- http://cacm.acm.org/blogs/blog-cacm/138907-john-mccarthy/fulltext

# The Chain Rule

$$P(X_1, X_2, \ldots X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\ldots$$

- Trivial decomposition:

$$P(\mathsf{Traffic}, \mathsf{Rain}, \mathsf{Umbrella}) =$$
$$P(\mathsf{Rain})P(\mathsf{Traffic}|\mathsf{Rain})P(\mathsf{Umbrella}|\mathsf{Rain}, \mathsf{Traffic})$$

- With assumption of conditional independence:

$$P(\mathsf{Traffic}, \mathsf{Rain}, \mathsf{Umbrella}) =$$
$$P(\mathsf{Rain})P(\mathsf{Traffic}|\mathsf{Rain})P(\mathsf{Umbrella}|\mathsf{Rain})$$

- Bayes' nets / graphical models help us express conditional independence assumptions

# Conditional Independence

- Reminder: independence
  - X and Y are independent if

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

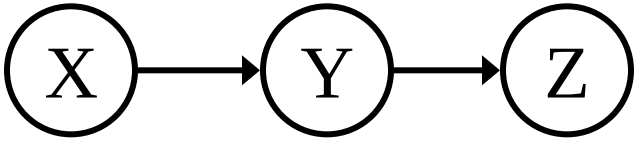
  - X and Y are conditionally independent given Z

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

  - (Conditional) independence is a property of a distribution

# Independence in a BN

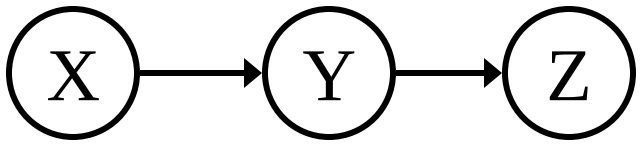
- Important question about a BN:
  - Are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counter example
  - Example:




  - 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)
    - Addendum: they *could* be independent: how?

# Causal Chains

- This configuration is a "causal chain"




X: Low pressure

Y: Rain

Z: Traffic

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

  - Is X independent of Z given Y?

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

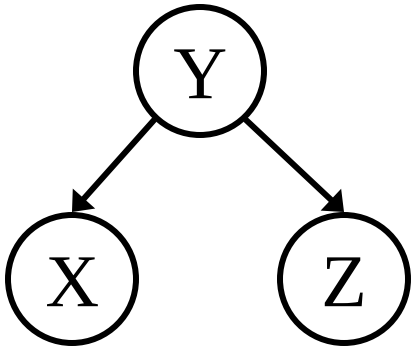
$$= P(z|y)$$

*Yes!*

  - Evidence along the chain "blocks" the influence

# Common Cause

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




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

$$= P(z|y)$$

*Yes!*

Y: Project due

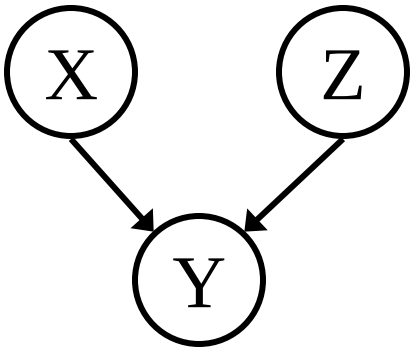
X: Newsgroup busy

Z: Lab full

  - Observing the cause blocks influence between effects.

# Common Effect

- 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
    - Still need to prove they must be (try it!)
  - 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

