Bayesian networks

Chapter 14

Section 1 – 2

Outline

- Syntax
- Semantics

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 called graphical models
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions

Bayesian networks

- A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

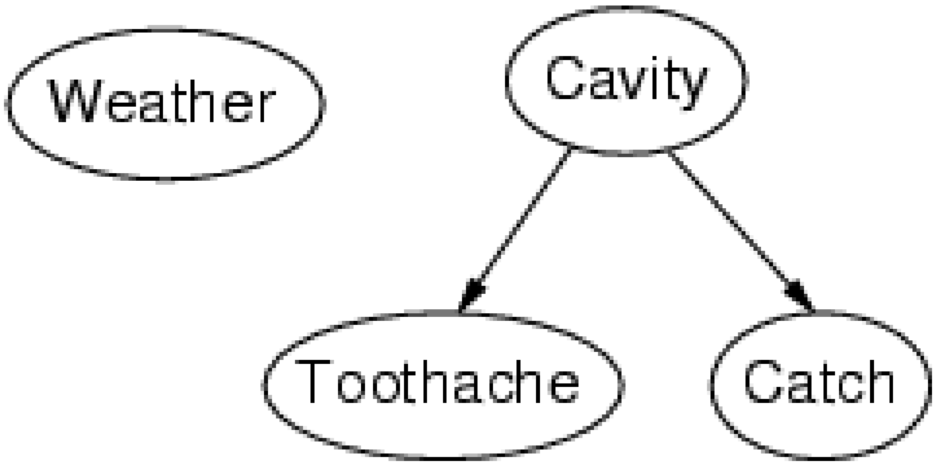
- Syntax:
  - a set of nodes, one per variable
  -
  - a directed, acyclic graph (link ≈ "directly influences")
  - a conditional distribution for each node given its parents:

$$\mathbf{P}\,(X_i \mid Parents\,(X_i))$$

- In the simplest case, conditional distribution represented as a conditional probability table (CPT) giving the distribution over X_i for each combination of parent values

Example

- Topology of network encodes conditional independence assertions:



- *Weather* is independent of the other variables
- *Toothache* and *Catch* are conditionally independent given *Cavity*

Example: Coin Flips

- N independent coin flips

- No interactions between variables: absolute independence

Example: Coin Flips

$P(X_1)$

h	0.5
t	0.5

$P(X_2)$

h	0.5
t	0.5

. . .

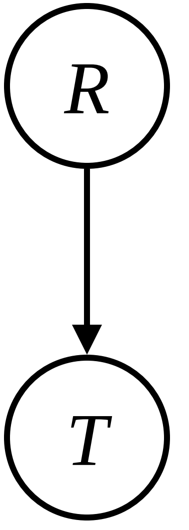
$P(X_n)$

h	0.5
t	0.5

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

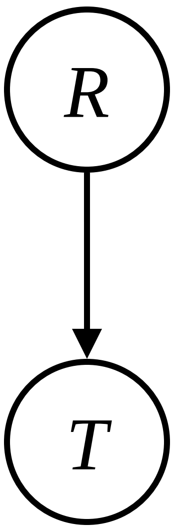
Example: Traffic

- Variables:
  - R: It rains
  - T: There is traffic



- Model 1: independence
- Model 2: rain causes traffic
- Why is an agent using model 2 better?

