复旦大学大数据学院
School of Data Science, Fudan University

魏忠钰

# Bayes' Net
# Representation and Independence

Data Intelligence and Social Computing Lab (DISC)

December 7th, 2021

# Outline

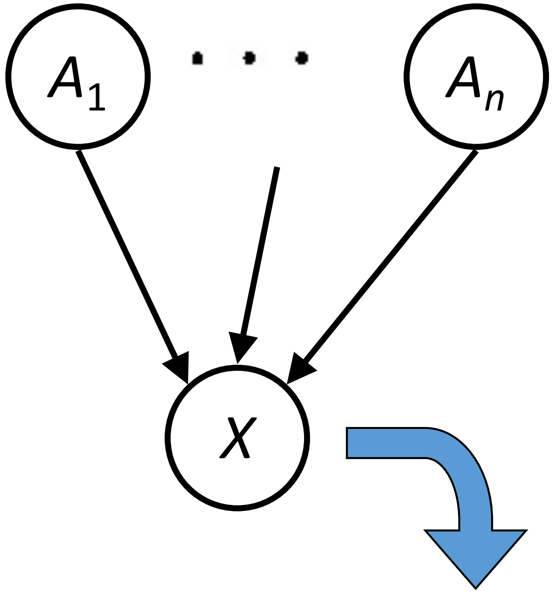
- Bayes' Net

# Bayes' Nets – Big picture

- A technique for describing complex joint distributions using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - Bayes' nets describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions

# Bayes' Net

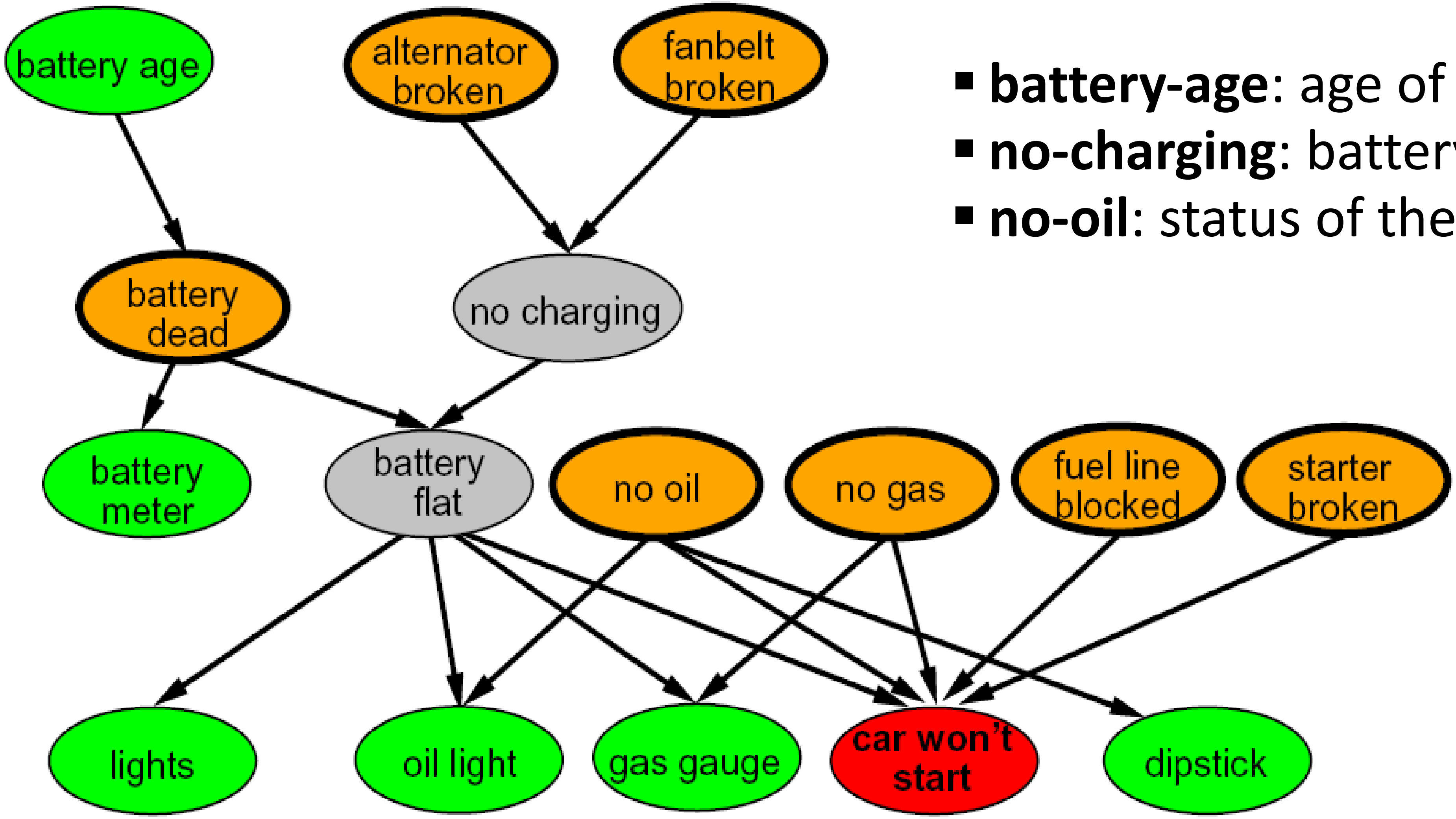
- Nodes: variables (with domains), assigned or unassigned
- Arcs: direct interactions between nodes
- Bayes' Net
  - A set of nodes, one per variable X
  - A directed, acyclic graph
  - A collection of CPTs, one for each variable
    - CPT: conditional probability table
    - Conditional distribution of X given all its parents. One entry for each combination of parents' values




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

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

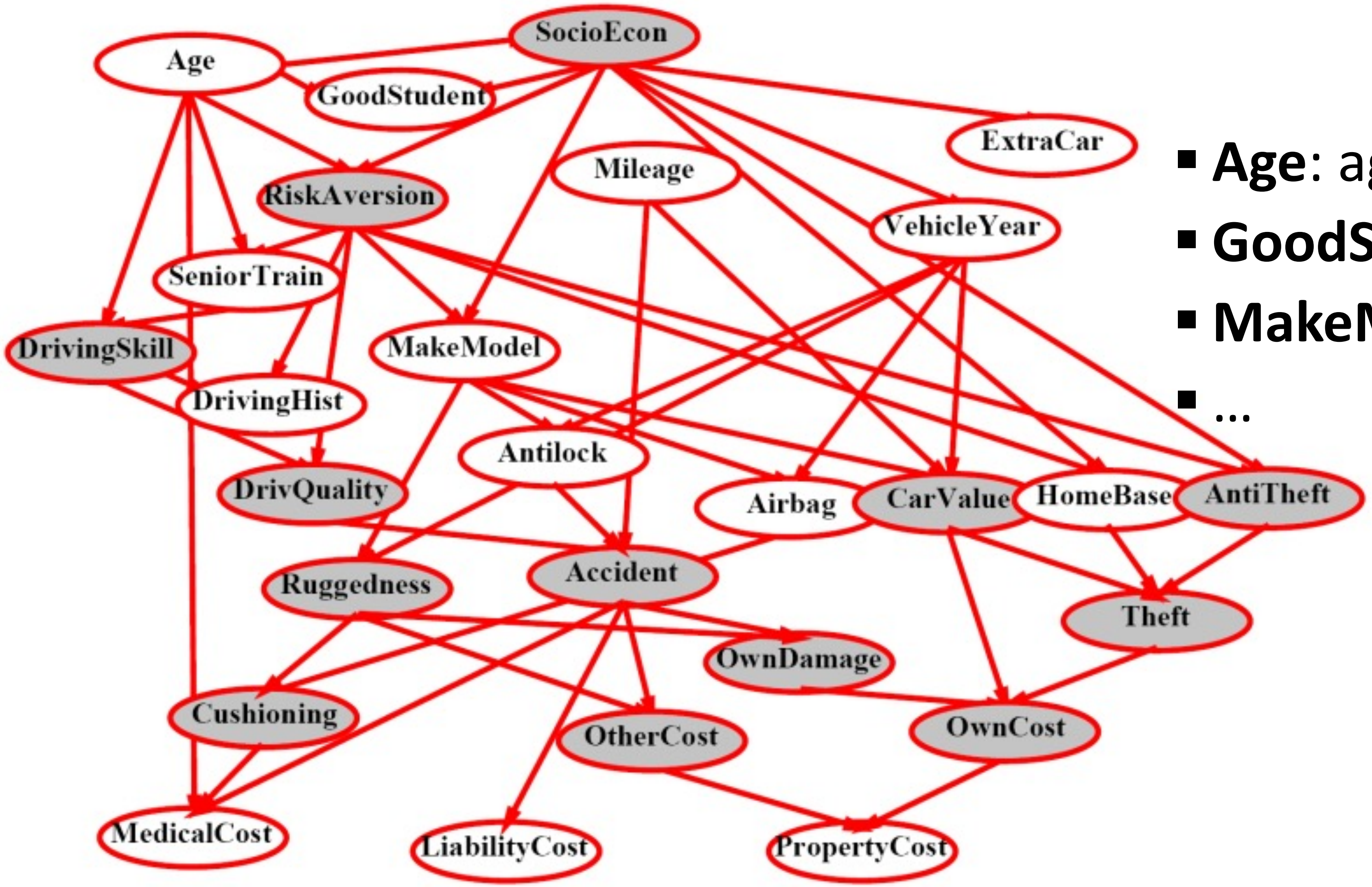
# Theme: Car Trouble Shooter




- **battery-age**: age of the battery of the car
- **no-charging**: battery is lack of power
- **no-oil**: status of the tank

