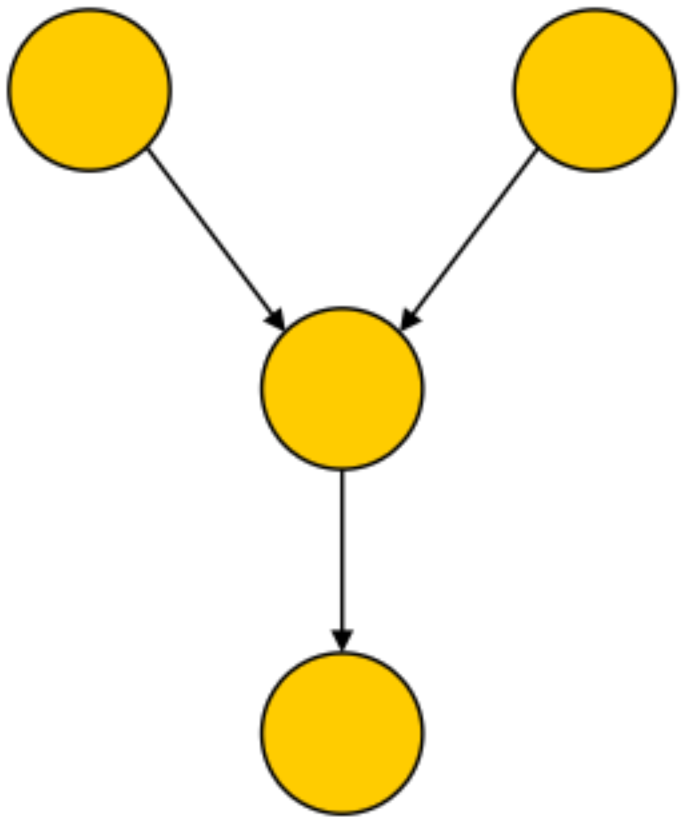
Bayes Net



Probability Primer

Probability Primer

- A random variable is the basic element of probability.
- Refers to an event where there is some degree of uncertainty as to the outcome of the event. For example, a coin toss.
- We will focus on Boolean random variables.
  Their outcome set can be represented as {0, 1}.
- Examples of Boolean rv, referred to as A:

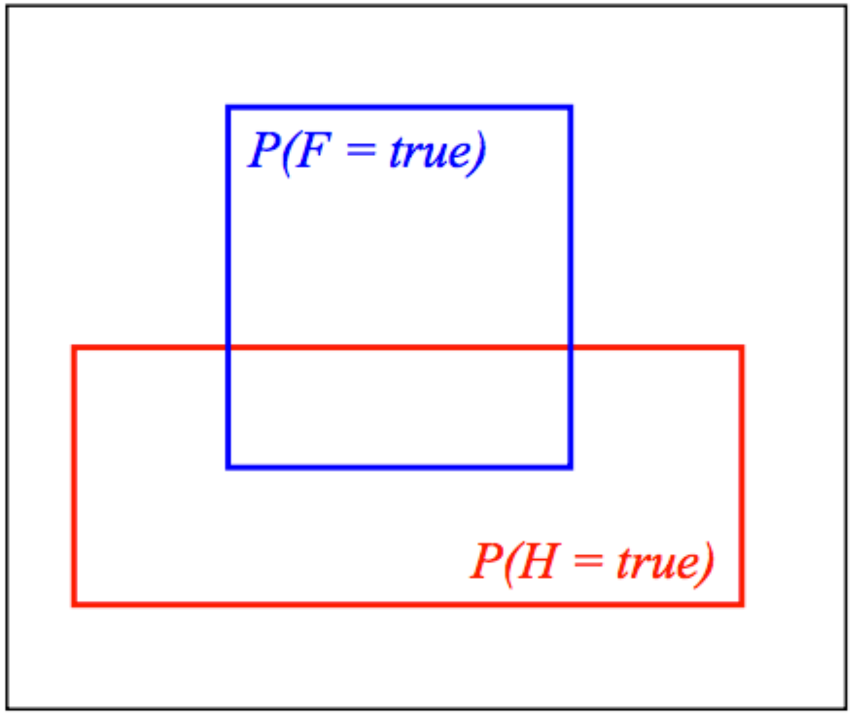
  A = Getting heads on a coin flip
  A = It will rain today
  A = Your team will win

Conditional Probability

- Out of all the outcomes in which B is true, how many also have A equal to true

$$P(A = \text{true} | B = \text{true})$$

- Read as: Probability of A conditioned on B or Probability of A given B



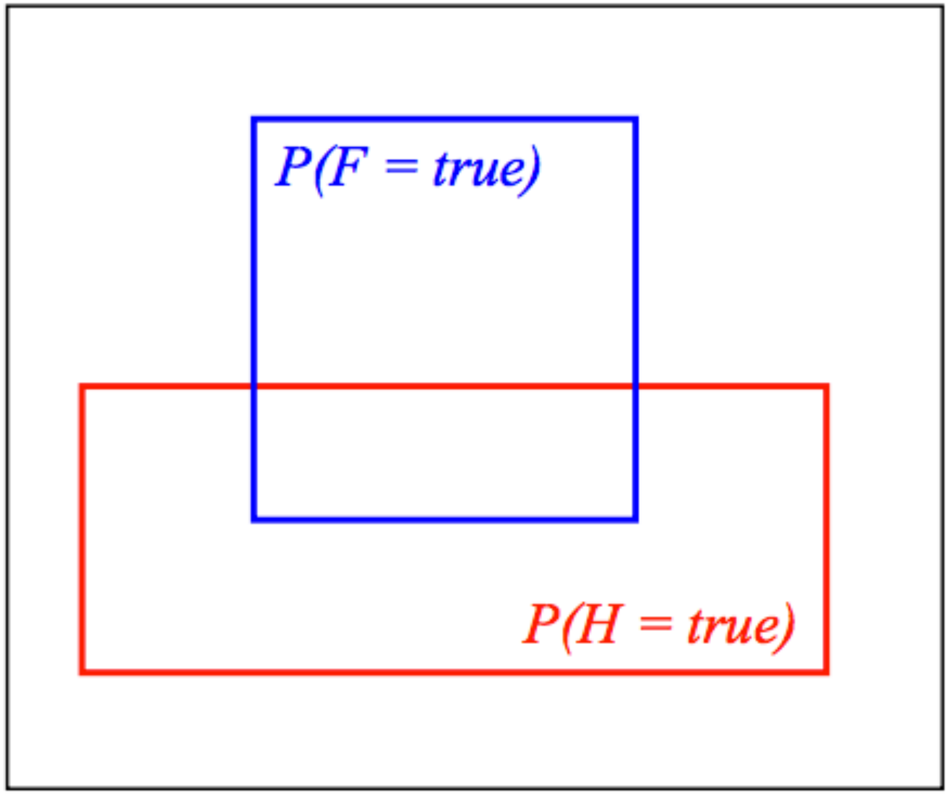
H = "Have a headache"
F = "Coming down with Flu"

$P(H = true) = 1/10$
$P(F = true) = 1/40$
$P(H = true \mid F = true) = 1/2$

"Headaches are rare and flu is rarer, but if you're coming down with flu there's a 50-50 chance you'll have a headache."

Joint Probability

- We will write P(A = true, B = true) to mean the joint probability of A = true and B = true.
- We know that $P(X|Y) = \frac{P(X,Y)}{P(Y)}$



$$P(H{=}true \mid F{=}true)$$

$$= \frac{\text{Area of "H and F" region}}{\text{Area of "F" region}}$$

$$= \frac{P(H = true, F = true)}{P(F = true)}$$

Joint Probability

- Joint probabilities can be between any number of variables
  eg. P(A = true, B = true, C = true)
- For each combination of variables, we need to say how probable that combination is
- The probabilities of these combinations need to sum to 1.
- How many entries will there be for joint distribution of n Booleans? 2^n
- How many of those will you need to figure out everything? $2^n - 1$

A	B	C	P(A,B,C)
false	false	false	0.1
false	false	true	0.2
false	true	false	0.05
false	true	true	0.05
true	false	false	0.3
true	false	true	0.1
true	true	false	0.05
true	true	true	0.15

Sums to 1

Joint Probability

- Use the table on the right to compute:

  P(A=true)

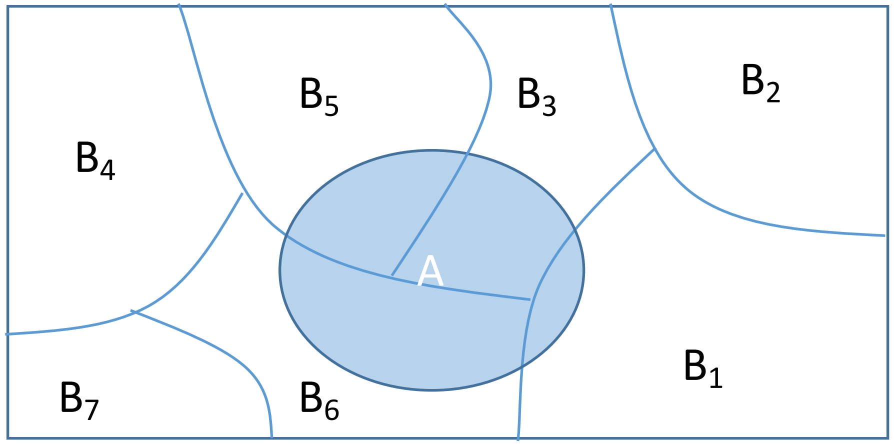
  P(A=true | C=true)

  P(A=true, B=true | C=true)

A	B	C	P(A,B,C)
false	false	false	0.1
false	false	true	0.2
false	true	false	0.05
false	true	true	0.05
true	false	false	0.3
true	false	true	0.1
true	true	false	0.05
true	true	true	0.15

$\underbrace{\hphantom{P(A,B,C)}}_{\text{Sums to 1}}$

Rule of total probability



$$p(A) = \sum P(B_i) P(A \mid B_i)$$

Marginal Probability

- Given two or more variables, the marginal probability of a random variable can be obtained by summing up over all joint probabilities.

- X can take values {x1, x2, x3, x4}, Y can take values {y1, y2,y3, y4}, Joint distribution is known.