Example: Traffic



$P(R)$

+r	1/4
¬r	3/4

$P(T|R)$

+r →

+t	3/4
¬t	1/4

¬r →

+t	1/2
¬t	1/2

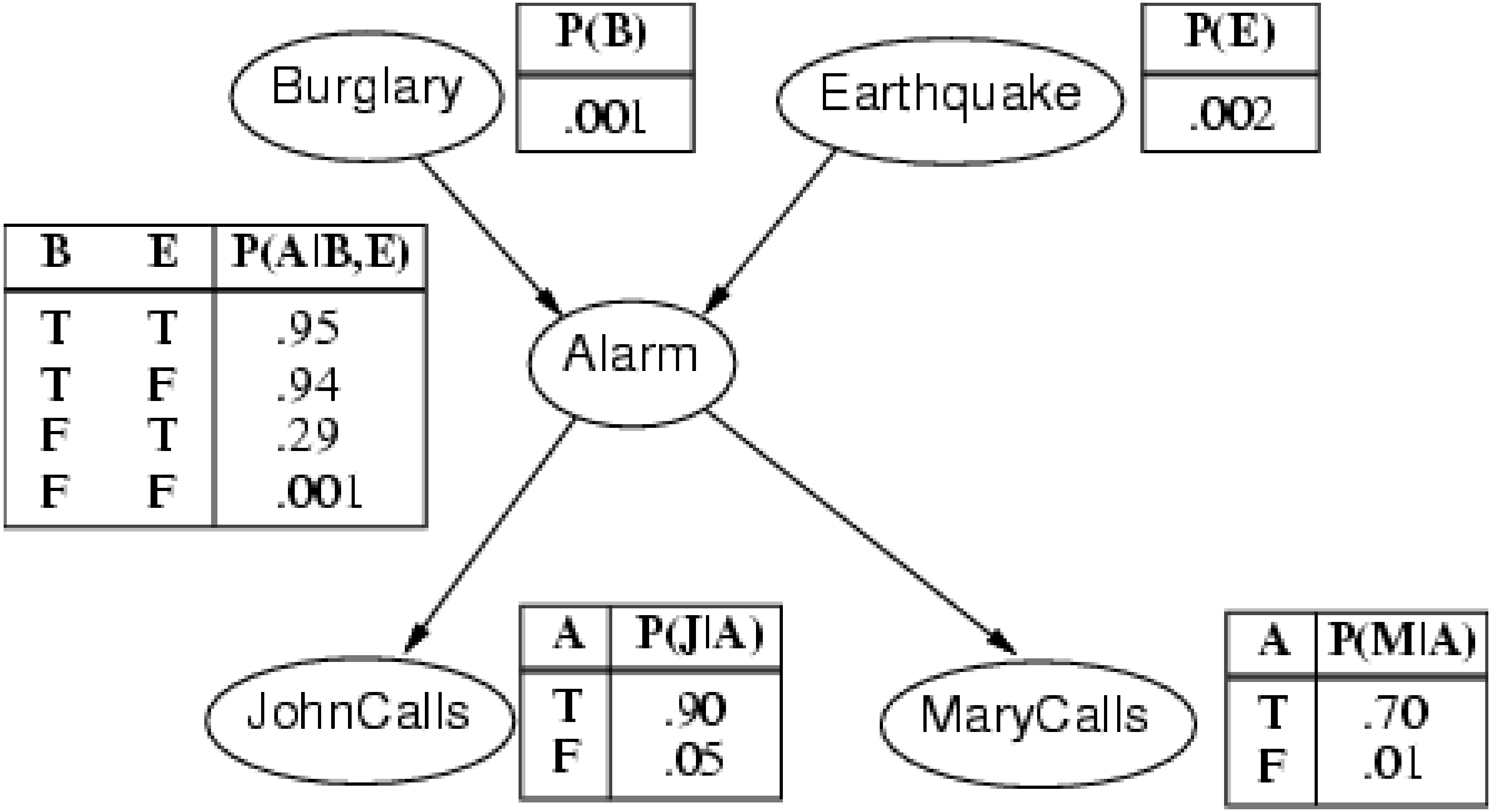
Example

- 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: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*

- Network topology reflects "causal" knowledge:
  - 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

Example contd.



P(B)
.001

P(E)
.002

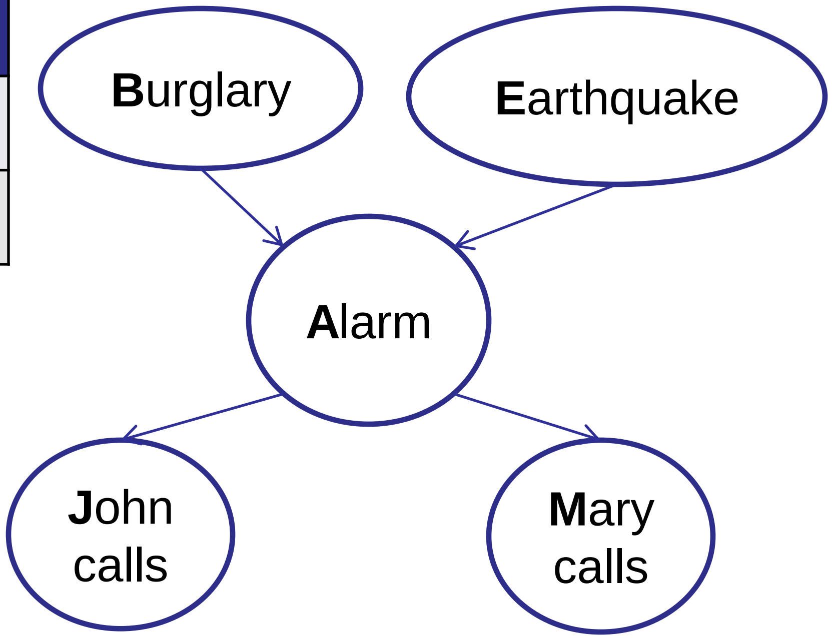
B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