- **no-charging → battery-flat**: forget to charge will cause the flat of battery
- **no-oil → oli-light**: lack of oil will be shown in the oil light

# Theme：Car Insurance Evaluation for student driver




- **Age**: age of the driver
- **GoodStudent**: quality of the driver
- **MakeModel**: model of the car
- …

- **Mileage → Accident**: higher mileage means more chances of accidents
- **Accident → LiabilityCost**: Cars with accident needs to pay more on liability cost

# Build a Bayes' Net

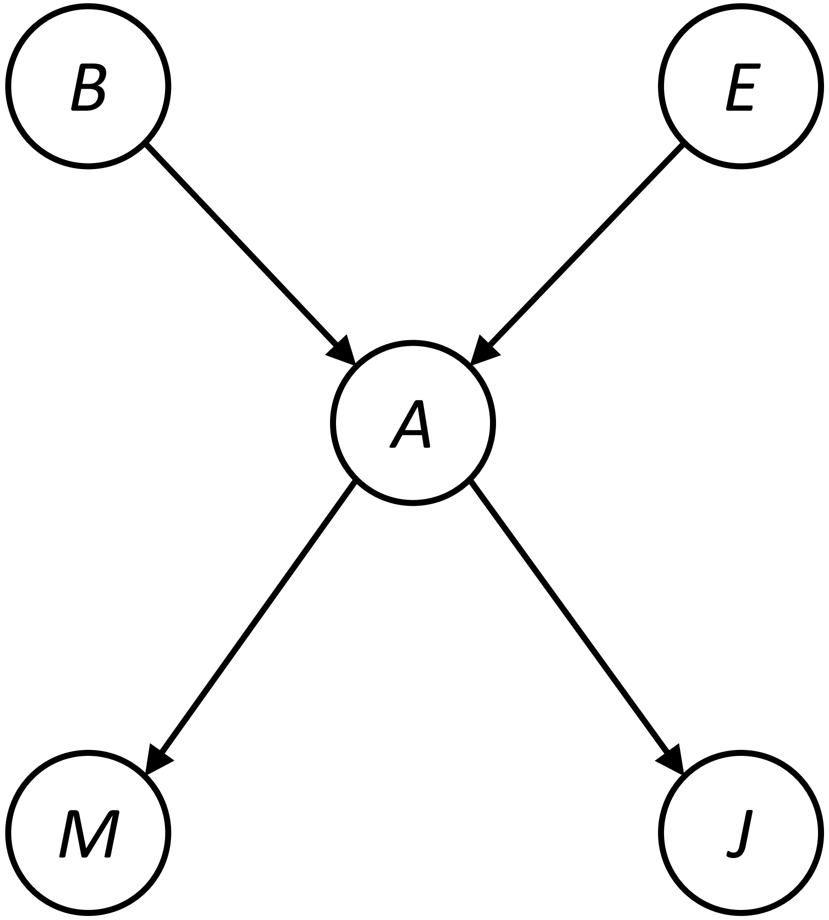
- Input: a set variables
- Output: a directed, acyclic graph

- Numbering nodes from small to large (1 to N)
- Add node with smallest number into the graph
- Add directed link from existing nodes to the new one if there is interaction between them

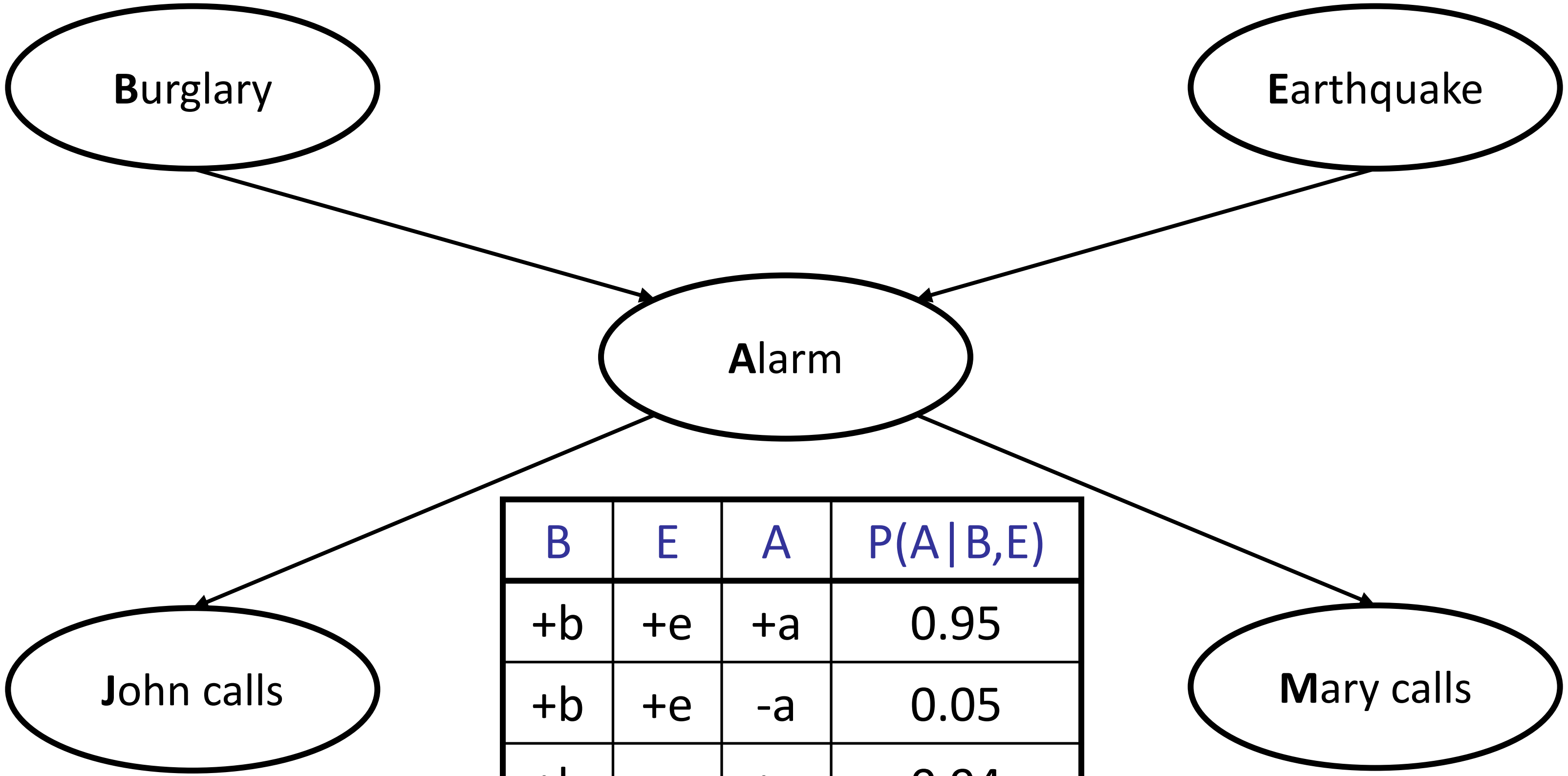
- What interaction to use?
  - Cause-effect can be the interaction mode
- How to guarantee there is no cycle in the graph?
  - link is from node with small number to node with larger number

# Example: Alarm Network

- Variables (add nodes from top to bottom)
  - B: Burglary
  - E: Earthquake
  - A: Alarm
  - M: Mary calls
  - J: John calls



