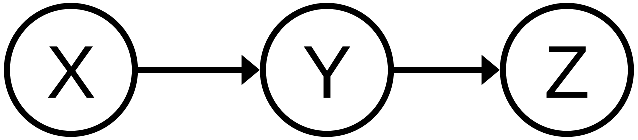
  - Question: are X and Z necessarily independent?
    - Answer: no. Example: low pressure causes rain, which causes traffic.
    - X can influence Z, Z can influence X (via Y)

D-separation: Outline

- D-Separation: a condition/algorithm for answering such queries
- Study independence properties for triples
- Analyze complex cases in terms of member triples – reduce big question to one of the base cases.

Causal Chains (1 of 3 structures)

- This configuration is a "causal chain"



X: Low pressure

Y: Rain

Z: Traffic

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

- Is X independent of Z given Y?

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

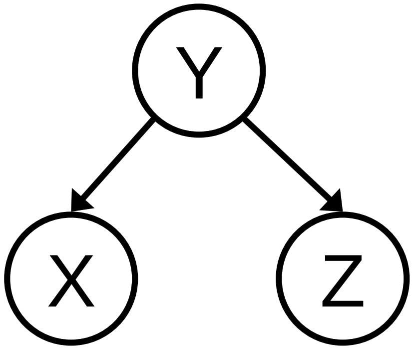
$$= P(z|y)$$

Yes!

- Evidence along the chain "blocks" the influence

Common Cause (2 of 3 structures)

- Another basic configuration: two effects of the same cause
  - Are X and Z independent?
  - Are X and Z independent given Y?



Y: Project due

X: Piazza busy

Z: Lab full

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

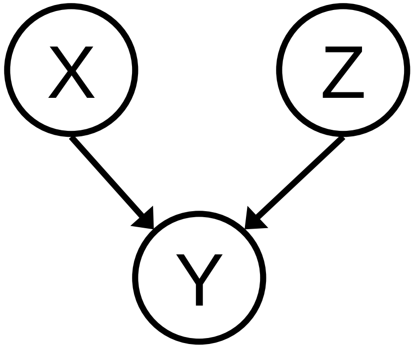
$$= P(z|y)$$

Yes!

  - Observing the cause blocks influence between effects.

Common Effect (3 of 3 structures)

- Last configuration: two causes of one effect (v-structures)
  - Are X and Z independent?
    - Yes: the ballgame and the rain cause traffic, but they are not correlated
  - Are X and Z independent given Y?
    - 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.