- Marginal of any value of X can be calculated:

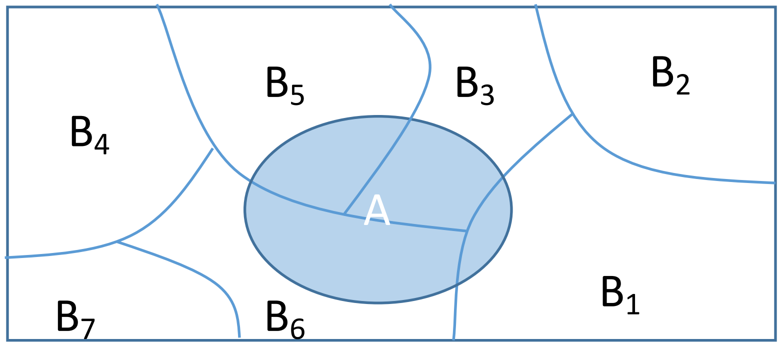
Y \ X	x_1	x_2	x_3	x_4	$p_y(Y)\downarrow$
y_1	$4/32$	$2/32$	$1/32$	$1/32$	$8/32$
y_2	$2/32$	$4/32$	$1/32$	$1/32$	$8/32$
y_3	$2/32$	$2/32$	$2/32$	$2/32$	$8/32$
y_4	$8/32$	0	0	0	$8/32$
$p_x(X) \rightarrow$	$16/32$	$8/32$	$4/32$	$4/32$	$32/32$

$$\Pr(X = x) = \sum_y \Pr(X = x, Y = y) = \sum_y \Pr(X = x \mid Y = y) \Pr(Y = y)$$

Marginalization (Also called conditioning)

- We know p(X,Y), what is P(X=x)?
- We can use the low of total probability, why?

$$p(x) = \sum_y P(x,y)$$
$$= \sum_y P(y)P(x \mid y)$$



Marginalization Cont.

- Another example

$$p(x) = \sum_{y,z} P(x, y, z)$$

$$= \sum_{z,y} P(y, z) P(x \mid y, z)$$

Marginalization Example

		temperature		
		Hot	Mild	Cold
weather	Sunny	0.10	0.20	0.10
	Cloudy	0.05	0.35	0.20

If we marginalize out *weather*, we get

	Hot	Mild	Cold
$P(temperature) =$	0.15	0.55	0.30

If we marginalize out *temperature*, we get

	Sunny	Cloudy
$P(weather) =$	0.40	0.60

Bayes Rule

You have seen this many times before

$$P(x \mid y) = \frac{P(x)P(y \mid x)}{P(y)}$$

Conditional Independence

Joint Probability

- To represent the joint probability of n Booleans, I need a table with 2^n rows filled.
- Need $2^n - 1$ parameters.
- What if we don't have enough data or resources to estimate that many parameters?
- Variable Independence to the rescue.

A	B	C	P(A,B,C)
false	false	false	0.1
false	false	true	0.2
false	true	false	0.05
false	true	true	0.05
true	false	false	0.3
true	false	true	0.1
true	true	false	0.05
true	true	true	0.15

Sums to 1

Independence

Variables A and B are independent if any of the following hold:

- $P(A, B) = P(A) \times P(B)$

- $P(A \mid B) = P(A)$

- $P(B \mid A) = P(B)$

This says that knowing the outcome of *A* does not tell me anything new about the outcome of *B*.

Independence

- How does independence help us?

- Suppose you flip n coins. If they were dependent, you would need a joint distribution. Example: If A = head, B = head, to find probability of C = heads, you would need to look up the table.

Coin A

Coin B

Coin C

If you assume that all the coin throws are independent, it makes it very easy.

$$P(A=H, B=H, C=H) = P(A=H)\ P(B=H)\ P(C=H)$$

Joint probability is simply product of individual probabilities.

Independence

- Suppose I have n Boolean events (C_1, C_2, …, C_n) that are independent, their joint probability can be expressed as:

$$P(C_1, C_2, \ldots, C_n) = \prod_i P(C_i)$$

- Each C_i has its own table :

C_1	$P(C_1)$
0	0.4
1	0.6

C_2	$P(C_2)$
0	0.7
1	0.3

…

C_n	$P(C_n)$
0	0.2
1	0.8

- Total of n tables, 2n entries, and n parameters.

Conditional Independence

Variables A and B are conditionally independent given C, if any of the following hold:

- $P(A, B \mid C) = P(A \mid C) \times P(B \mid C)$
- $P(A \mid B, C) = P(A \mid C)$
- $P(B \mid A, C) = P(B \mid C)$

Notation: $A \perp\!\!\!\perp B \mid C$

Knowing *C* tells me everything about *B*. I don't gain anything by knowing *A* (either because *A* doesn't influence *B* or because knowing *C* provides all the information knowing *A* would give)

Conditional Independence

- In real life, the scenario is somewhere in between total dependence and total independence.

- Some variables may be dependent, while others may not be.

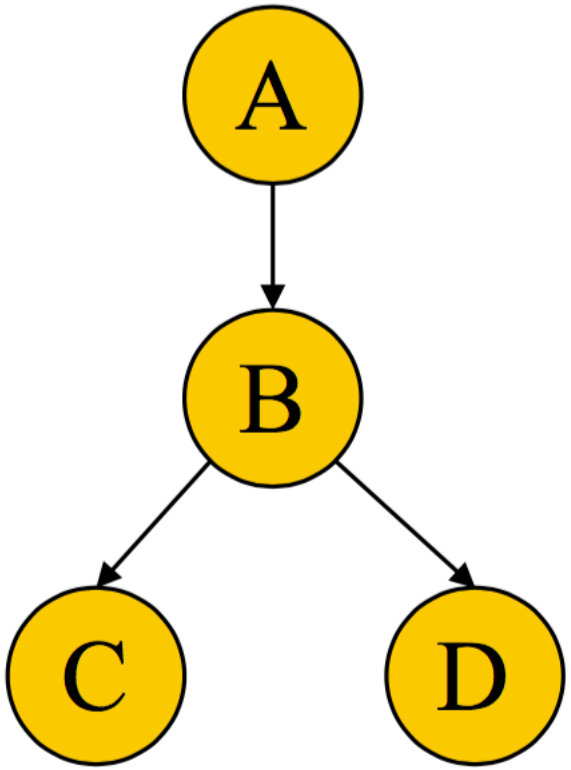
- We need a way to represent these relationships.

- This is where Bayesian networks come in.

Bayesian Network

- A Bayes Net is made up of:

1. A Directed Acyclic Graph (DAG) representing dependencies



2. A set of tables for each node

A	**P(A)**
false	0.6
true	0.4

A	**B**	**P(B|A)**
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

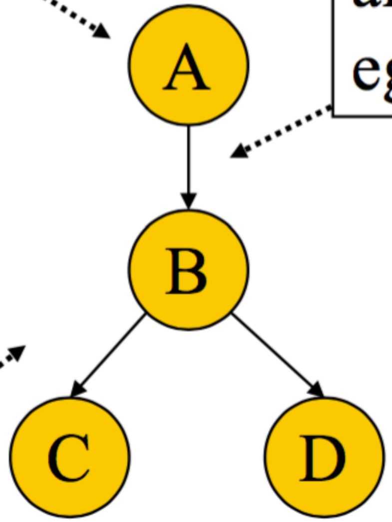
B	**D**	**P(D|B)**
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

B	**C**	**P(C|B)**
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

DAG

Each node in the graph is a random variable

A node X is a parent of another node Y if there is an arrow from node X to node Y eg. A is a parent of B



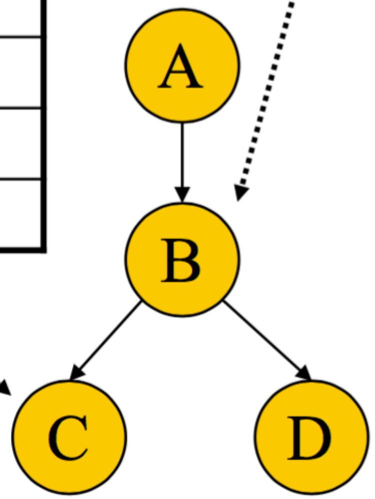
Informally, an arrow from node X to node Y means X has a direct influence on Y

A Set of Tables for Each Node

A	P(A)
false	0.6
true	0.4

A	B	P(B|A)
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1



B	D	P(D|B)
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

Each node X_i has a conditional probability distribution P(X_i | Parents(X_i)) that quantifies the effect of the parents on the node

The parameters are the probabilities in these conditional probability tables (CPTs)

A Set of Tables for Each Node

Conditional Probability Distribution for C given B

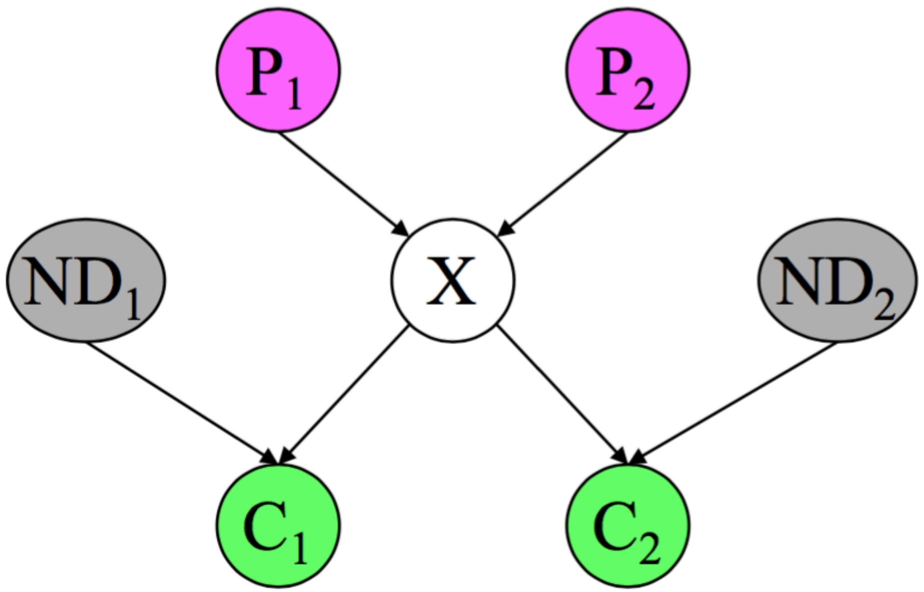
B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

For a given combination of values of the parents (B in this example), the entries for P(C=true | B) and P(C=false | B) must add up to 1
eg. P(C=true | B=false) + P(C=false |B=false)=1

If you have a Boolean variable with k Boolean parents, this table has 2^{k+1} probabilities (but only 2^k need to be stored)

Conditional Independence using Bayes Net

The Markov condition: given its parents (P_1, P_2), a node (X) is conditionally independent of its non-descendants (ND_1, ND_2)



Technically, this is first order Markov property – each node only depends on one previous level.