# Example: Alarm Network




| B | P(B) |
|---|---|
| +b | 0.001 |
| -b | 0.999 |

| E | P(E) |
|---|---|
| +e | 0.002 |
| -e | 0.998 |

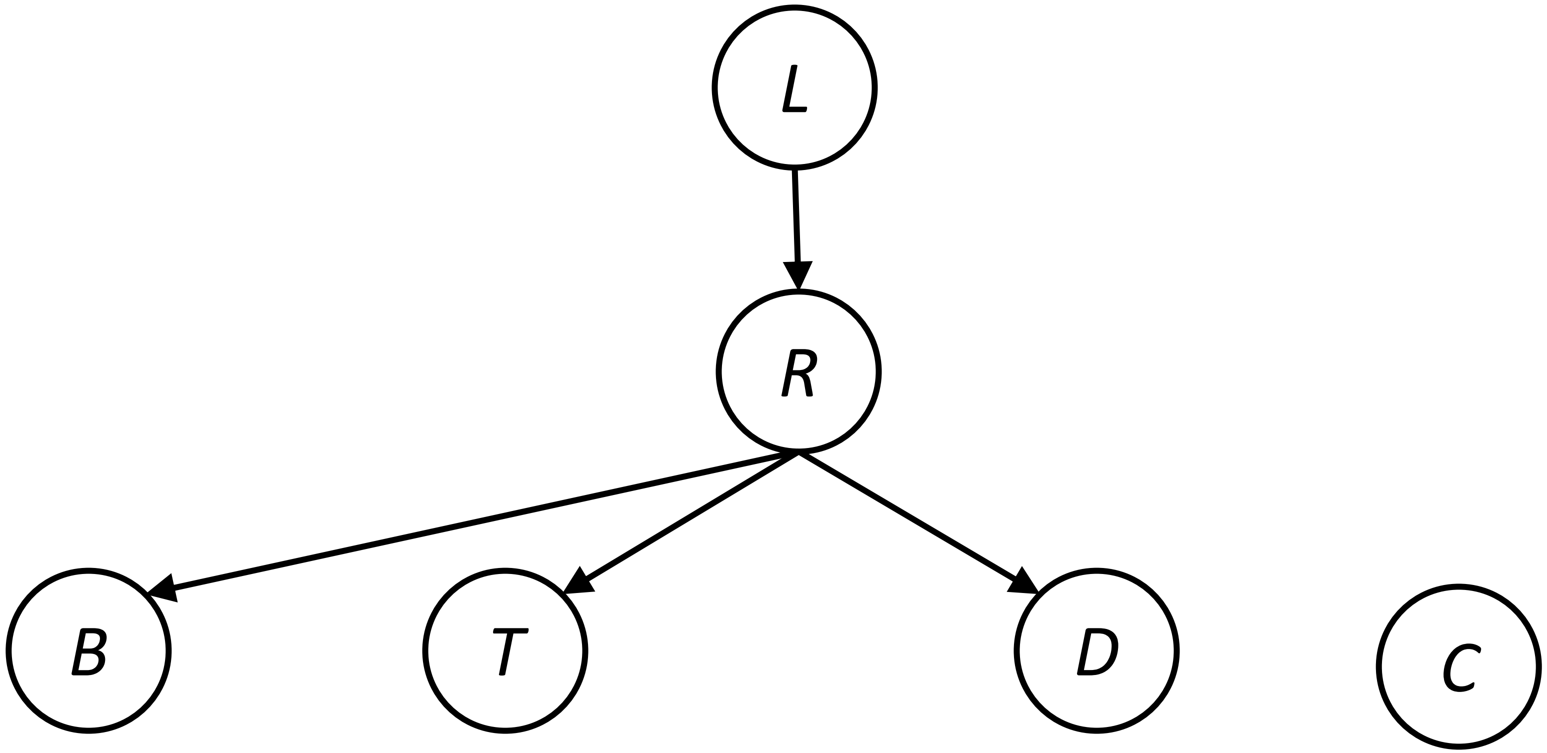
| B | E | A | P(A\|B,E) |
|---|---|---|---|
| +b | +e | +a | 0.95 |
| +b | +e | -a | 0.05 |
| +b | -e | +a | 0.94 |
| +b | -e | -a | 0.06 |
| -b | +e | +a | 0.29 |
| -b | +e | -a | 0.71 |
| -b | -e | +a | 0.001 |
| -b | -e | -a | 0.999 |

| 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 |

# Example: Traffic II

- Variables
  - L: Low pressure
  - R: It rains
  - D: Roof drips
  - T: Traffic
  - B: Ballgame
  - C: Cavity

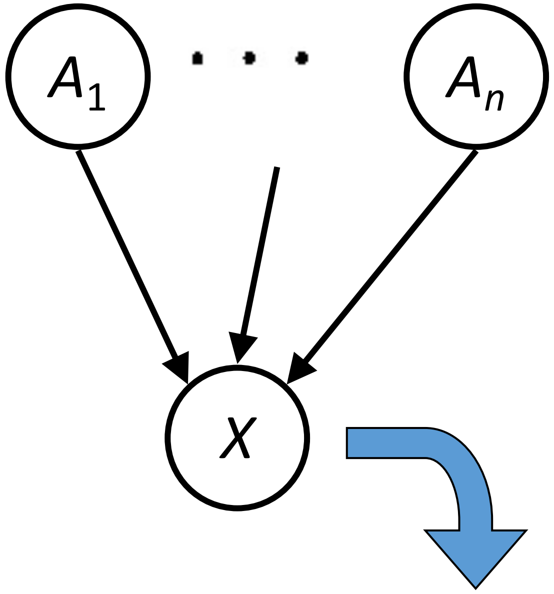


# Outline

- Bayes' Net
- Joint Probability in Bayes' Net

# Bayes' Net

- Nodes: variables (with domains), assigned or unassigned
- Arcs: direct interactions between nodes
- Bayes' Net
  - A set of nodes, one per variable X
  - A directed, acyclic graph
  - A collection of CPTs, one for each variable
    - CPT: conditional probability table
    - Conditional distribution of X given all its parents. One entry for each combination of parents' values




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

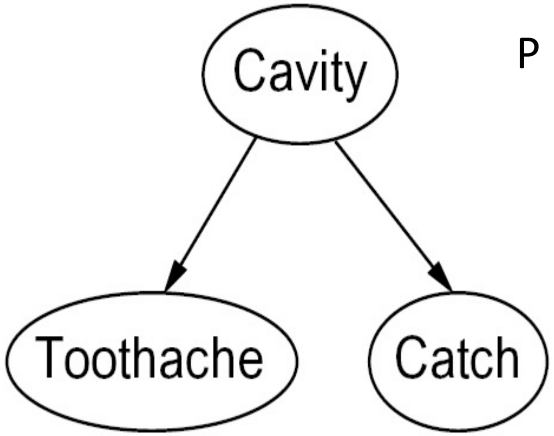
*A Bayes net = Topology (graph) + Local Conditional Probabilities*

# Joint Probabilities in BNs

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

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

  - Example:




P (Cavity)

P (Toothache | Cavity)

P (Catch|Cavity)

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

P(+cav., +cat., -t) = P(+cav.) * P (+cat. | +cav.) * (+t | +cav.)

# Recall Conditional Independence

- X and Y are independent if

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

- 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
  - If it holds for every assignment variables, we can capitalize the letter.

# Conditional Independence in BNs

- Why are we guaranteed that setting?

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

- Chain rule (valid for all distributions):

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

- <u>Assume</u> conditional independences:

$$x_i \perp \{x_1, \ldots x_{i-1} / parents(x_i)\} | \; parents(x_i)$$

→ Consequence:

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

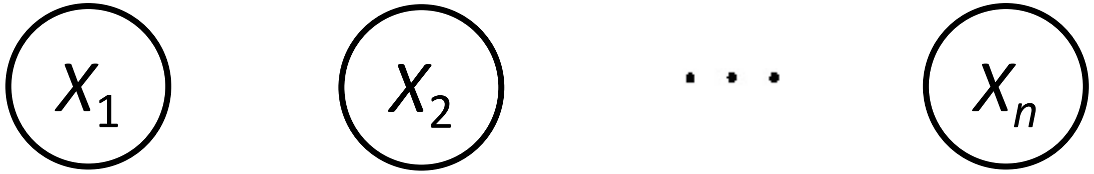
$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | \mathit{parents}(X_i))$$

# Size of a Bayes' Net

- To calculate joint distribution over N Boolean variables
  - How big is a full joint distribution table?
    - $2^N$
  - How big is the bayes' net if variables have up to k parents?
    - $O(N * 2^{k+1})$

- Both full joint distribution table and bayes' net are able to calculate joint distribution
- BNs: save storage spaces and faster computation