Slightly different notation

B	P(B)
+b	0.001
¬b	0.999



E	P(E)
+e	0.002
¬e	0.998

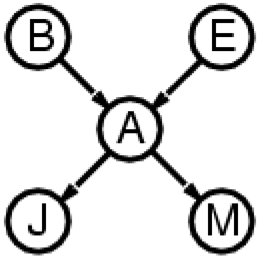
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	¬a	0.05
+b	¬e	+a	0.94
+b	¬e	¬a	0.06
¬b	+e	+a	0.29
¬b	+e	¬a	0.71
¬b	¬e	+a	0.001
¬b	¬e	¬a	0.999

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

Compactness

- A CPT for Boolean X_i with k Boolean parents has 2^k rows for the combinations of parent values



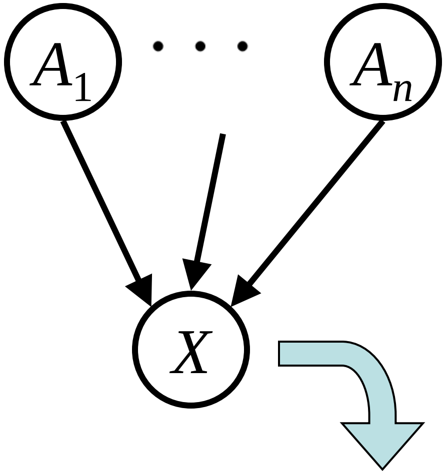
- Each row requires one number p for X_i = *true* (the number for X_i = *false* is just $1-p$)
- If each variable has no more than k parents, the complete network requires $O(n \cdot 2^k)$ numbers
- I.e., grows linearly with n, vs. $O(2^n)$ for the full joint distribution
- For burglary net, 1 + 1 + 4 + 2 + 2 = 10 numbers (vs. $2^5-1 = 31$)
- BNs: Huge space savings
- Also easier to elicit local CPTs
- Also turns out to be faster to answer queries (coming)

Bayes' Net Semantics

- Let's formalize the semantics of a Bayes' net
- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents' values

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

  - CPT: conditional probability table
  - Description of a noisy "causal" process

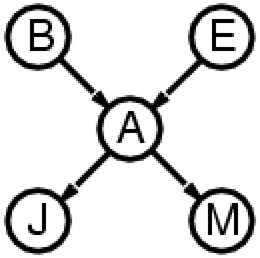


$$P(X|A_1 \ldots A_n)$$

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

Semantics

The full joint distribution is defined as the product of the local conditional distributions:



$$\mathbf{P}(X_1, \ldots, X_n) = \pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i))$$

e.g., $\mathbf{P}(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$= \mathbf{P}(j \mid a)\, \mathbf{P}(m \mid a)\, \mathbf{P}(a \mid \neg b, \neg e)\, \mathbf{P}(\neg b)\, \mathbf{P}(\neg e)$

To emphasize: every BN over a domain implicitly defines a joint distribution over that domain, specified by local probabilities and graph structure

Constructing Bayesian networks

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

$$\mathbf{P}(X_i \mid Parents(X_i)) = \mathbf{P}(X_i \mid X_1, \ldots X_{i-1})$$

This choice of parents guarantees:

$$\mathbf{P}(X_1, \ldots, X_n) = \pi_{i=1}^{n} \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1}) \text{ (chain rule)}$$

$$= \pi_{i=1}^{n} \mathbf{P}(X_i \mid Parents(X_i)) \text{ (by construction)}$$

Causality?