Joint Probability Distribution

- Using a Bayes Net and Markov condition, we can compute the joint probability distribution over all the variables $X_1, X_2, \ldots, X_n$

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i \mid Parents(X_i))$$

Where Parents(X_i) means the values of the Parents of the node X_i with respect to the graph

Example

- Compute the following joint probability:

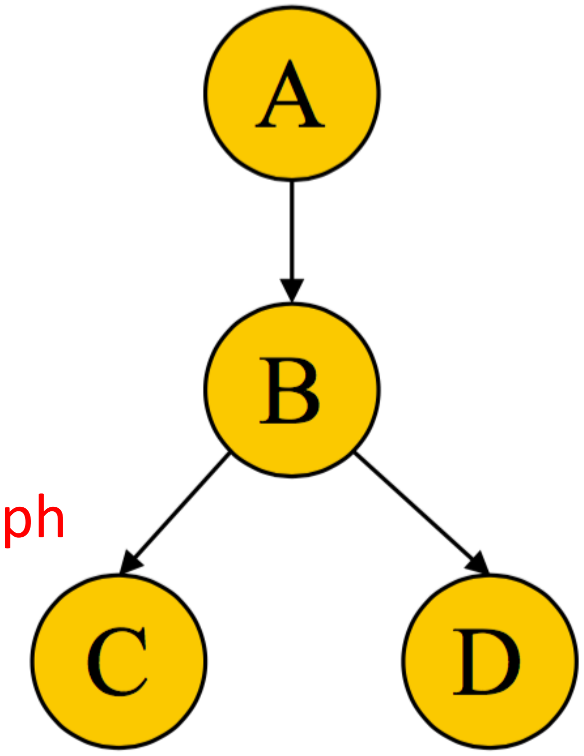
  P(A = true, B = true, C = true, D = true)

  = P(A = true) * P(B = true | A = true) * P(C = true | B = true) * P(D = true | B = true)

  = (0.4) * (0.3) * (0.1) * (0.95)

  = 0.0114

This knowledge comes from graph structure



This knowledge comes from CPT

A	P(A)
false	0.6
true	0.4

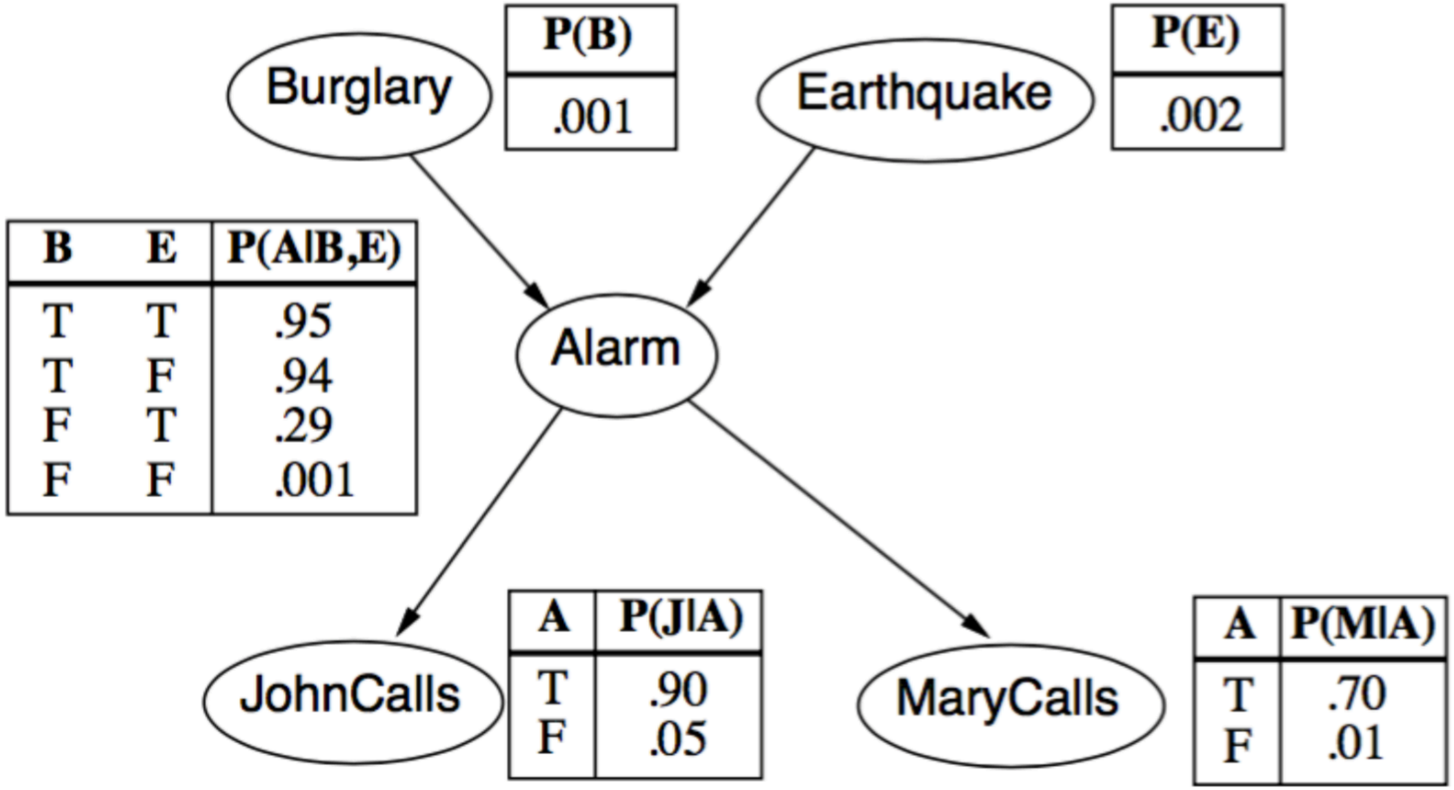
A	B	P(B|A)
false	false	0.01
false	true	0.99
true	false	0.7
true	true	0.3

B	D	P(D|B)
false	false	0.02
false	true	0.98
true	false	0.05
true	true	0.95

B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

Example

- Consider the Bayes Net:



- How many parameters? 10

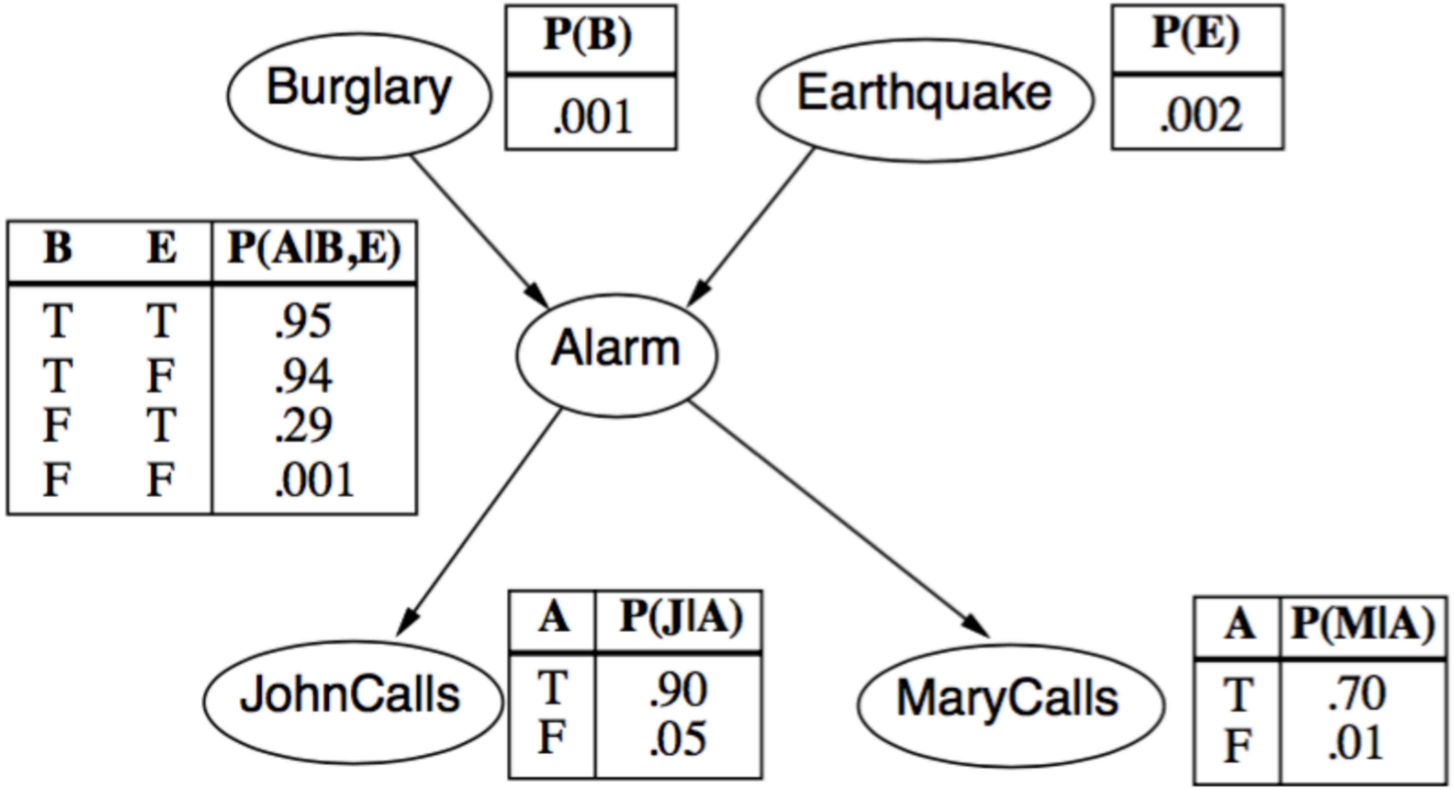
- Without CI assumption, we would require 2^5 -1 parameters.