# Example: Coin Flips




$P(X_1)$

| | |
|---|---|
| h | 0.5 |
| t | 0.5 |

$P(X_2)$

| | |
|---|---|
| h | 0.5 |
| t | 0.5 |

$\cdots$

$P(X_n)$

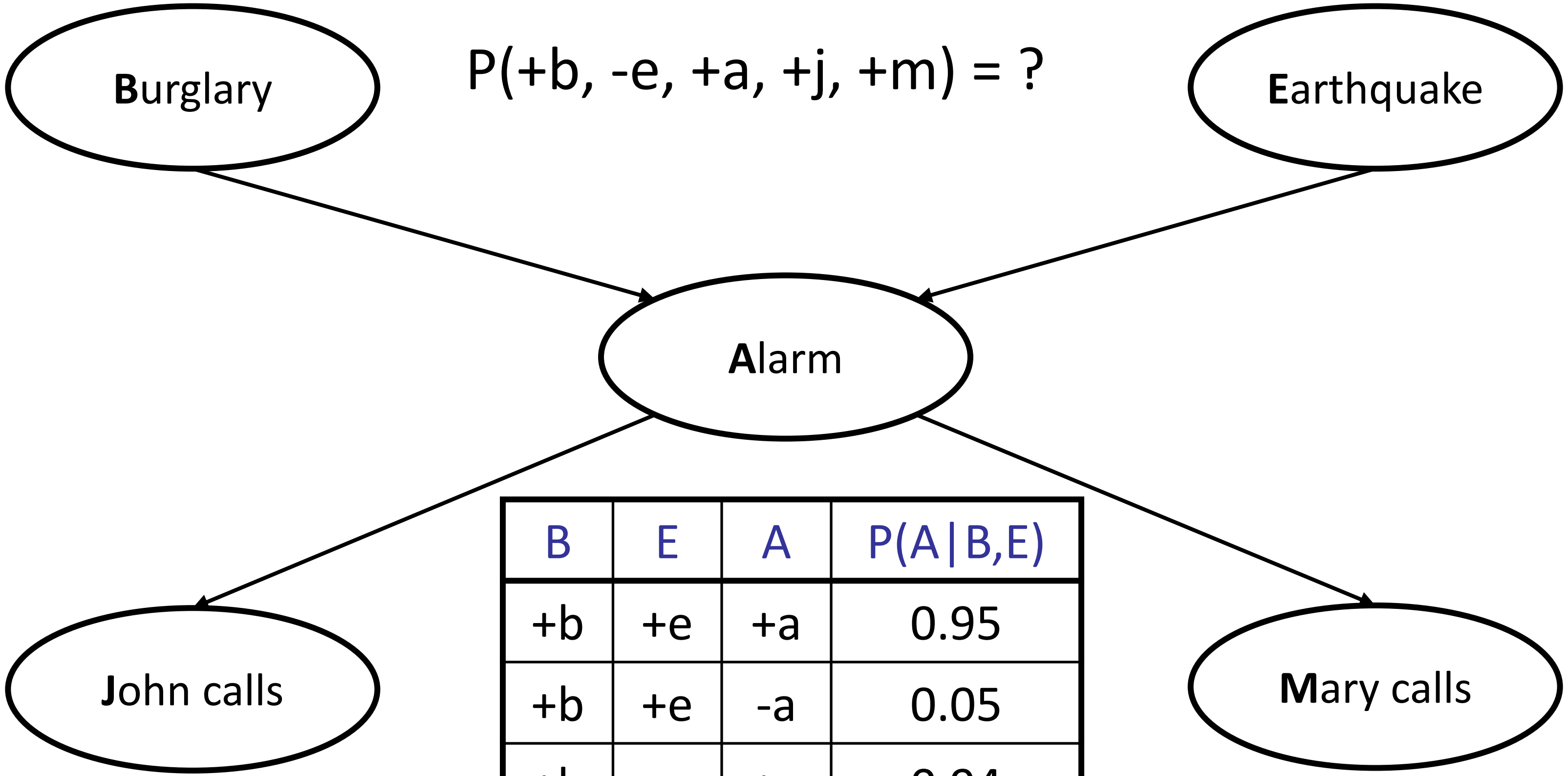
| | |
|---|---|
| h | 0.5 |
| t | 0.5 |

$$P(h,h,t,h) = P(x_1 = h) * P(x_2 = h) * P(x_3 = t) * P(x_4 = h)$$

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

# Alarm Network

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




| 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 |

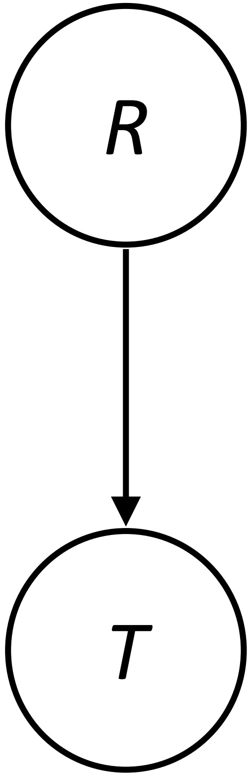
| A | M | P(M\|A) |
|---|---|---|
| +a | +m | 0.7 |
| +a | -m | 0.3 |
| -a | +m | 0.01 |
| -a | -m | 0.99 |

# Outline

- Bayes’ Net
- Joint Probability in Bayes’ Net
- Independence in Bayes’ Net

# Causal Relation in Bayes’ Net

## Causal direction



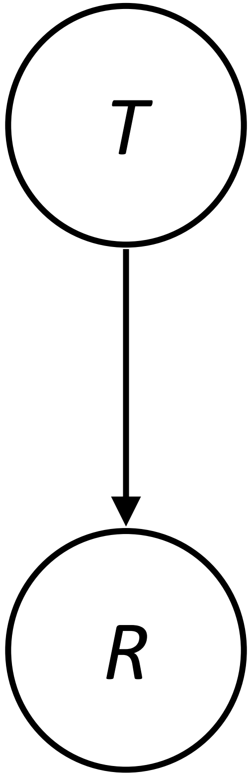

$P(R)$

| | |
|---|---|
| +r | 1/4 |
| -r | 3/4 |

$P(T|R)$

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

## Reverse causality




$P(T)$

| | |
|---|---|
| +t | 9/16 |
| -t | 7/16 |

$P(R|T)$

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

$P(T, R)$

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

**R： Rain** **T： Traffic**

**BNs need not actually be causal**

# Causality?

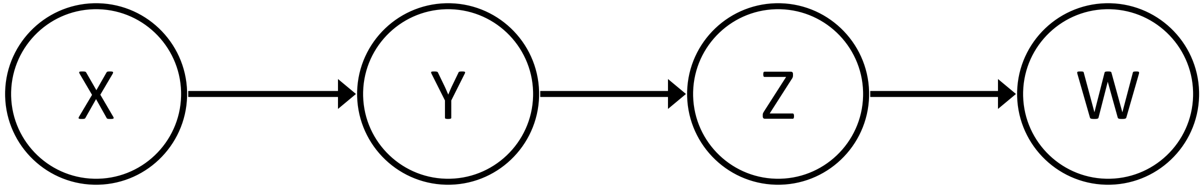
- When Bayes' nets reflect the true causal patterns:
  - Make the graph simpler (nodes have fewer parents)
  - Easier to understand the graph
  - Easier to construct the graph (elicit from experts)
- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables 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?

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

  - Topology really encodes **conditional independence**

$$x_i \perp \{x_1, \ldots x_{i-1} / parents(x_i)\} | \; parents(x_i)$$

# Example




- Conditional independence assumptions can be derived directly from simplifications in chain rule:

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

- Implied conditional independence:

$$X \perp Z|Y \qquad X \perp W, Y \perp W\ |Z$$

- Additional implied conditional independence assumptions?

$$X \perp W\ ? \qquad X \perp W|Y\ ?$$

# Independence in a BN

- General question: in a given BN, are two variables independent (given evidence)?

- Question:
- Are X and Y conditionally independent given evidence variables {Z}?

# Outline

- Bayes’ Net
- Bayes’ Net Semantics
- Independence of Bayes’ Net
- D-separation

# D-Separation

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

- Check all paths between $X_i$ and $X_j$
  - If one or more active, then **independence not guaranteed**

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

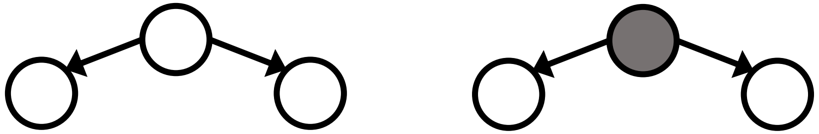
  - Otherwise (i.e. if all paths are inactive), then independence is guaranteed

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

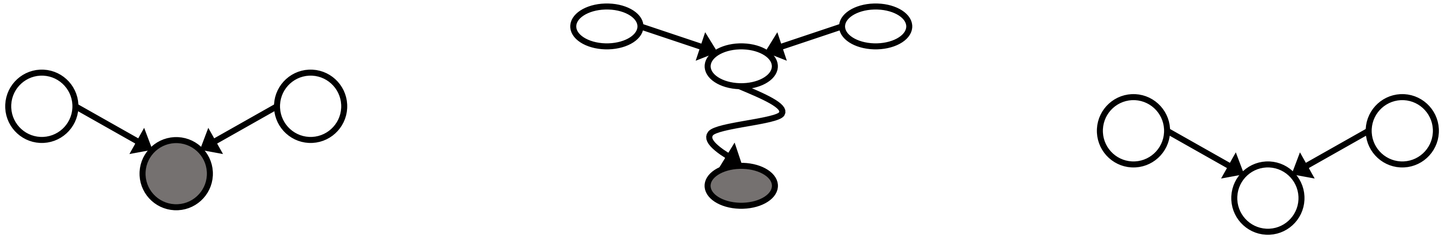
# Active / Inactive Paths

- **A path is active if every triple is active:**
  - Causal chain

  - Common cause

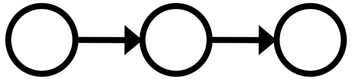
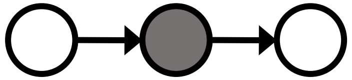
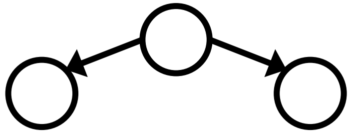
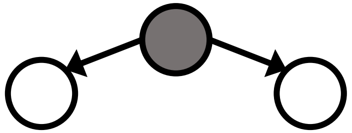
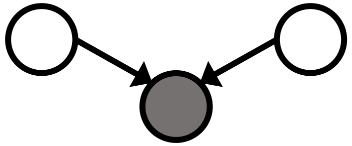
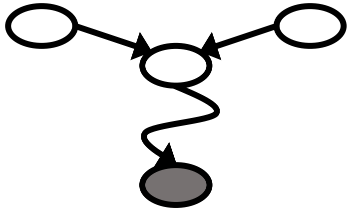
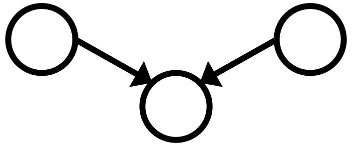