X: Raining

Z: Ballgame

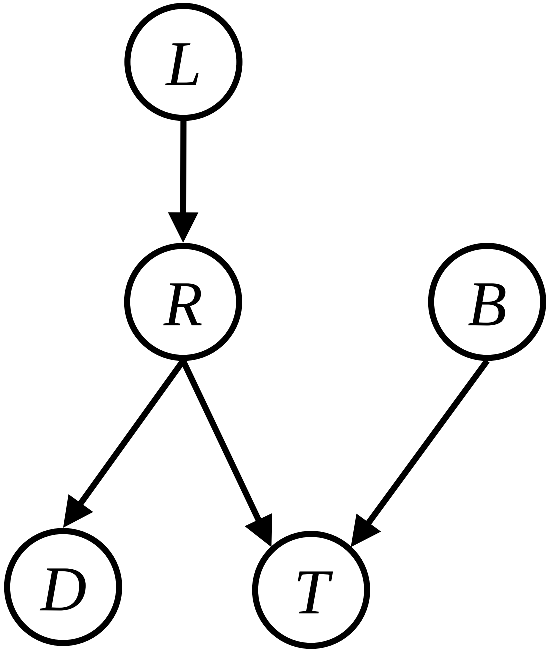
Y: Traffic

The General Case

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

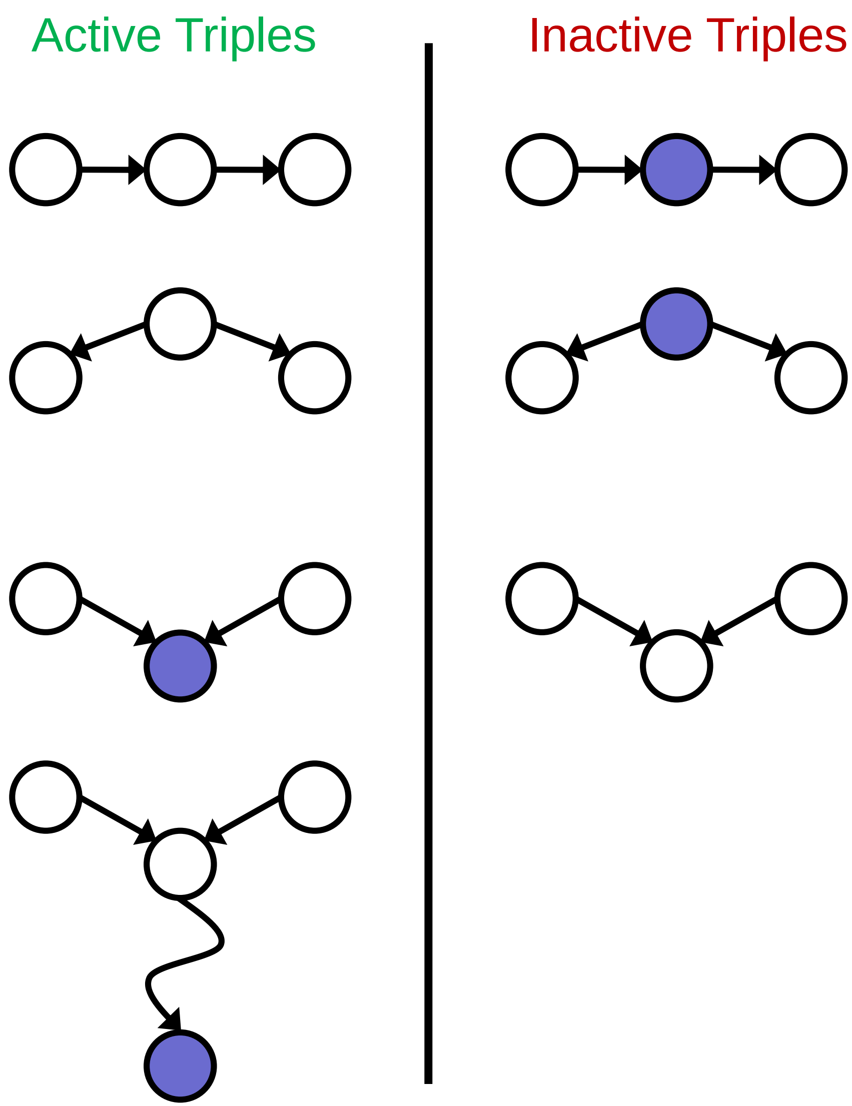
Reachability

- Recipe: shade evidence nodes, look for paths in the resulting graph
- Attempt 1: if two nodes are connected by an undirected path 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"