Example

- Consider the Bayes Net:

- What's the joint probability of
  P(J, M, A, ⌐B, ⌐E)

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



Example:

- Consider the BN:

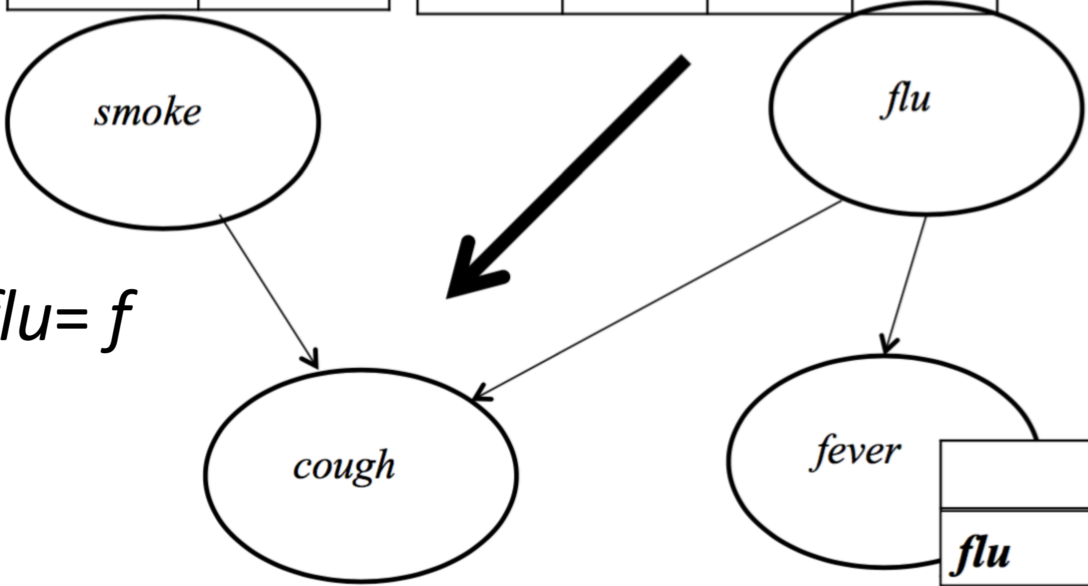
Example:

smoke	
true	0.2
false	0.8

flu	smoke	cough: true	cough: false
True	True	0.95	0.05
True	False	0.8	0.2
False	True	0.6	0.4
false	false	0.05	0.95

Conditional probability tables for each node

flu	
true	0.01
false	0.99



flu	fever: true	fever: false
true	0.9	0.1
false	0.2	0.8

- Use it to evaluate:

$$P(cough{=}t \wedge fever{=}f \wedge flu{=}f \wedge smoke{=}f)$$

Example:

$$P(cough = t \wedge fever = f \wedge flu = f \wedge smoke = f)$$

$$= \prod_{i=1}^{n} P(X_i = x_i \mid parents(X_i))$$

$$= P(cough = t \mid flu = f \wedge smoke = f)$$

$$\times P(fever = f \mid flu = f)$$

$$\times P(flu = f)$$

$$\times P(smoke = f)$$

$$= .05 \times .8 \times .99 \times .8$$

$$= .032$$

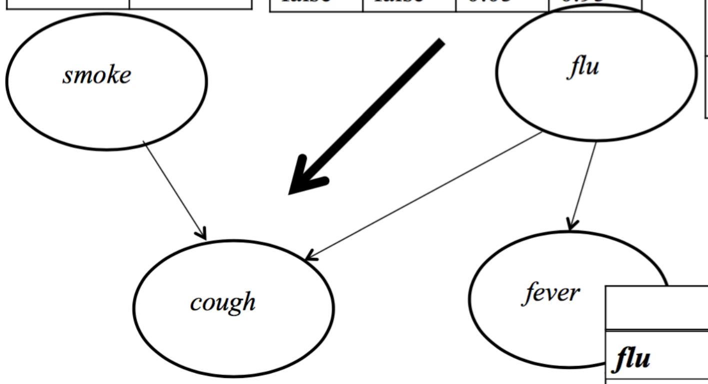
Example:

Conditional probability tables for each node

	smoke
true	0.2
false	0.8

		cough	
flu	***smoke***	true	false
True	True	0.95	0.05
True	False	0.8	0.2
False	True	0.6	0.4
false	false	0.05	0.95

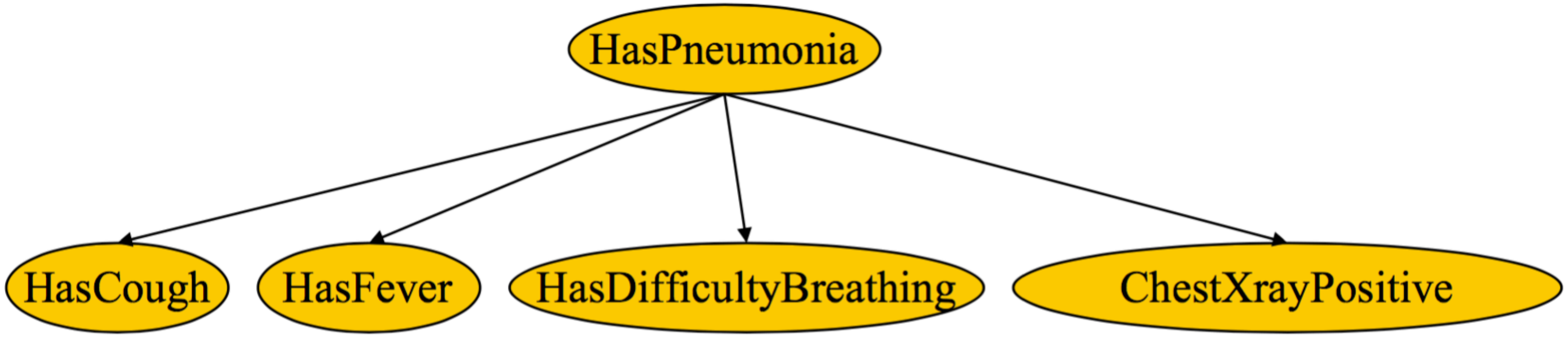
	flu
true	0.01
false	0.99



	fever	
flu	true	false
true	0.9	0.1
false	0.2	0.8

Inference Queries using BN

- The previous 2 were examples of inference queries using BN.
- In general, we can have queries like: P(X | E)
  where X = query variable, E = evidence variable.
- Another Example:



P(HasPneumonia = T | HasFever = T, HasCough = T)

Query

Evidence

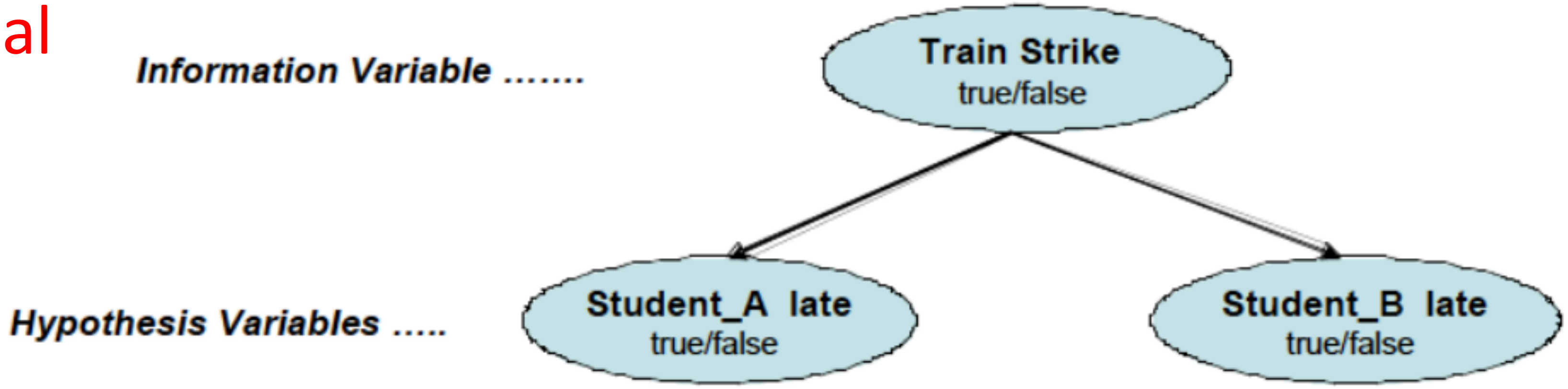
Another Bayesian Belief Network (BBN)

- What's the probability that A is late?
  What's the probability that B is late?

This is called unconditional or marginal probability

A could be late with a train strike or without a train strike

Train Strike	
true	0.1
false	0.9



	Train Strike	
Student_A late	*true*	*false*
true	0.8	0.1
false	0.2	0.9

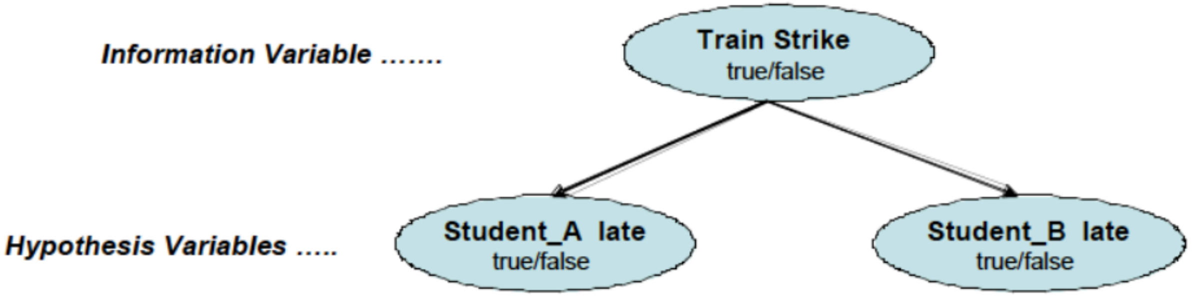
	Train Strike	
Student_B late	*true*	*false*
true	0.6	0.5
false	0.4	0.5