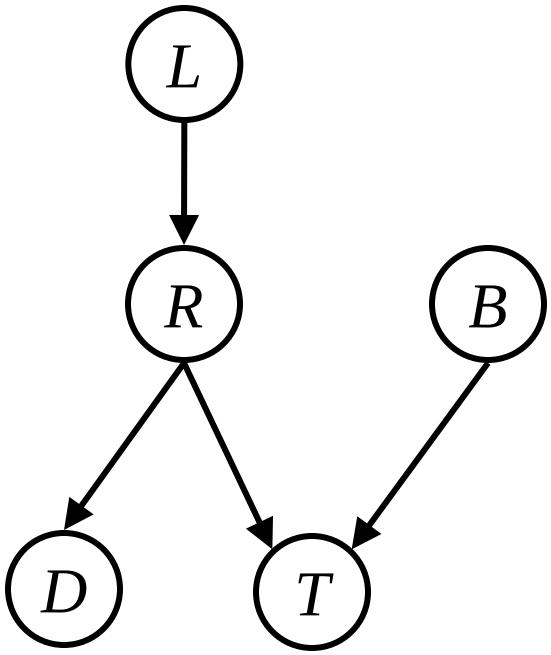
Y: Traffic

# The General Case

- 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

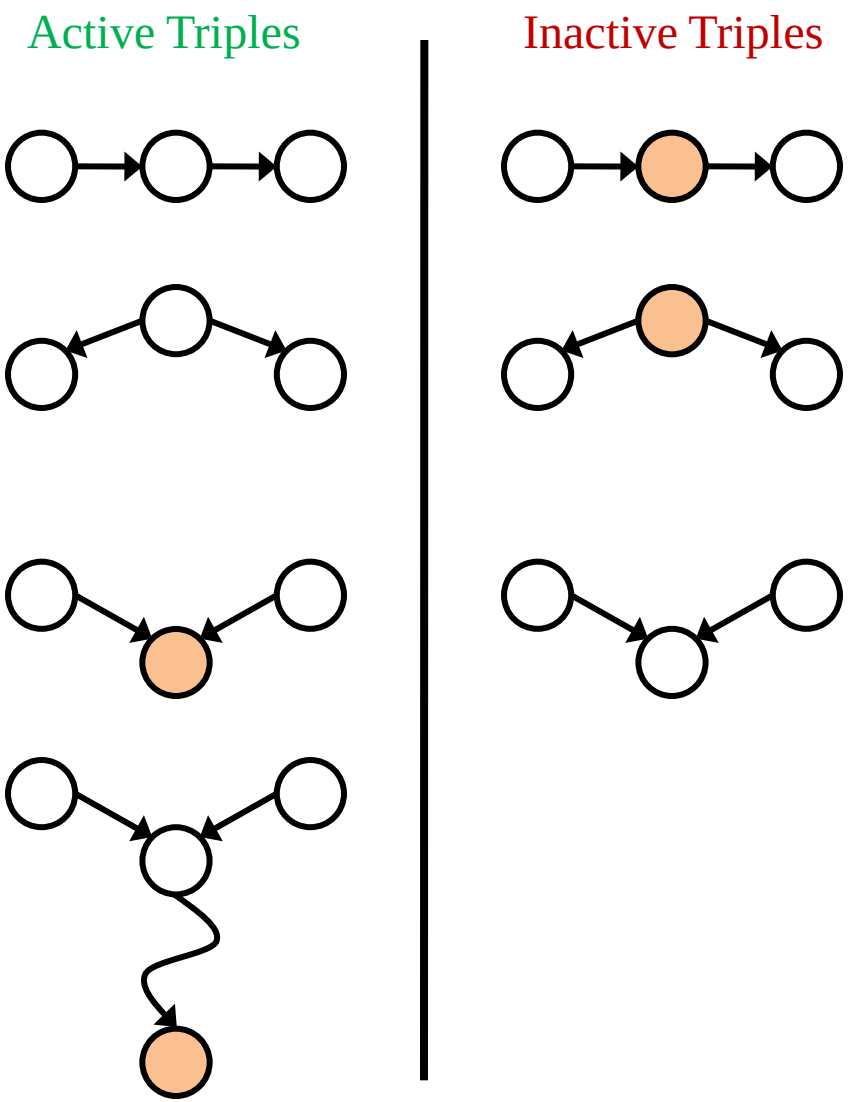
# Reachability

- Recipe: shade evidence nodes
- 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"