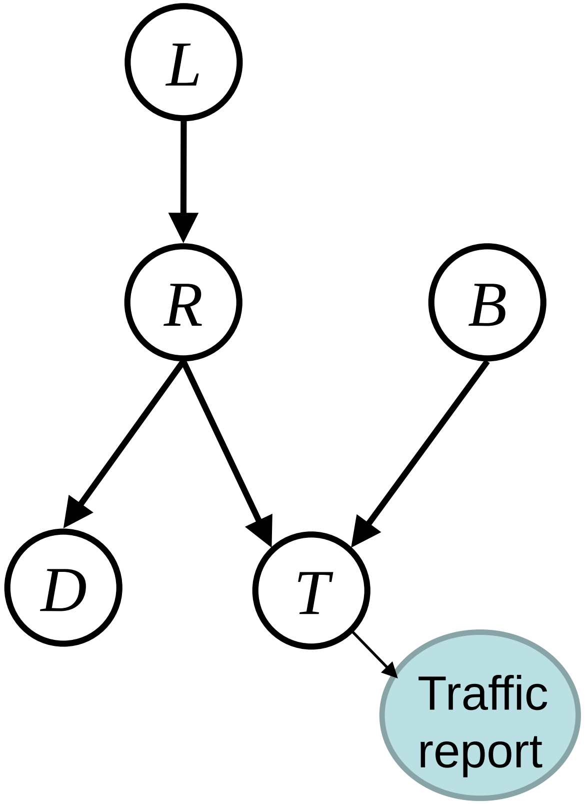
Active / Inactive paths

- Question: Are X and Y conditionally independent given evidence vars {Z}?
  - Yes, if X and Y “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 → 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 descendents* is observed
- All it takes to block a path is a single inactive segment



Reachability

- Recipe: shade evidence nodes, look for paths in the resulting graph

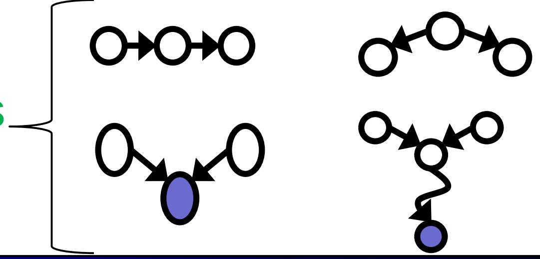


D-Separation

- Given query $X_i \overset{?}{\perp\!\!\!\perp} X_j | \{X_{k_1}, ..., X_{k_n}\}$
- For all (undirected!) paths between X_i and X_j
  - Check whether path is active
    - If active return $X_i \not\perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$
- Otherwise (i.e., if all paths are inactive) then independence is guaranteed.
  - Return $X_i \perp\!\!\!\perp X_j | \{X_{k_1}, ..., X_{k_n}\}$