Marginal Probability

$$P(StudentALate) = P(StudentALate \mid TrainStrike)P(TrainStrike)$$
$$+P(StudentALate \mid \neg TrainStrike)P(\neg TrainStrike)$$
$$= 0.8 \times 0.1 + 0.8 \times 0.9 = 0.17$$

$$P(StudentBLate) = P(StudentBLate \mid TrainStrike)P(TrainStrike)$$
$$+P(StudentBLate \mid \neg TrainStrike)P(\neg TrainStrike)$$
$$= 0.6 \times 0.1 + 0.5 \times 0.9 = 0.51$$

Train Strike	
true	0.1
false	0.9



	Train Strike	
Student_A late	*true*	*false*
true	0.8	0.1
false	0.2	0.9

	Train Strike	
Student_B late	*true*	*false*
true	0.6	0.5
false	0.4	0.5

Evidence About Parent Given

- Given that there is a train strike, what's the probability that A is late. P(Student_A late | Train Strike)

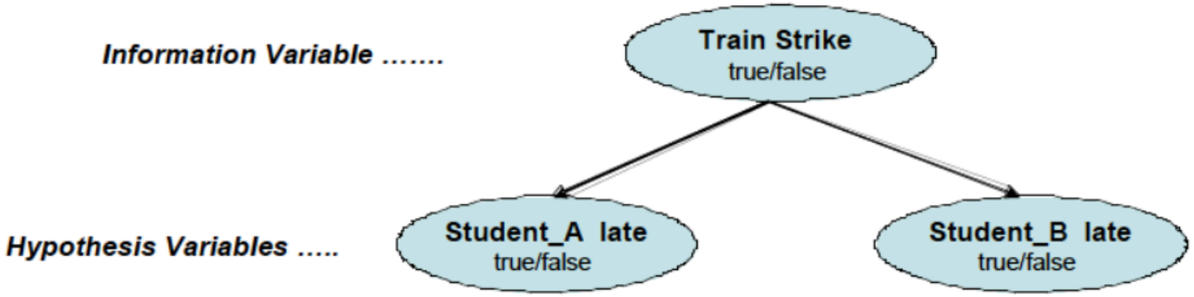
- Simple – just look up the table

Train Strike	
true	0.1
false	0.9

Evidence: There is a train strike.

$$P(StudentALate) = 0.8$$

$$P(StudentBLate) = 0.6$$



	Train Strike	
Student_A late	*true*	*false*
true	0.8	0.1
false	0.2	0.9

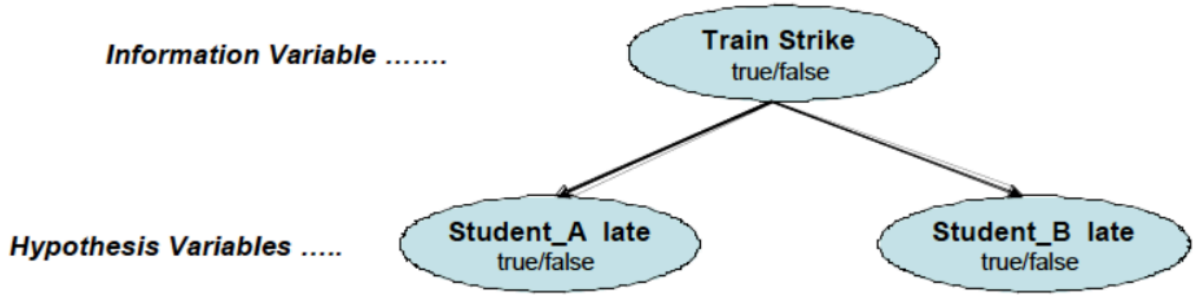
	Train Strike	
Student_B late	*true*	*false*
true	0.6	0.5
false	0.4	0.5

Evidence About Child Node Given

- Suppose we know that student A was late, how does it revise probability of train being late, and student B being late?

Train Strike	
true	0.1
false	0.9

- This is idea behind belief propagation.



- Evidence: Student A late
  Query: Train Strike

	Train Strike	
Student_A late	*true*	*false*
true	0.8	0.1
false	0.2	0.9

	Train Strike	
Student_B late	*true*	*false*
true	0.6	0.5
false	0.4	0.5

$$P(TrainStrike \mid StudentALate) = \frac{P(StudentALate \mid TrainStrike)P(TrainStrike)}{P(StudentALate)}$$ by Bayes Theorem

$$= \frac{0.8 \times 0.1}{0.17} = 0.47$$

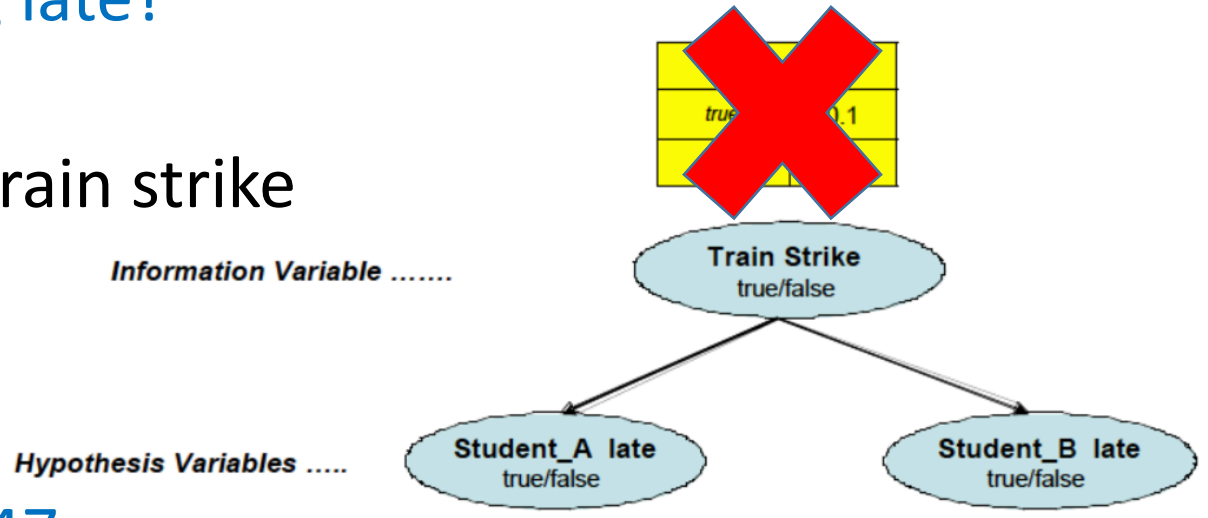
Evidence About Child Node Given

P(TrainStrike) = 0.47

- How does it affect probability of B being late?

- We need to use updated probability of train strike

- Evidence: Student A late
  Query: Student B late
  We saw that updated P(TrainStrike) = 0.47



	Train Strike	
Student_A late	true	false
true	0.8	0.1
false	0.2	0.9

	Train Strike	
Student_B late	true	false
true	0.6	0.5
false	0.4	0.5

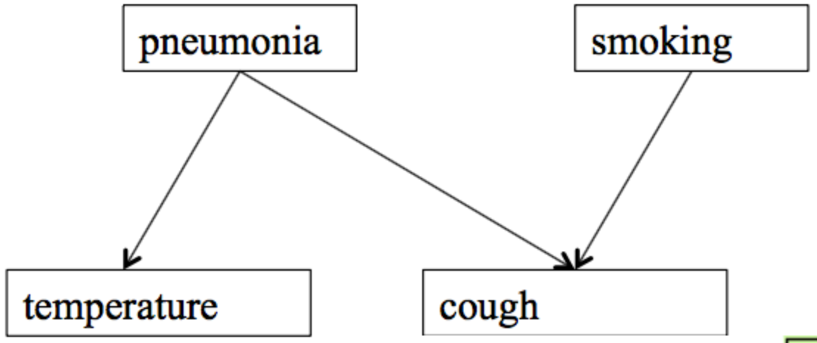
$$P(StudentBLate) = P(StudentBLate \mid TrainStrike)P(TrainStrike)$$
$$+P(StudentBLate \mid \neg TrainStrike)P(\neg TrainStrike)$$
$$= 0.6 \times 0.47 + 0.5 \times 0.53 = 0.55$$

Practice Question

- Consider the Bayes Net shown

pneumonia	
true	0.1
false	0.9

smoking	
yes	0.2
no	0.8



	temperature	
pneumonia	yes	no
yes	0.9	0.1
no	0.2	0.8

		cough	
pneumonia	smoking	true	false
true	yes	0.95	0.05
true	no	0.8	0.2
false	yes	0.6	0.4
false	no	0.05	0.95

- Evaluate:

  P(Cough | Smoking = T)

- How do you proceed?
  You need to marginalize over pneumonia.

$$P(Cough \mid Smoking) = \sum_{p \in Pneumonia} P(Cough, p \mid Smoking)$$

Practice Question

- $P(Cough \mid \ Smoking)$

$$= \sum_{p \in Pneumonia} P(Cough, p \mid Smoking)$$

$$= \sum_{p \in Pneumonia} P(Cough \mid Smoking, p)\ P(Pneumonia = p)$$

$$= P(Cough \mid Smoking\ , Pneumonia)\ P(Pneumonia) + P(Cough \mid Smoking, \neg Pneumonia)\ P(\neg Pneumonia)$$

= 0.95 * 0.1 + 0.6 * 0.9

= 0.635

Practice Question

- Consider the Bayes Net shown

- Evaluate:

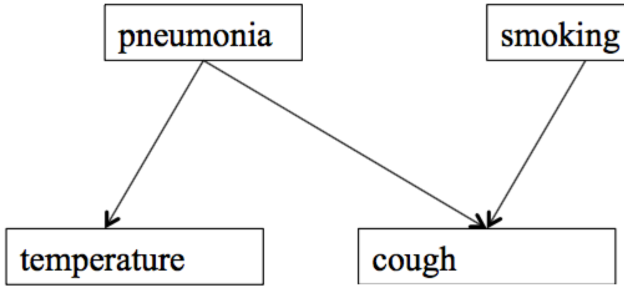
  P(Cough | Pneumonia = F)

  P(Cough | Smoking = T, Pneumonia = F)