  - Common effect



- All it takes to block a path is a single inactive segment

# independence properties for triples

- Triples (with or without evidence)
  - Causal chain
     
  - Common cause
     
  - Common effect
      
- Are two nodes independent given certain evidence?
  - If yes, can prove using algebra
  - If no, can prove with a counter example

# Causal Chains w/o evidence

$$X \rightarrow Y \rightarrow Z$$

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

X: Low pressure   Y: Rain   Z: Traffic

- X independent of Z ? $P(+x, -z) = P(+x)P(-z)?$
- **NO! Give a counter-case**
  - Low pressure causes rain causes traffic certainly
  - High pressure causes no rain causes no traffic certainly

| | |
|---|---|
| +x | 0.5 |
| -x | 0.5 |

P(X)

| | | |
|---|---|---|
| +x | +y | 1 |
| | -y | 0 |
| -x | +y | 0 |
| | -y | 1 |

P(Y|X)

| | | |
|---|---|---|
| +y | +z | 1 |
| | -z | 0 |
| -y | +z | 0 |
| | -z | 1 |

P(Z|Y)

| | | | |
|---|---|---|---|
| +x | +y | +z | 0.5 |
| +x | +y | -z | 0 |
| +x | -y | +z | 0 |
| +x | -y | -z | 0 |
| -x | +y | +z | 0 |
| -x | +y | -z | 0 |
| -x | -y | +z | 0 |
| -x | -y | -z | 0.5 |

P(X,Y,Z)

P(+x, -z) = 0   P(+x) * P(-z) != 0

# Causal Chains with evidence

$$X \rightarrow Y \rightarrow Z$$

X: Low pressure    Y: Rain    Z: Traffic

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

- Given Y, X independent of Z ?

$$P(X|Z,y) = P(X|y) \quad ?$$

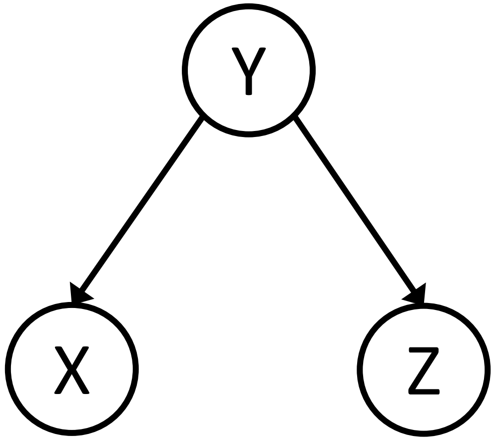
$$P(X|Z,y) = \frac{P(X,Z,y)}{P(Z,y)} = \frac{P(Z|y)P(y|X)P(X)}{\sum_x P(X,y,Z)} = \frac{P(Z|y)P(y|X)P(X)}{P(Z|y)\sum_x P(y|x)P(x)} = \frac{P(y|X)P(X)}{\sum_x P(y|x)P(x)} = \frac{P(y|X)P(X)}{P(y)} = P(X|y)$$

*Yes!*

Evidence along the chain "blocks" the influence

# Common Cause w/o evidence

Y: Project due




X: Unhappy Z: Library full

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

| | |
|---|---|
| +y | 0.5 |
| -y | 0.5 |

P(X)

| | | |
|---|---|---|
| +y | +x | 1 |
| | -x | 0 |
| -y | +x | 0 |
| | -x | 1 |

P(X|Y)

| | | |
|---|---|---|
| +y | +z | 1 |
| | -z | 0 |
| -y | +z | 0 |
| | -z | 1 |

P(Z|Y)

- X independent of Z ? $P(+x, -z) = P(+x)P(-z)$?

  **NO! Give a counter-case**

- Project due causes both unhappy and library full

P(+x, -z) = 0 P(+x) * P(-z) != 0

| | | | |
|---|---|---|---|
| +x | +y | +z | 0.5 |
| +x | +y | -z | 0 |
| +x | -y | +z | 0 |
| +x | -y | -z | 0 |
| -x | +y | +z | 0 |
| -x | +y | -z | 0 |
| -x | -y | +z | 0 |
| -x | -y | -z | 0.5 |

P(X, Y, Z)

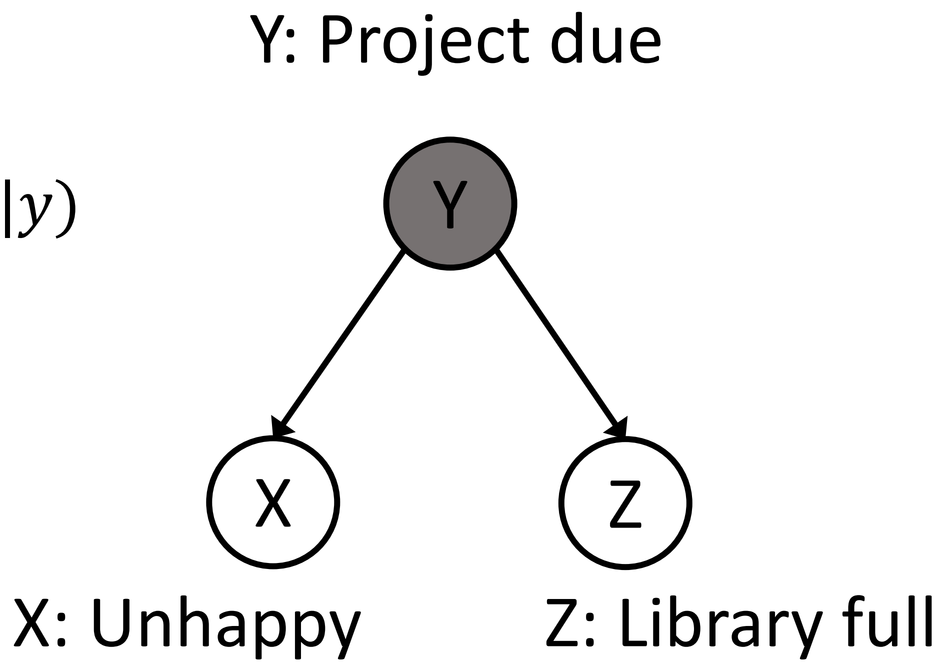
# Common Cause with evidence

- X and Z independent given Y ?

$$P(X|Z,y) = P(X|y) \quad ?$$

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

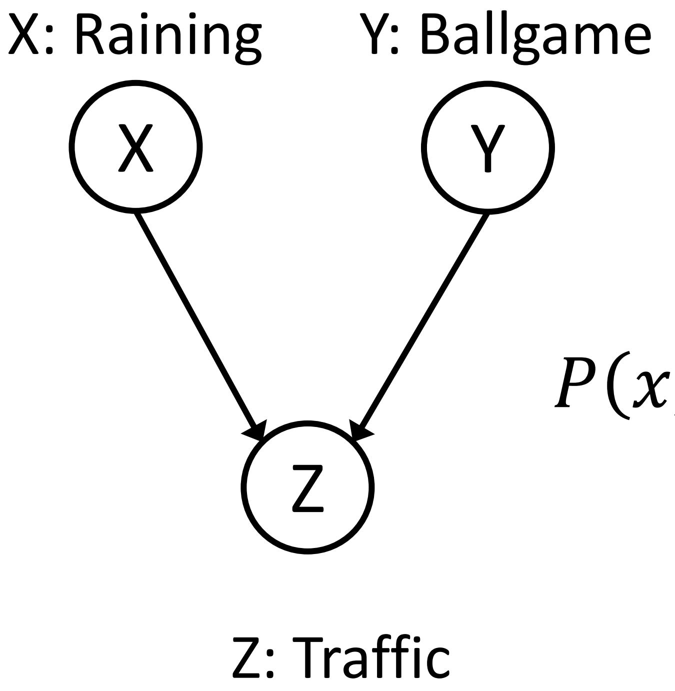
*Yes!*




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

Observing the cause blocks influence between effects.