Example 1

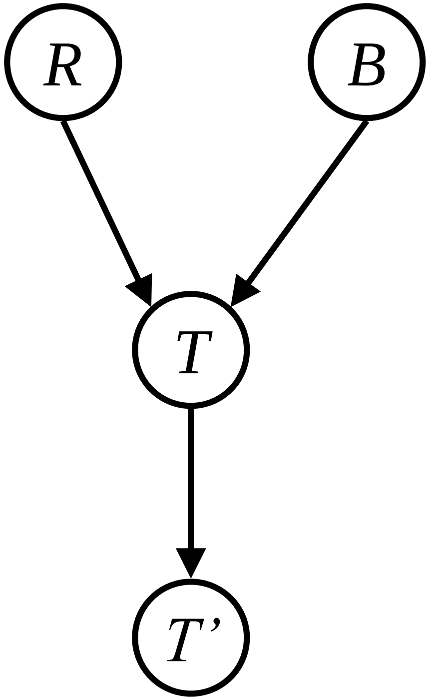
Active Triples



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

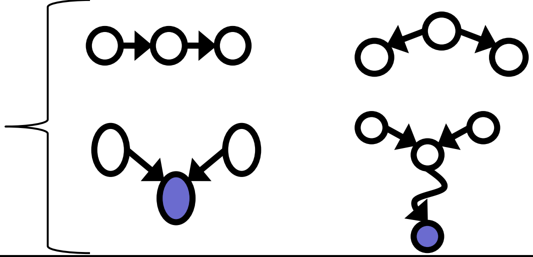
$R \perp\!\!\!\perp B | T$

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



Example 2

Active Triples



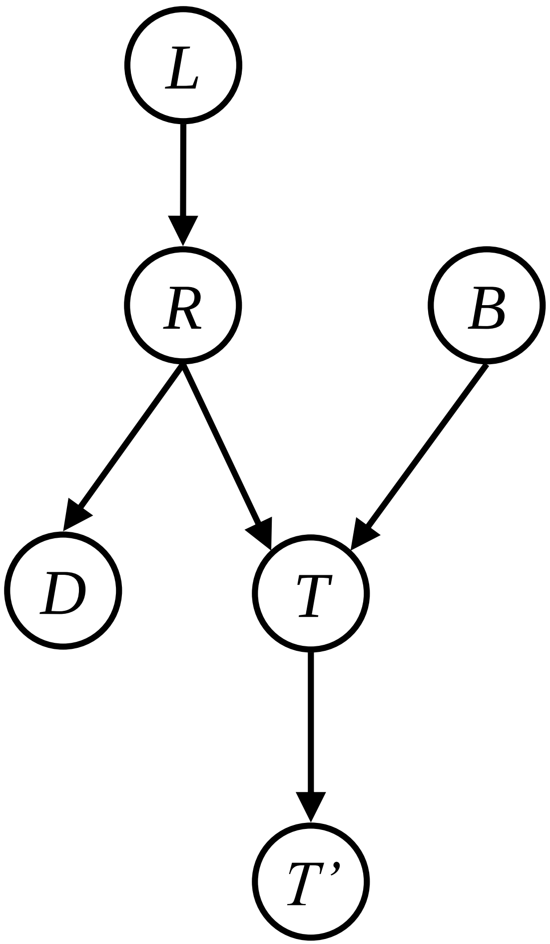
$L \perp\!\!\!\perp T' | T$ *Yes*

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

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

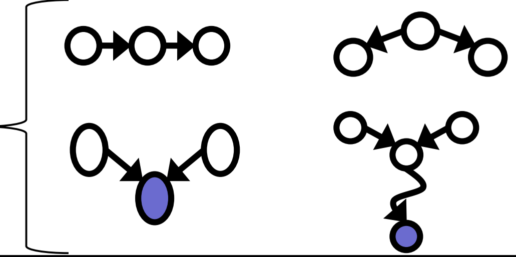
$L \perp\!\!\!\perp B | T'$

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

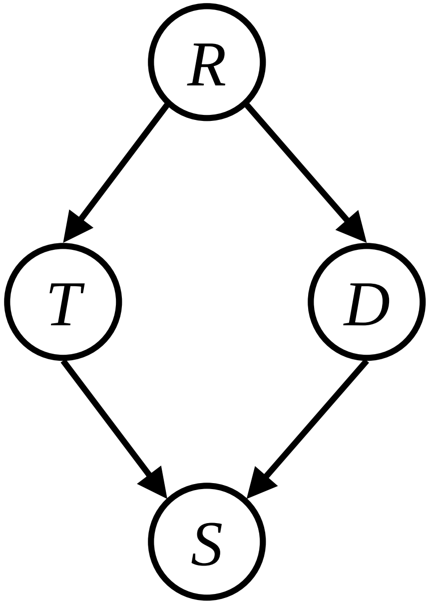


Example 3

Active Triples



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



- Questions:

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

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

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

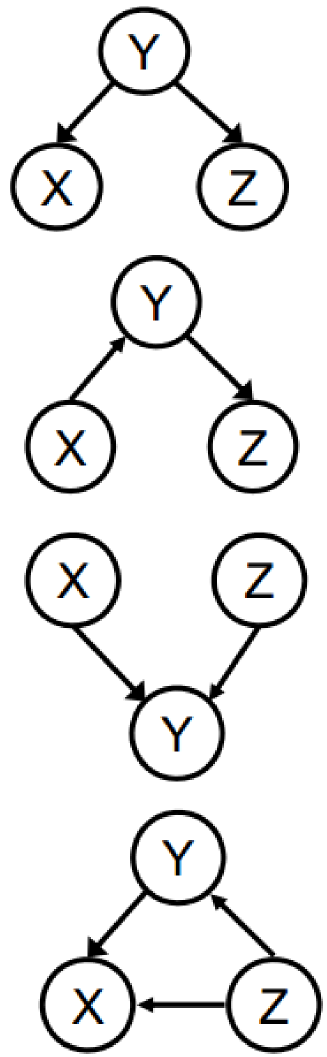
Structure implications

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

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

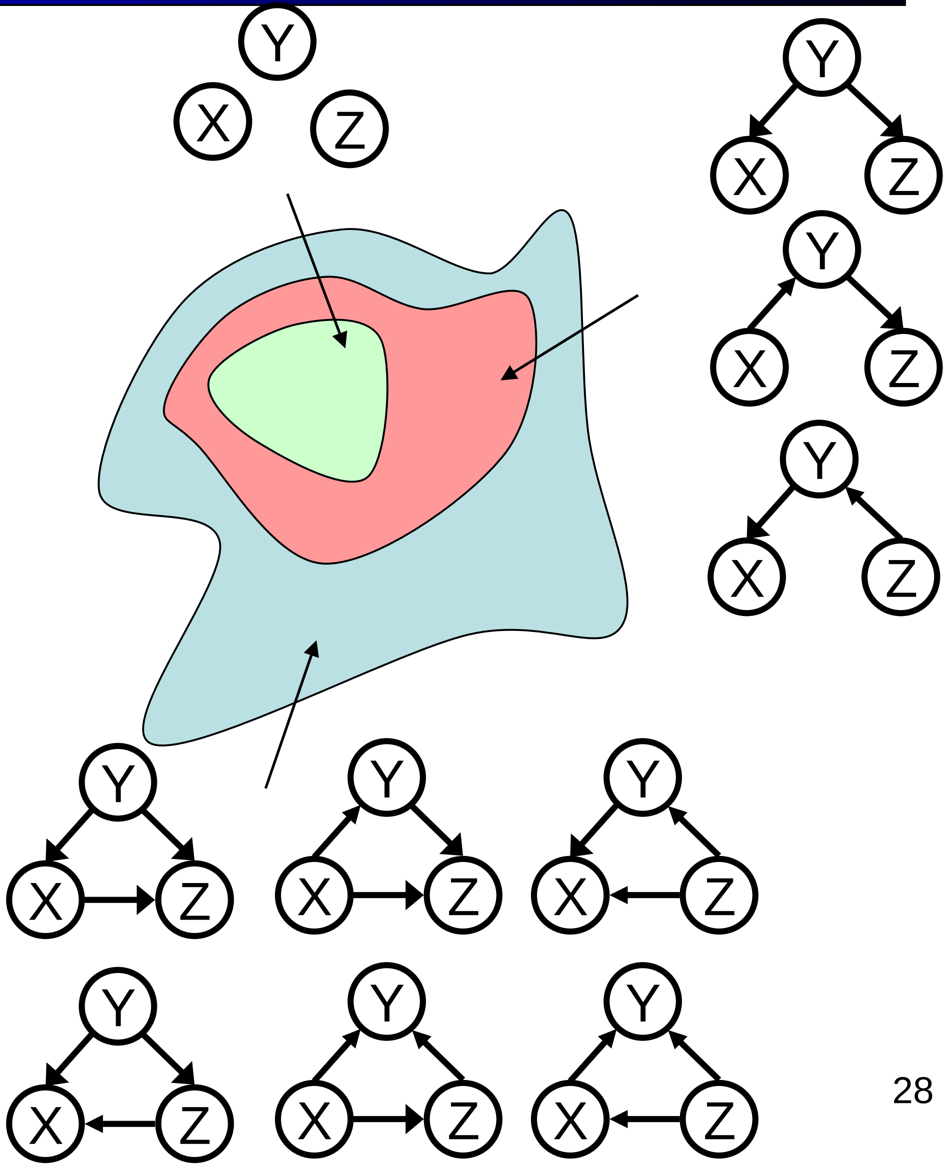
- This list determines the set of probability distributions that can be represented by this BN

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



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

Bayes' Net

- ✓ Representation
- ✓ Conditional independences
- Probabilistic inference
- Learning Bayes' Nets from data