  P(Cough)

pneumonia	
true	0.1
false	0.9

smoking	
yes	0.2
no	0.8



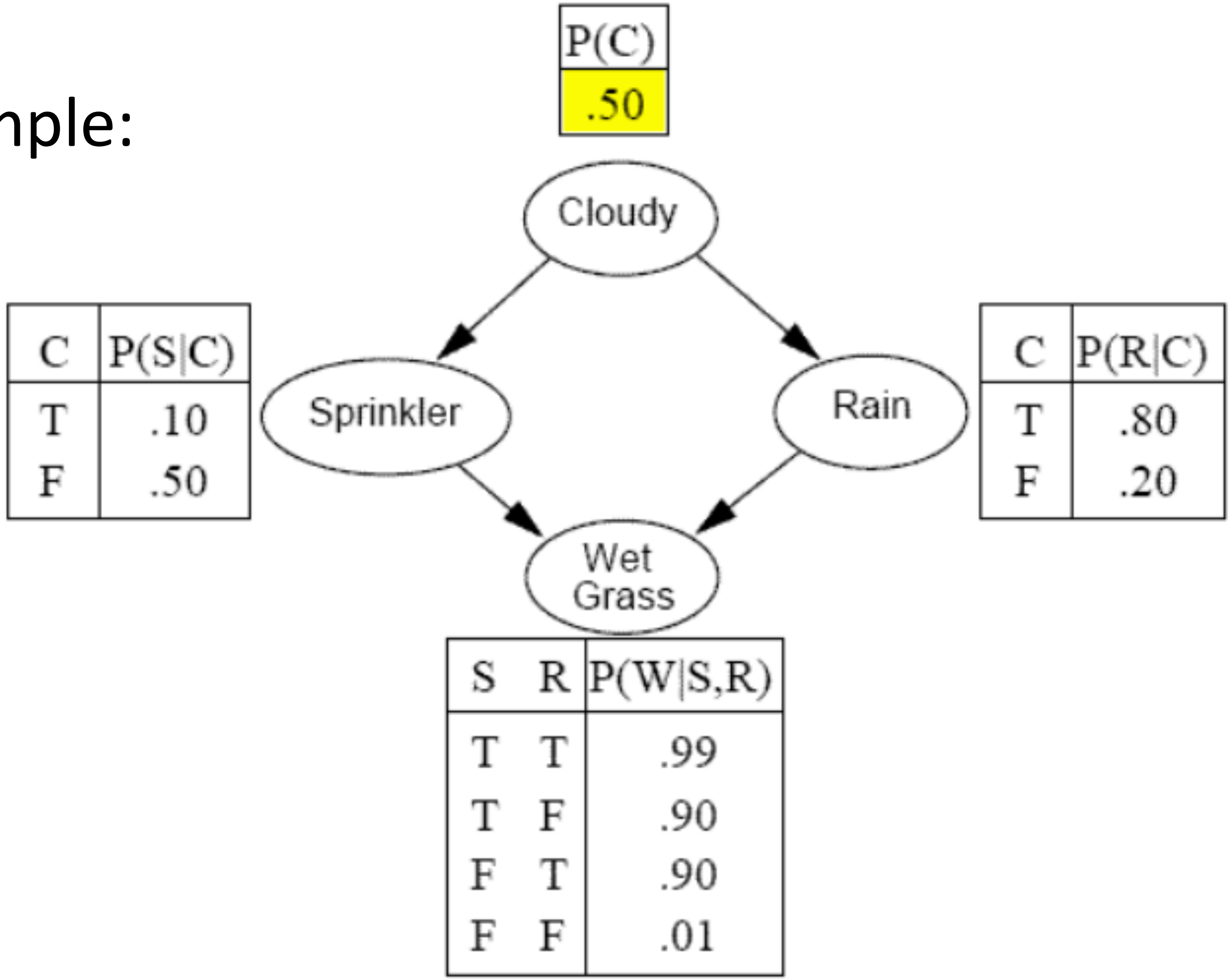
	temperature	
pneumonia	yes	no
yes	0.9	0.1
no	0.2	0.8

		cough	
pneumonia	smoking	true	false
true	yes	0.95	0.05
true	no	0.8	0.2
false	yes	0.6	0.4
false	no	0.05	0.95

More Exercises:

- Consider the famous sprinkler example:

  Evaluate:
  P(C, R, ¬S, W)
  = P(C) P(R|C) P(¬S|C) P(W|¬S, R)
  = (0.50)(0.80)(0.90)(0.90)
  = 0.324



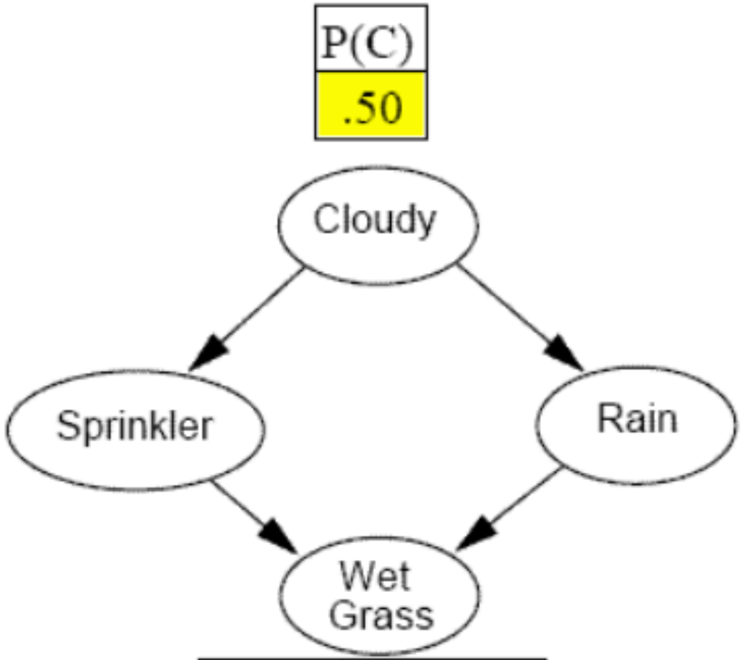
More Exercises:

- Now, some complex queries:

What is P(*Cloudy*| *Sprinkler*)?

What is P(*Cloudy* | *Rain*)?

What is P(*Cloudy* | *Wet Grass*)?



C	P(S|C)
T	.10
F	.50

C	P(R|C)
T	.80
F	.20

S	R	P(W|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Inference Queries on a Bayes Net

We can run 3 types of queries on a Bayes Net

- **Diagnostic:** Use evidence of an effect to infer probability of a cause.
  - E.g., **Evidence:** *cough=true*. What is *P*(*pneumonia* | *cough*)?

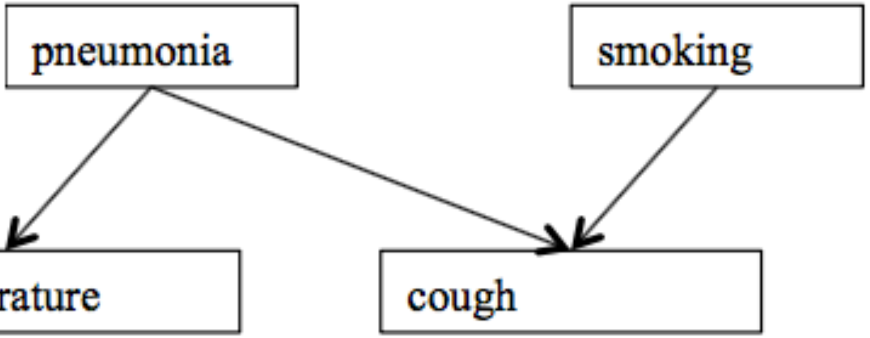
- **Causal inference:** Use evidence of a cause to infer probability of an effect
  - E.g., **Evidence:** *pneumonia=true*. What is *P*(*cough* | *pneumonia*)?

- **Inter-causal inference:** "Explain away" potentially competing causes of a shared effect.
  - E.g., Evidence: *smoking=true.* What is *P*(*pneumonia* | *cough* and *smoking*)?

Diagnostic Queries

- **Diagnostic: Evidence:** *cough=true*. What is *P*(*pneumonia* | *cough*)?

pneumonia	
true	0.1
false	0.9



smoking	
yes	0.2
no	0.8

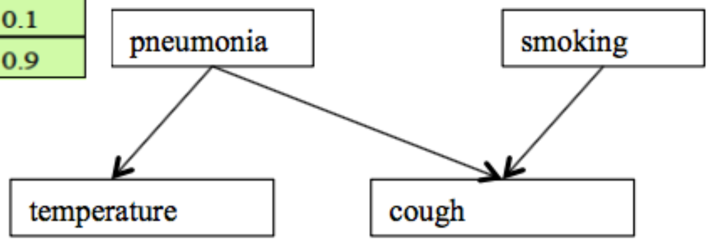
	temperature	
pneumonia	yes	no
yes	0.9	0.1
no	0.2	0.8

		cough	
pneumonia	smoking	true	false
true	yes	0.95	0.05
true	no	0.8	0.2
false	yes	0.6	0.4
false	no	0.05	0.95

Diagnostic Queries

- **Diagnostic: Evidence:** *cough=true*. What is *P*(*pneumonia* | *cough*)?

pneumonia	
true	0.1
false	0.9



smoking	
yes	0.2
no	0.8

	temperature	
pneumonia	yes	no
yes	0.9	0.1
no	0.2	0.8

		cough	
pneumonia	smoking	true	false
true	yes	0.95	0.05
true	no	0.8	0.2
false	yes	0.6	0.4
false	no	0.05	0.95

$$P(pneumonia \mid cough) = \frac{P(cough \mid pneumonia)P(pneumonia)}{P(cough)}$$

$$= \frac{[P(cough \mid pneumonia, smoking)P(smoking) + P(cough \mid pneumonia, \neg smoking)P(\neg smoking)]P(pneumonia)]}{P(cough)}$$

$$= \frac{[(.95)(.2)+(.8)(.8)](.1)}{P(cough)} = \frac{.083}{P(cough)}$$

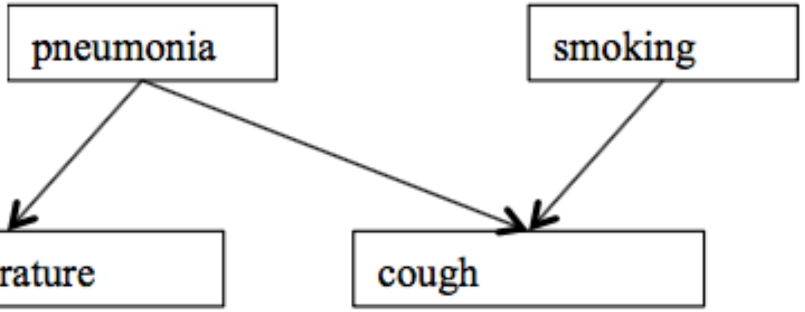
$$= \frac{.083}{P(cough)} = \frac{.083}{.227} = .366$$