# Common Effect w/o evidence




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

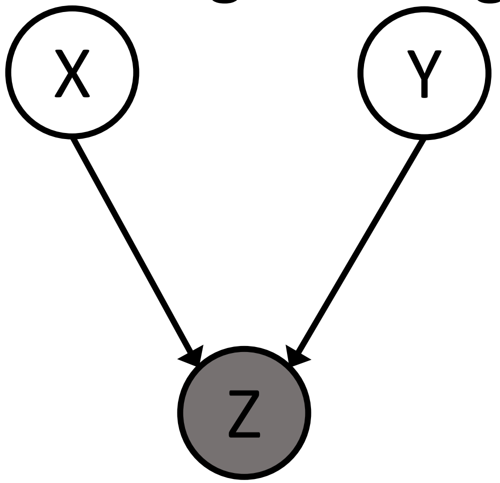
- Are X and Y independent? $P(x,\mathrm{y}) = P(\mathrm{x})P(\mathrm{y})$?
- *Yes*

$$P(X,Y) = \sum_Z P(X,Y,Z) = P(X,Y,+z) + P(Z,Y,-z) = P(X)P(Y)\sum_Z P(Z|X,Y) = P(X)P(Y)$$

# Common Effect with evidence

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

X: Raining Y: Ballgame




Z: Traffic

| | |
|---|---|
| +y | 0.5 |
| -y | 0.5 |

P(Y)

| | |
|---|---|
| +x | 0.5 |
| -x | 0.5 |

P(X)

| | | | |
|---|---|---|---|
| +x | +y | +z | 1 |
| +x | +y | -z | 0 |
| +x | -y | +z | 0 |
| +x | -y | -z | 1 |
| -x | +y | +z | 0 |
| -x | +y | -z | 1 |
| -x | -y | +z | 1 |
| -x | -y | -z | 0 |

P(Z|X,Y)

| | | | |
|---|---|---|---|
| +x | +y | +z | 0.25 |
| +x | +y | -z | 0 |
| +x | -y | +z | 0 |
| +x | -y | -z | 0.25 |
| -x | +y | +z | 0 |
| -x | +y | -z | 0.25 |
| -x | -y | +z | 0.25 |
| -x | -y | -z | 0 |

P(X,Y,Z)

- Given Z, X and Y independent? $P(X|Y,z) = P(X|z)$ ?
  - *No*: seeing traffic puts the rain and the ballgame in competition as explanation

**P(+x|+z) = 0.5, P(+x|+y, +z) = 1**

Observing an effect activates influence between possible causes.

# Outline

- Bayes’ Net
- Bayes’ Net Semantics
- Independence of Bayes’ Net
- D-separation
- D-separation – Demonstration

# Formalization of the independence question in BNs net

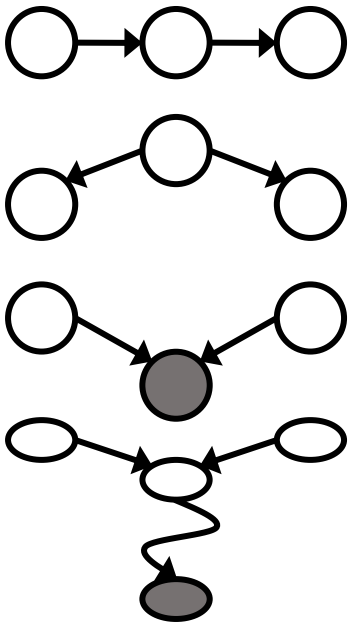
- Question: Are X and Y conditionally independent given evidence variables {Z}?
- D separation: look for paths in the resulting graph
  - Shade evidence nodes
  - **Consider all paths** from X to Y
  - No **active paths** = independence

- **Active** Path means **not independent**.
  - **Break path down to triples to check the independence.**

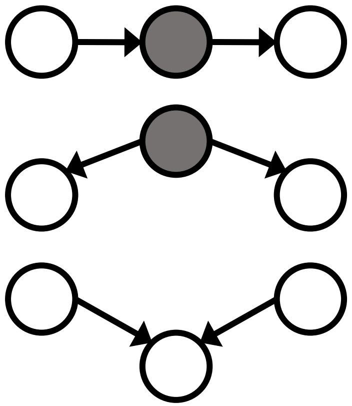
# Active / Inactive Paths

- **A path is active if every triple is active:**
  - Causal chain $A \rightarrow B \rightarrow C$
    - B is **unobserved**, A is dependent with C
    - Otherwise,  A is independent with C
  - Common cause $A \leftarrow B \rightarrow C$
    - B is **unobserved**, A is dependent with C
    - Otherwise,  A is independent with C
  - Common effect $A \rightarrow B \leftarrow C$
    - B *or one of its descendents* is **observed**, A is dependent with C
    - Otherwise,  A is independent with C
- All it takes to block a path is a single inactive segment

Active Triples

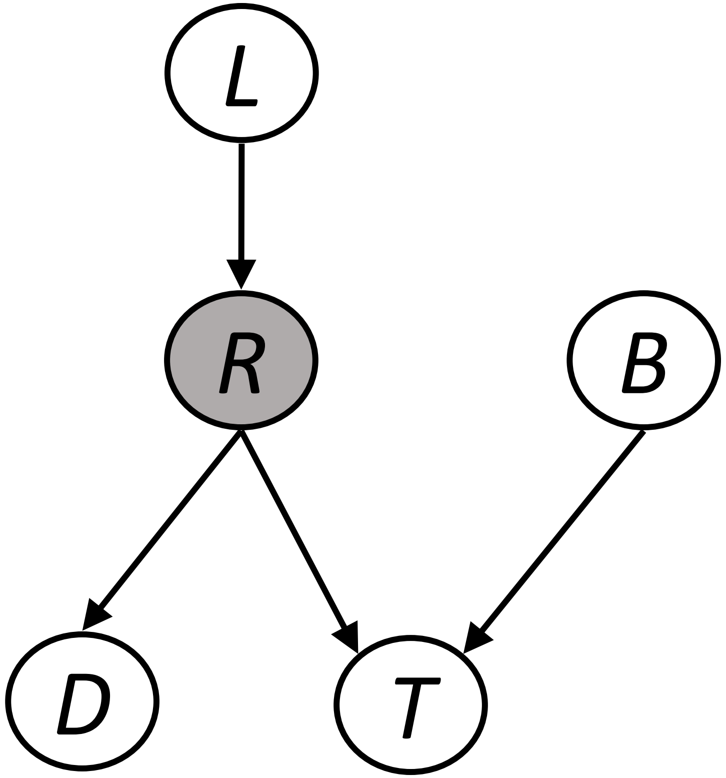


Inactive Triples



# Examples of D-separation

- Given R, is L independent with B?
  - **Paths**: L->R*->T<-B
  - **Triples:** L->R*->T; R*->T<-B




$L \perp T | R$ ? ⇒ Independent = inactive

$B \perp T | R$ ? ⇒ dependent = active

⇒ Inactive path ⇒ $L \perp B | R$ ***Yes!***

- **Break it down to triples and see if independence holds.**

# 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 by the BN.

# Markov Blanket

- A node is conditionally independent of all other nodes in the network, given its parents, children, and children's’ parents.

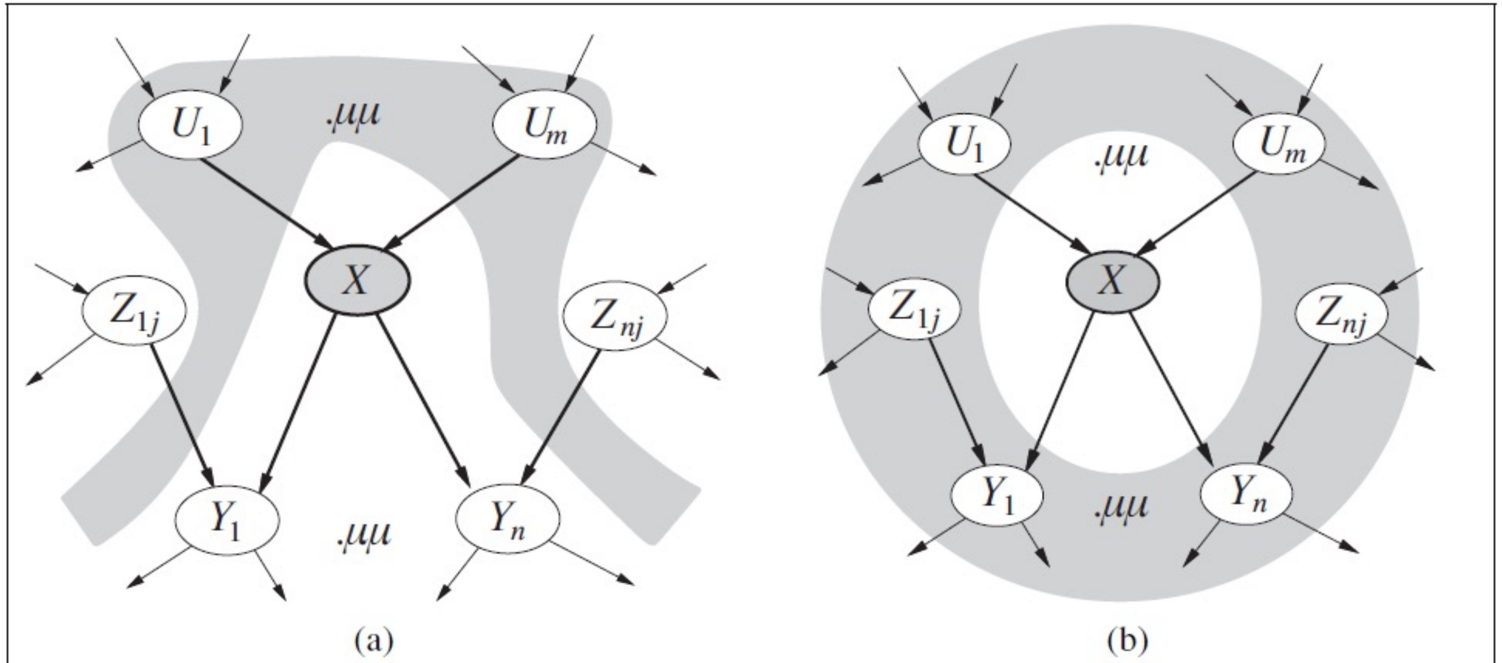
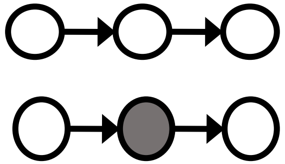



**Figure 14.4** (a) A node $X$ is conditionally independent of its non-descendants (e.g., the $Z_{ij}$s) given its parents (the $U_i$s shown in the gray area). (b) A node $X$ is conditionally independent of all other nodes in the network given its Markov blanket (the gray area).