# Reachability (D-Separation)

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

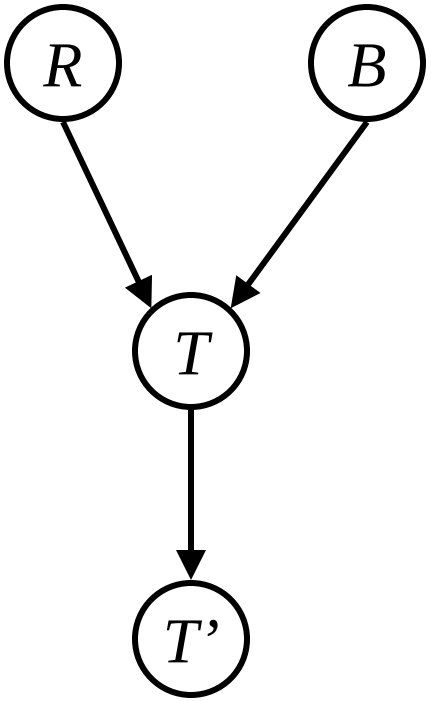


# Example

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

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

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



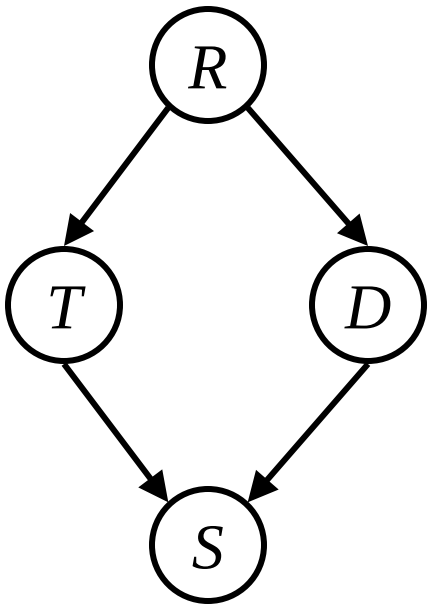
# Example

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

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

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

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

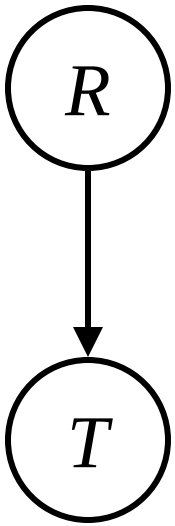


# Causality?

- 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
  - 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 only guaranteed to encode conditional independence

# Example: Traffic

- Basic traffic net
- Let's multiply out the joint




$P(R)$

| | |
|---|---|
| r | 1/4 |
| ¬r | 3/4 |

$P(T|R)$

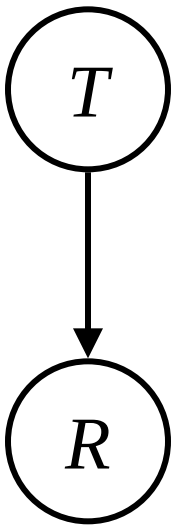
| | | |
|---|---|---|
| r | t | 3/4 |
| | ¬t | 1/4 |
| ¬r | t | 1/2 |
| | ¬t | 1/2 |

$P(T, R)$

| | | |
|---|---|---|
| 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 |

$P(R|T)$

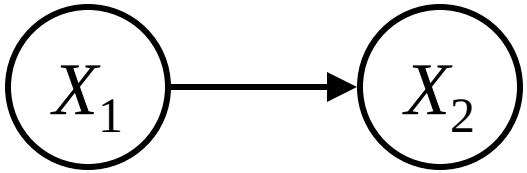
| | | |
|---|---|---|
| t | r | 1/3 |
| | ¬r | 2/3 |
| ¬t | r | 1/7 |
| | ¬r | 6/7 |

$P(T, R)$

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

# Example: Coins

- Extra arcs don't prevent representing independence, just allow non-independence


$P(X_1)$

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

$P(X_2)$

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

$P(X_1)$

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

$P(X_2|X_1)$

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

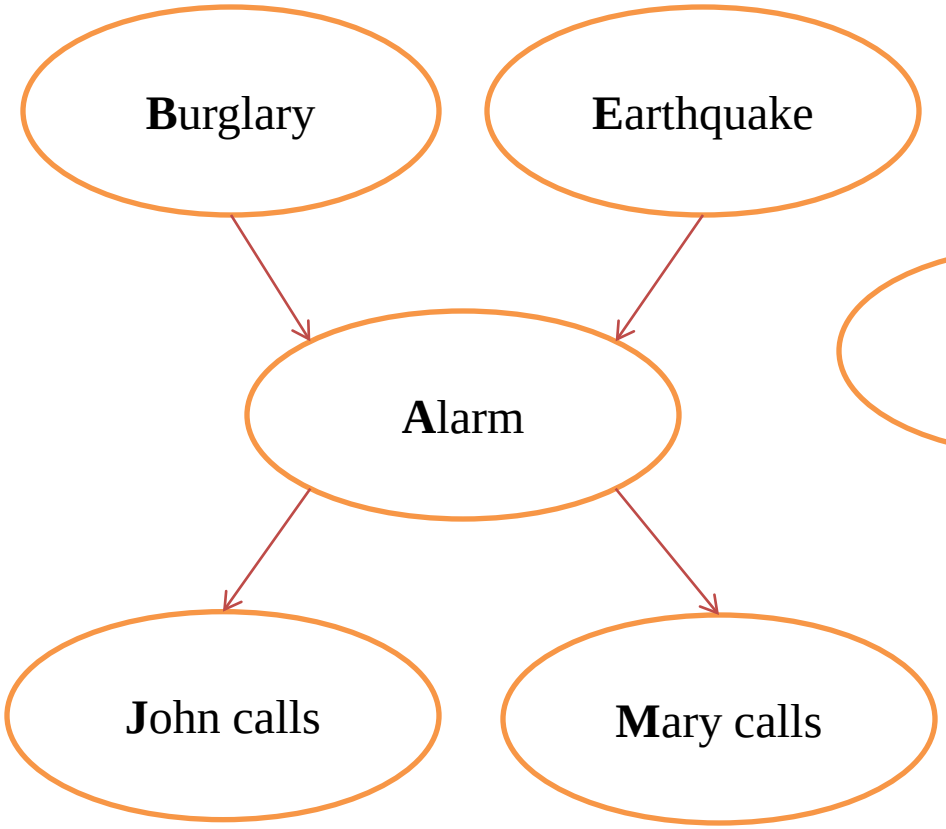
- Adding unneeded arcs isn't wrong, it's just inefficient