We are marginalizing over or "summing out" the smoking variable

Causal Queries

- **Causal: Evidence:** *pneumonia=true*. What is *P*(*cough* | *pneumonia*)?

pneumonia	
true	0.1
false	0.9



smoking	
yes	0.2
no	0.8

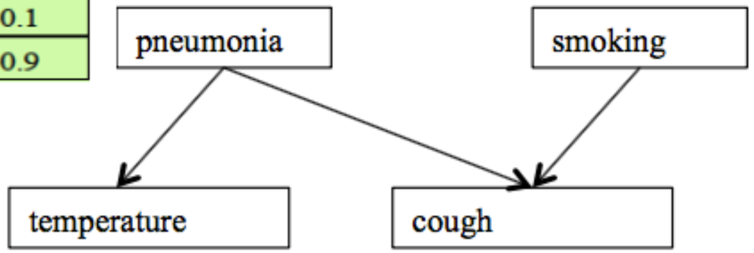
	temperature	
pneumonia	yes	no
yes	0.9	0.1
no	0.2	0.8

		cough	
pneumonia	smoking	true	false
true	yes	0.95	0.05
true	no	0.8	0.2
false	yes	0.6	0.4
false	no	0.05	0.95

Causal Queries

- **Causal: Evidence:** *pneumonia=true*. What is *P*(*cough* | *pneumonia*)?

pneumonia	
true	0.1
false	0.9



smoking	
yes	0.2
no	0.8

	temperature	
pneumonia	yes	no
yes	0.9	0.1
no	0.2	0.8

		cough	
pneumonia	smoking	true	false
true	yes	0.95	0.05
true	no	0.8	0.2
false	yes	0.6	0.4
false	no	0.05	0.95

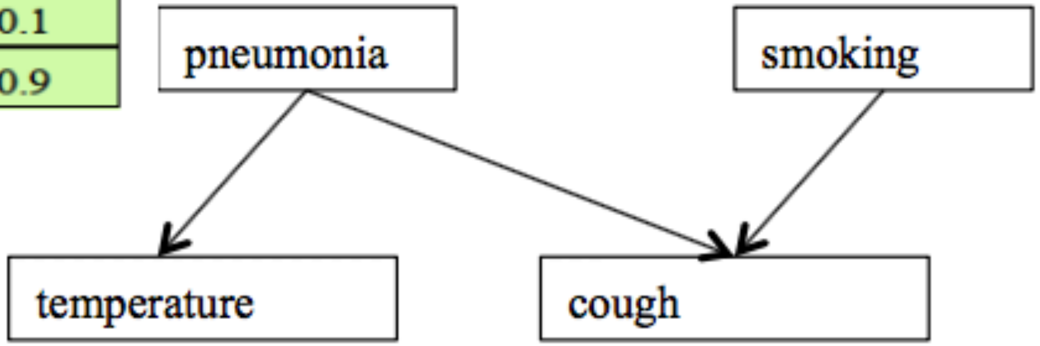
$$P(cough \mid pneumonia) = P(cough \mid pneumonia, smoking)P(smoking)$$
$$+P(cough \mid pneumonia, \neg smoking)P(\neg smoking)$$
$$= \left[(.95)(.2) + (.8)(.8)\right] = .83$$

We are marginalizing over or "summing out" the smoking variable

Inter-Causal Queries

- **Inter-causal:** Evidence: *smoking=true*. What is *P*(*pneumonia* | *cough* and *smoking*)?

pneumonia	
true	0.1
false	0.9



smoking	
yes	0.2
no	0.8

	temperature	
pneumonia	yes	no
yes	0.9	0.1
no	0.2	0.8

		cough	
pneumonia	smoking	true	false
true	yes	0.95	0.05
true	no	0.8	0.2
false	yes	0.6	0.4
false	no	0.05	0.95

Inter-Causal Queries

$$P(pneumonia \mid cough \wedge smoking) = \frac{P(cough \wedge smoking \mid pneumonia)P(pneumonia)}{P(cough \wedge smoking)}$$

$$= \frac{P(cough \wedge smoking \wedge pneumonia)}{P(pneumonia)} \frac{P(pneumonia)}{P(cough \wedge smoking)}$$

$$= \frac{P(cough \wedge smoking \wedge pneumonia)}{P(cough \wedge smoking)} = \frac{P(cough \mid pneumonia, smoking)P(smoking)P(pneumonia)}{P(cough \mid smoking)P(smoking)}$$

$$= \frac{(.95)(.2)(.1)}{\begin{array}{l}[P(cough \mid smoking, pneumonia)P(pneumonia) \\ +P(cough \mid smoking, \neg pneumonia)P(\neg pneumonia)]P(smoking)\end{array}}$$

$$= \frac{.019}{[(.95)(.1)+(.6)(.9)](.2)} = .15$$

"Explaining away"

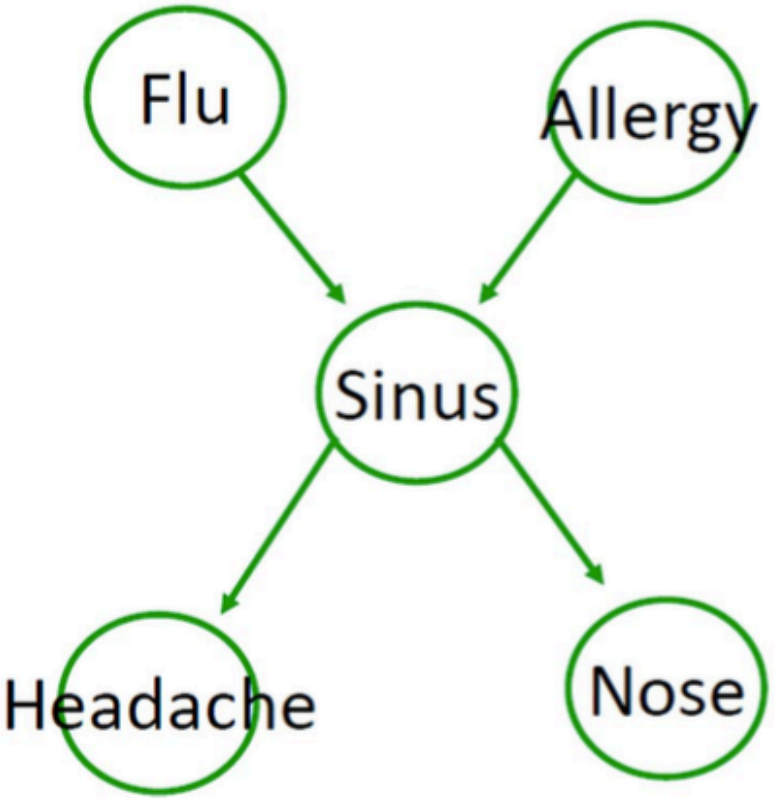
Dependencies and
Conditional Independence in a BN

Markov Assumption

- Remember our Markov Assumption:
  A variable X is independent of its non- descendants given (only) its parents

	parents	non-desc	assumption
S	F,A	-	-
H	S	F,A,N	$H \perp \{F,A,N\} \mid S$
N	S	F,A,H	$N \perp \{F,A,H\} \mid S$
F	-	A	$F \perp A$
A	-	F	$A \perp F$

$F \perp A, \quad H \perp \{F,A\} \mid S, \quad N \perp \{F,A,H\} \mid S$



Markov Assumption

- How does it help us run inference queries?

$P(F, A, S, H, N)$

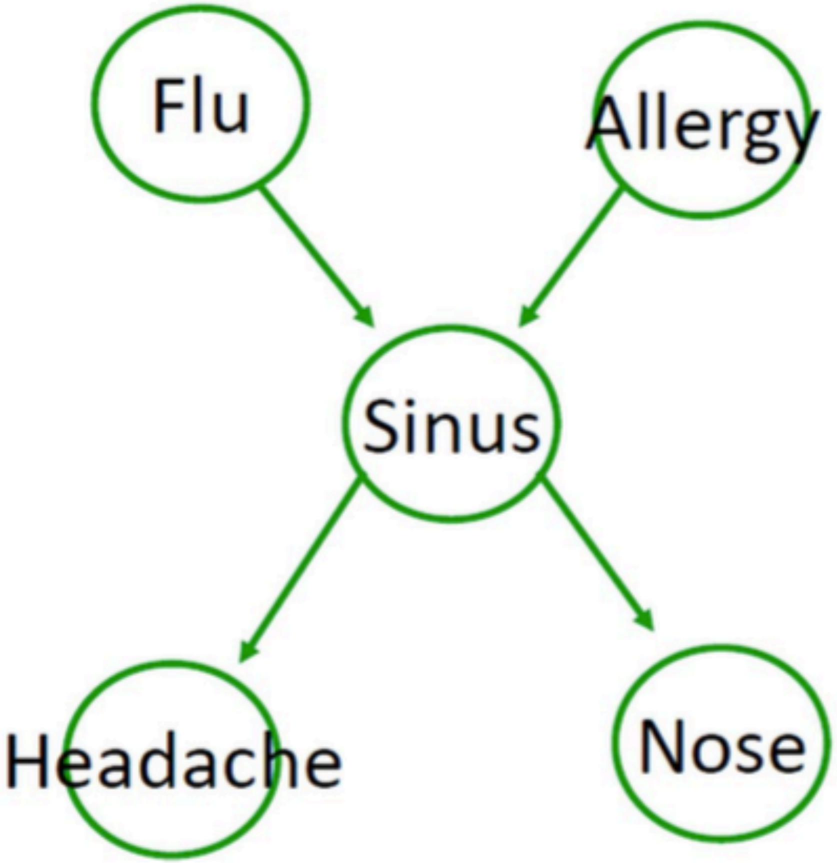
$= P(F)\ P(F|A)\ P(S|F,A)\ P(H|S,F,A)\ P(N|S,F,A,H)$

Chain rule

$= P(F)\ P(A)\ P(S|F,A)\ P(H|S)\ P(N|S)$

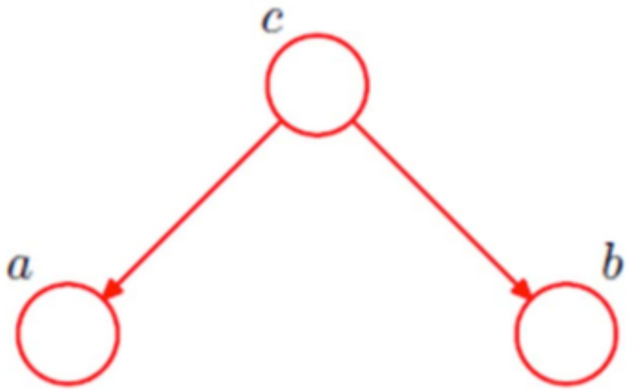
Markov Assumption

$F \perp A, \quad H \perp \{F,A\}|S, \quad N \perp \{F,A,H\}|S$



Inferring CI from Factored Joint Distribution

- For the BN on the right, how would you factor the joint distribution?
- p(a, b, c) = p(a|c) p(b|c) p(c)
- Can you infer that a and b are CI given c?