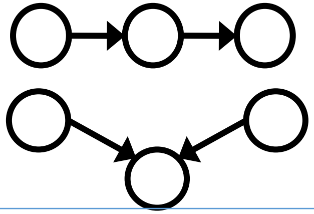
# Exercise :

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



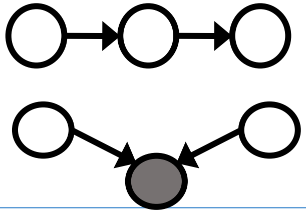
*Yes*

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



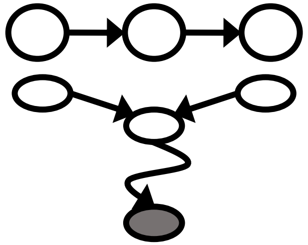
*Yes*

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



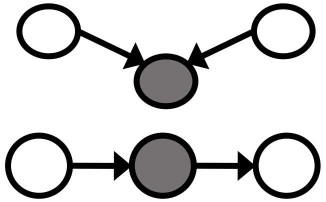
*No*

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

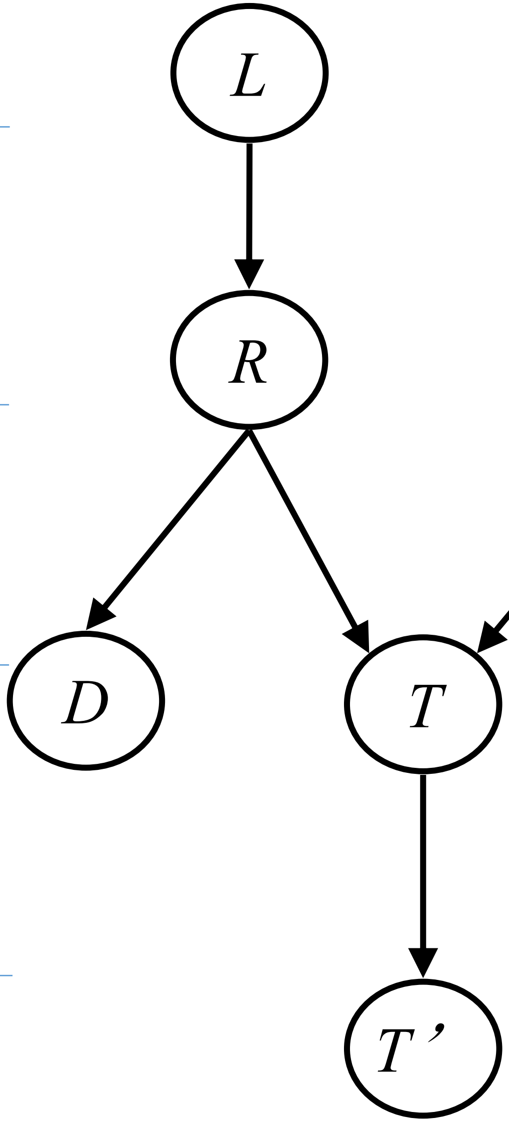


*No*

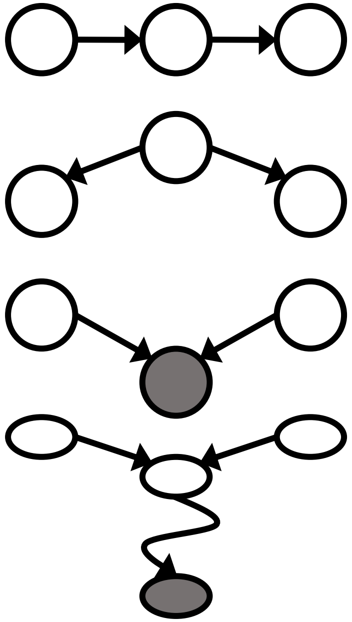
$L \perp\!\!\!\perp B | T, R$



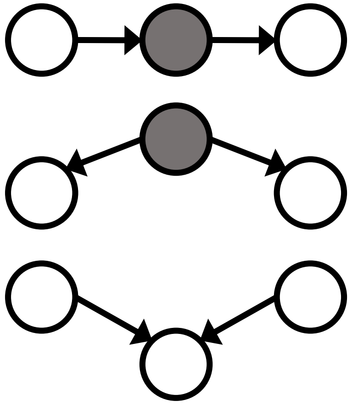
*Yes*




Active Triples



Inactive Triples

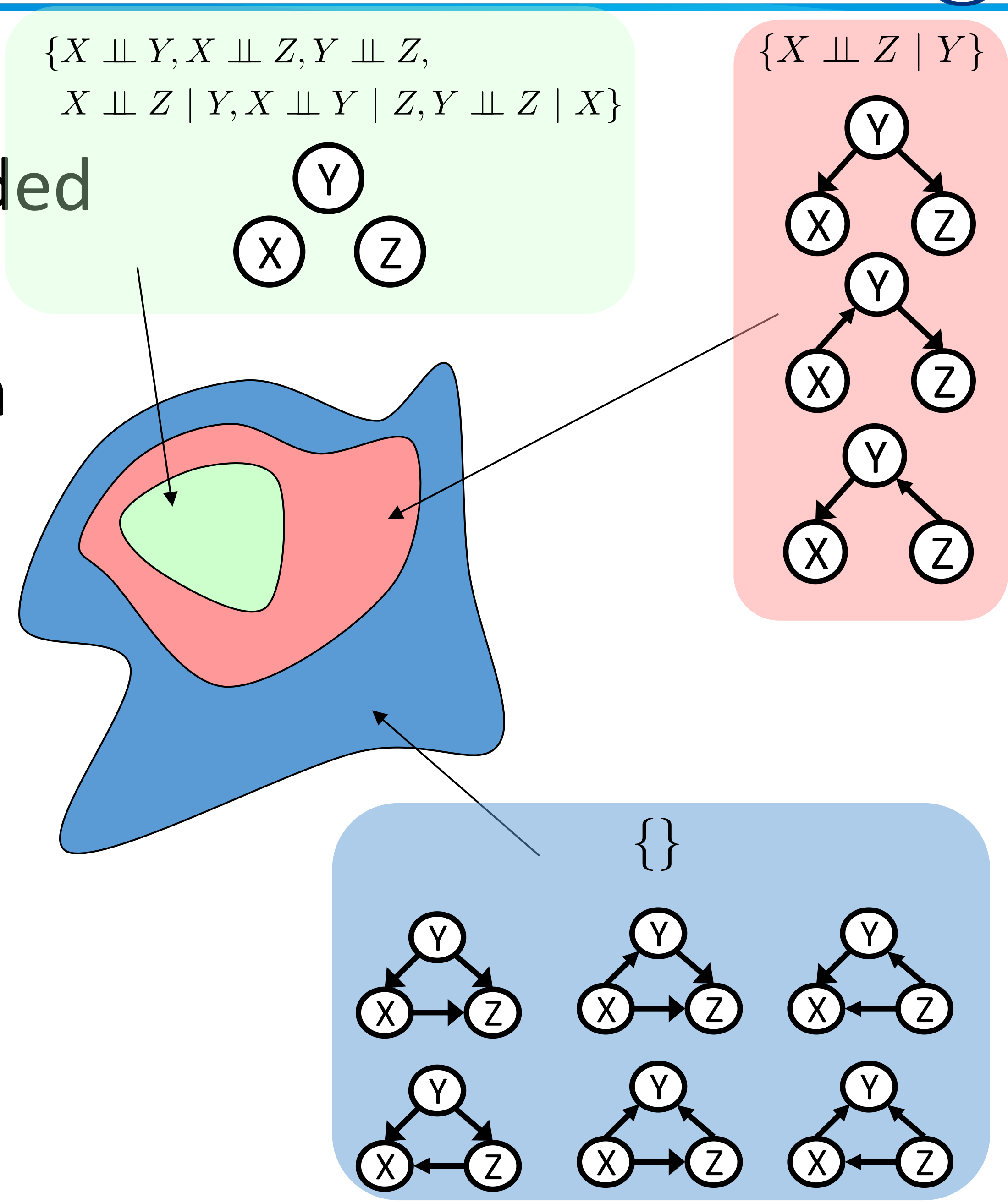


# Outline

- Bayes’ Net
- Bayes’ Net Semantics
- Independence of Bayes’ Net
- D-separation
- D-separation – Demonstration
- Structure Implications of Bayes’ Net

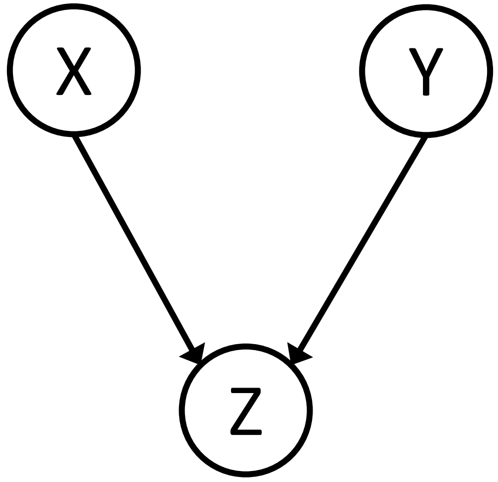
# Topology Limits Distributions

- Given some graph topology G, only certain joint distributions can be encoded
- The graph structure guarantees certain (conditional) independences
- Adding arcs increases the set of distributions, but has several costs
- Full conditioning can encode any distribution



# Structure Implications Example

X: Raining   Y: Ballgame




Can the graphical model encode the joint distribution in the table?