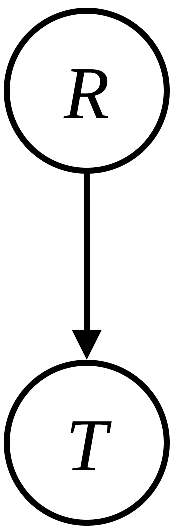
- When Bayes' nets reflect the true causal patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about
  - Often easier to elicit from experts

- BNs need not actually be causal
  - Sometimes no causal net exists over the domain
  - 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

$P(T|R)$

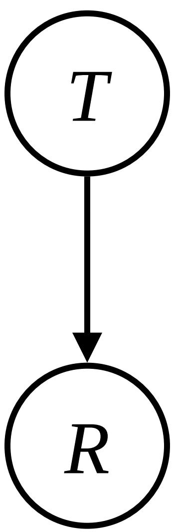
r	t	3/4
	¬t	1/4
¬r	t	1/2
	¬t	1/2

$P(T, R)$

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

$P(T, R)$

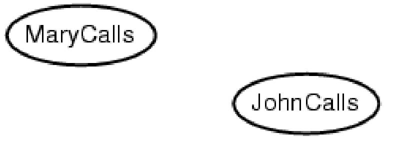
r	t	3/16
r	¬t	1/16
¬r	t	6/16
¬r	¬t	6/16

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

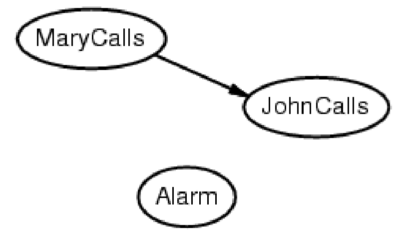
Example

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

$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

Example

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



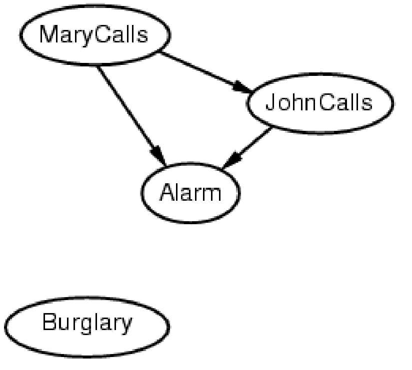
$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

No

$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$?

Example

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



$\mathbf{P}(J \mid M) = \mathbf{P}(J)$?

No

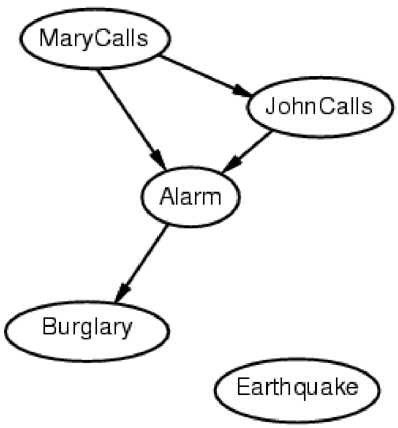
$\mathbf{P}(A \mid J, M) = \mathbf{P}(A \mid J)$? $\mathbf{P}(A \mid J, M) = \mathbf{P}(A)$? **No**

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B \mid A)$?

$\mathbf{P}(B \mid A, J, M) = \mathbf{P}(B)$?

Example

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



P(*J* | *M*) = ***P***(*J*)?

No

P(*A* | *J*, *M*) = ***P***(*A* | *J*)? ***P***(*A* | *J*, *M*) = ***P***(*A*)? **No**

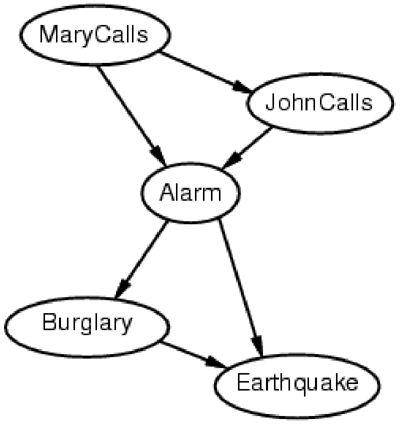
P(*B* | *A*, *J*, *M*) = ***P***(*B* | *A*)? **Yes**

P(*B* | *A*, *J*, *M*) = ***P***(*B*)? **No**

P(*E* | *B*, *A* ,*J*, *M*) = ***P***(*E* | *A*)?