# Changing Bayes' Net Structure

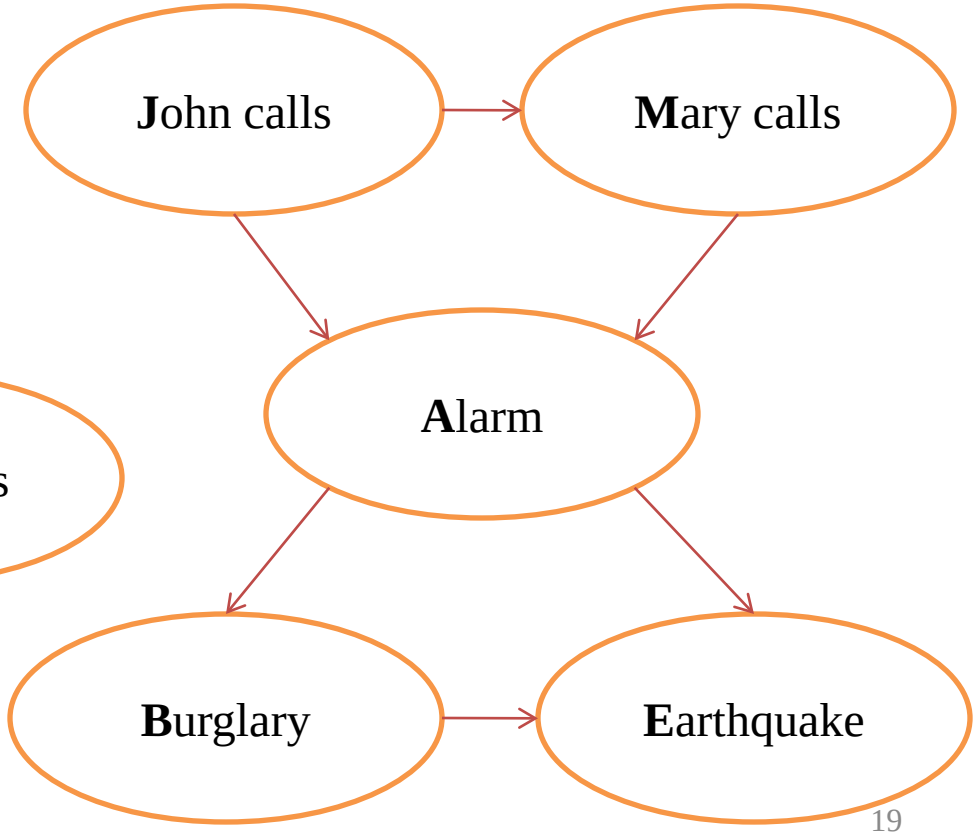
- The same joint distribution can be encoded in many different Bayes' nets
  - Causal structure tends to be the simplest

- Analysis question: given some edges, what other edges do you need to add?
  - One answer: fully connect the graph
  - Better answer: don't make any false conditional independence assumptions

# Example: Alternate Alarm




If we reverse the edges, we make different conditional independence assumptions




To capture the same joint distribution, we have to add more edges to the graph

# 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