Z: Traffic

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

| | | | |
|---|---|---|---|
| +x | +y | +z | 0.15 |
| +x | +y | -z | 0.25 |
| +x | -y | +z | 0.05 |
| +x | -y | -z | 0.05 |
| -x | +y | +z | 0.05 |
| -x | +y | -z | 0.2 |
| -x | -y | +z | 0.01 |
| -x | -y | -z | 0.24 |

**P(X,Y,Z)**

**P(+x) = 0.5; P(+y) = 0.65**

**P (+x, +y) = 0.325**

**No!**

**The graph imply the dependence of X and Y.**

**However, X and Y are not independent in the table.**

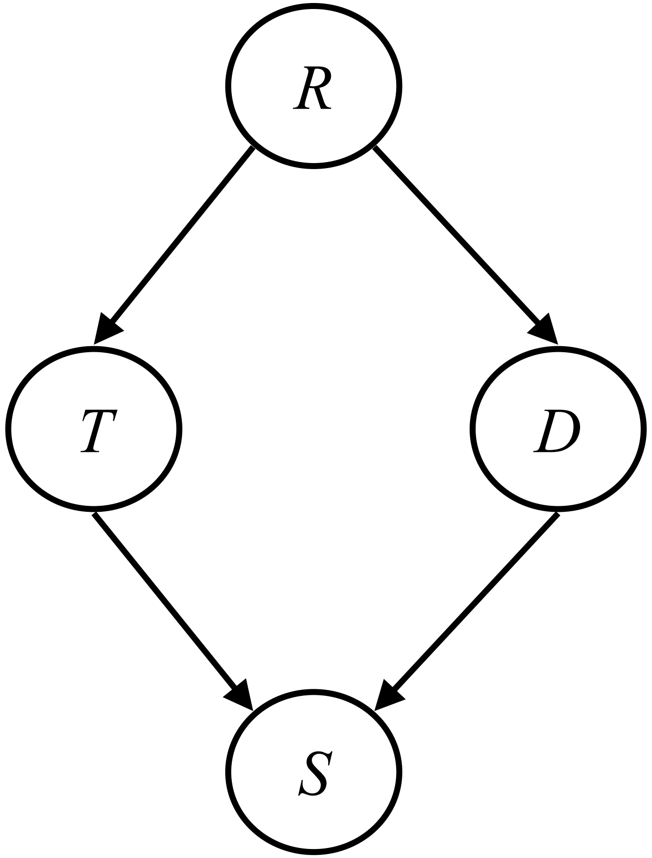
# More Examples for D-Separation

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

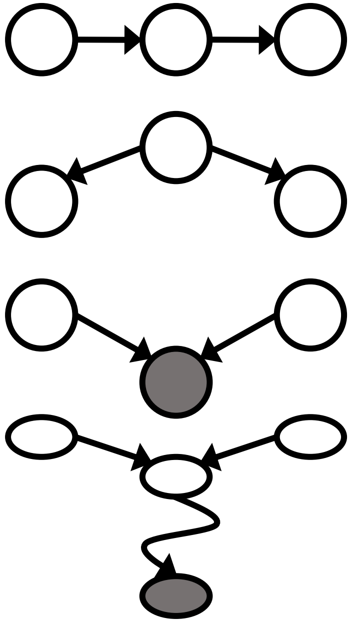
$T \perp\!\!\!\perp D$ No

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

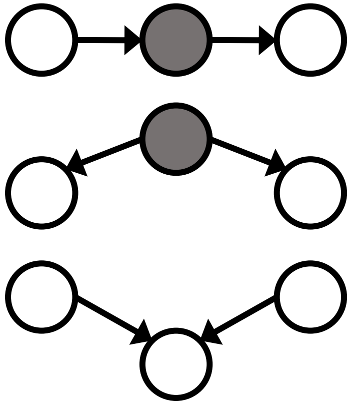
$T \perp\!\!\!\perp D | R, S$ No




Active Triples

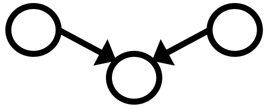


Inactive Triples

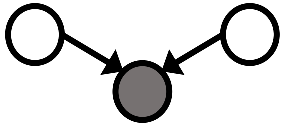


# More Examples for D-Separation

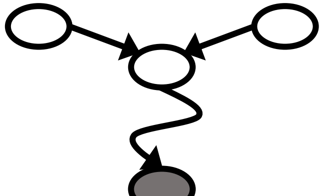
$R \perp\!\!\!\perp B$ *Yes*

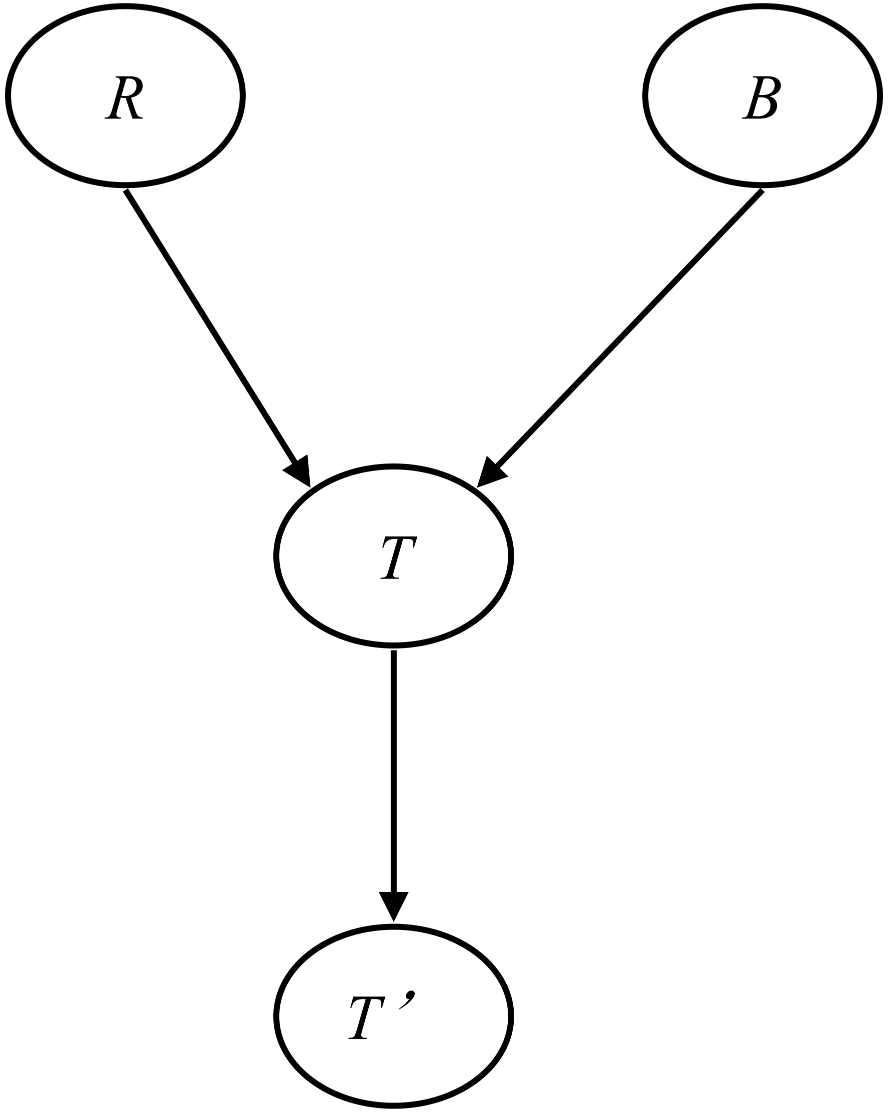


$R \perp\!\!\!\perp B | T$ *No*

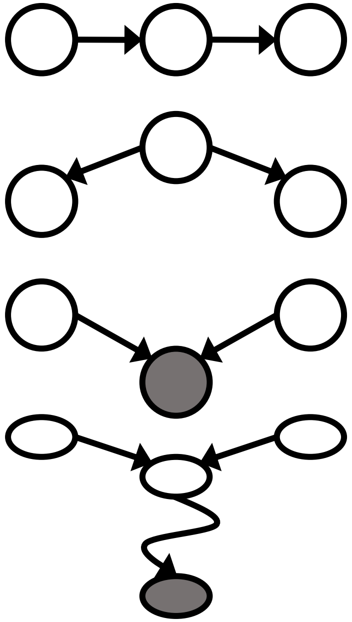


$R \perp\!\!\!\perp B | T'$ *No*






Active Triples



Inactive Triples