Example

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



P(*J* | *M*) = ***P***(*J*)?

No

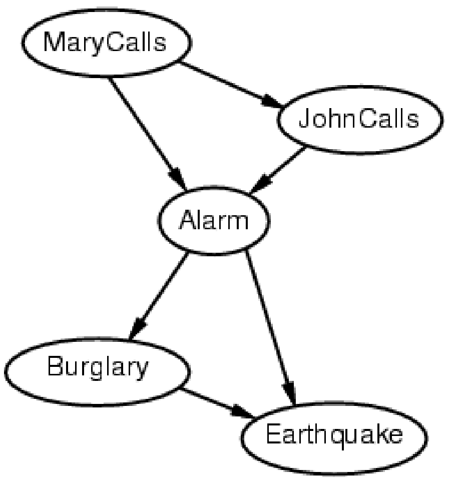
P(*A* | *J, M*) = ***P***(*A* | *J*)? ***P***(*A* | *J, M*) = ***P***(*A*)? **No**

P(*B* | *A, J, M*) = ***P***(*B* | *A*)? **Yes**

P(*B* | *A, J, M*) = ***P***(*B*)? **No**

P(*E* | *B, A ,J, M*) = ***P***(*E* | *A*)? **No**

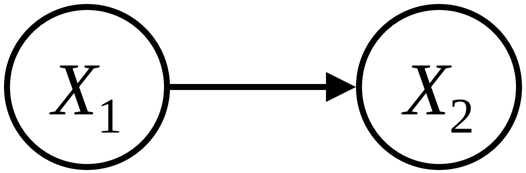
Example contd.



- Deciding conditional independence is hard in noncausal directions
-
- (Causal models and conditional independence seem hardwired for humans!)
-
- Network is less compact: 1 + 2 + 4 + 2 + 4 = 13 numbers needed
-

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

Summary

- Bayesian networks provide a natural representation for (causally induced) conditional independence
- Topology + CPTs = compact representation of joint distribution
- Generally easy for domain experts to construct

Bayes' Nets So Far

- We now know:
  - What a Bayes' net is
  - What joint distribution a Bayes' net encodes
- Briefly: properties of that joint distribution (independence)
  - Previously: assembled BNs using an intuitive notion of conditional independence as causality
  - Main goal: answer queries about conditional independence
- Next: how to compute posteriors quickly (inference)

Conditional Independence

- Reminder: independence
  - X and Y are independent if

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

  - X and Y are conditionally independent given Z

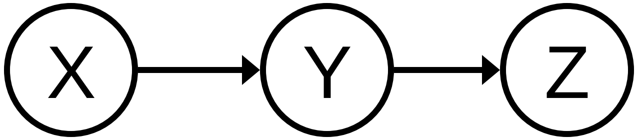
$$\forall x, y, z \;\; P(x, y|z) = P(x|z)P(y|z)$$

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

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?

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

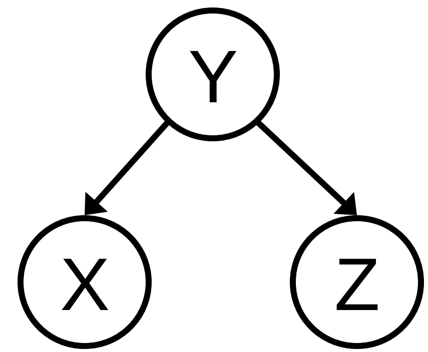
$$= P(z|y)$$

Yes!

  - Evidence along the chain “blocks” the influence

Common Cause

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

  - Observing the cause blocks influence between effects.

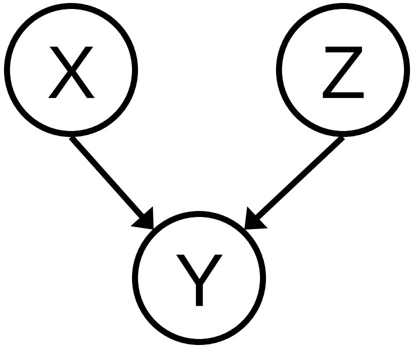
Y: Project due

X: Newsgroup busy

Z: Lab full

Common Effect

- Last configuration: two causes of one effect
  - 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