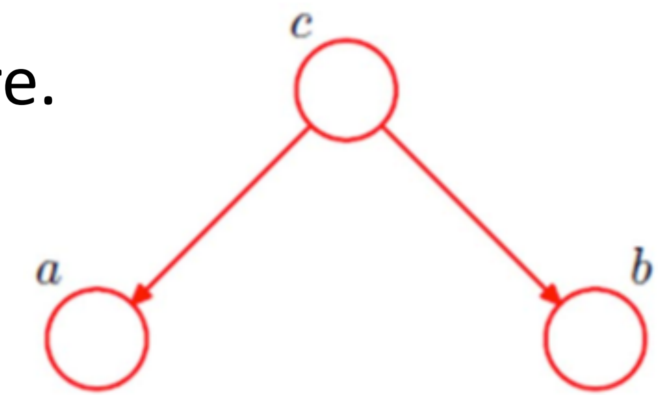


Show that $a \perp\!\!\!\perp b \mid c$

$$
\begin{aligned}
p(a, b|c) &= \frac{p(a, b, c)}{p(c)} \\
&= \frac{p(a|c)p(b|c)p(c)}{p(c)} \\
&= p(a|c)p(b|c)
\end{aligned}
$$

Inferring CI from Factored Joint Distribution

- Note that we used the Markov property and BN structure.



Do we have $a \perp\!\!\!\perp b$? In general, no.

$$p(a, b) = \sum_c p(a, b, c)$$
$$= \sum_c p(a|c)p(b|c)p(c)$$

- Cannot be written into two separate terms of *a* and *b*

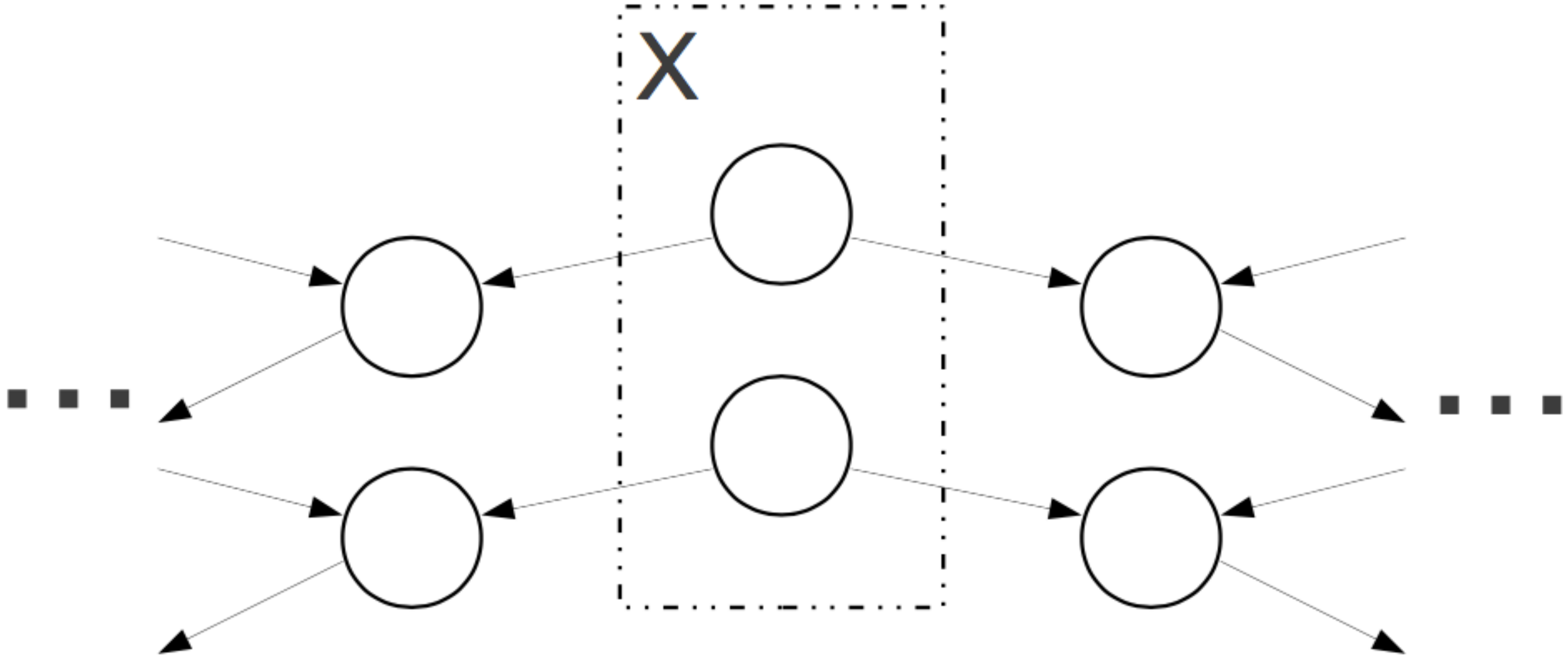
Dependency Flow

- Consider the BN on right, how can we tell if x_1 and x_M are dependent?
  It depends on whether the node x_2 allows the flow of dependency or blocks the flow of dependency.

- How do we infer whether a node in a BN allows the flow of dependency or blocks it?
- This is the idea behind D-separation.

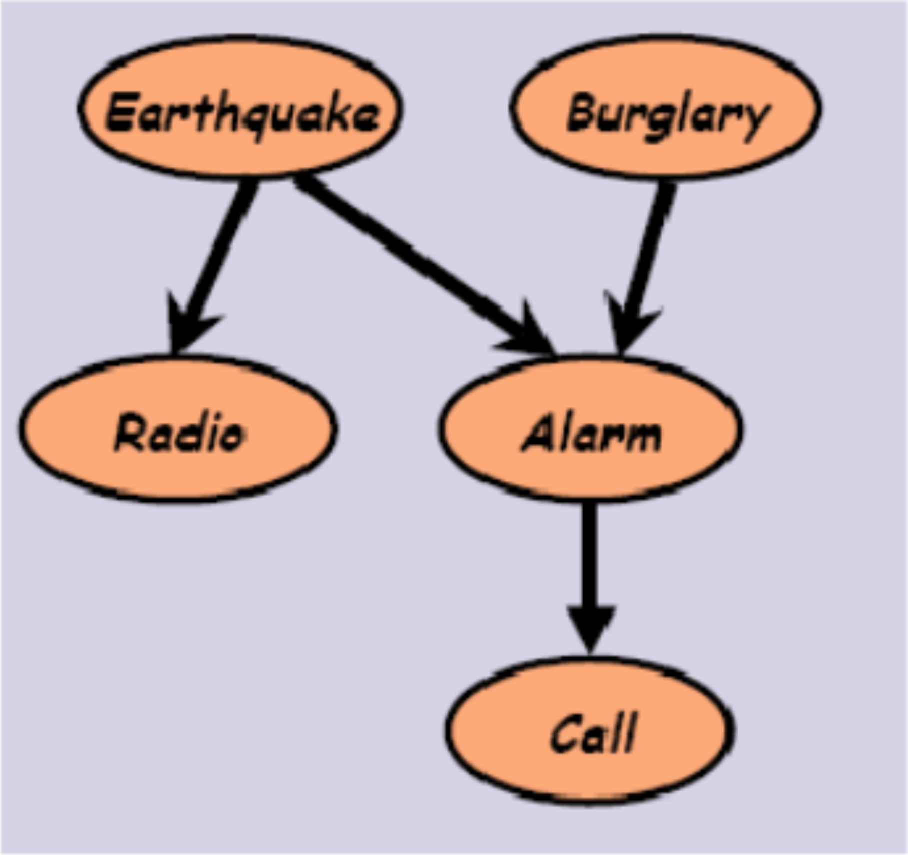
d-Separation

- Helps us understand the dependencies implied by a graph
- Helps us perform inference efficiently



Path in a Bayes Net

- We would like to know if dependency flows along paths on a BN
- By a "path" we mean any consecutive sequence of edges, disregarding their directionalities.
- It can also be defined as a sequence of neighboring variables.
- For the BN on the right, for finding paths, we consider consecutive sequence of undirected edges.
  We will also use the direction arrows later.
- Examples of paths:
  R – E – A – B
  C – A – E – R



d-Separation

- Let's see under what conditions a node blocks flow of dependency.

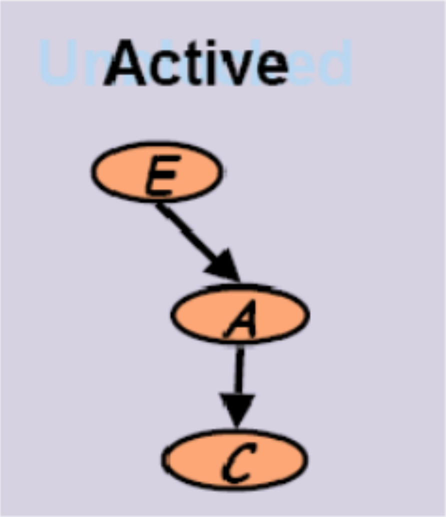
  Definition: If X1, X2, and X3 are three disjoint subsets of nodes in a DAG, then X2 is said to d-separate X1 from X3 if every undirected path from X1 to X3 is blocked by X2. A path is **blocked** if it contains a node Z such that:

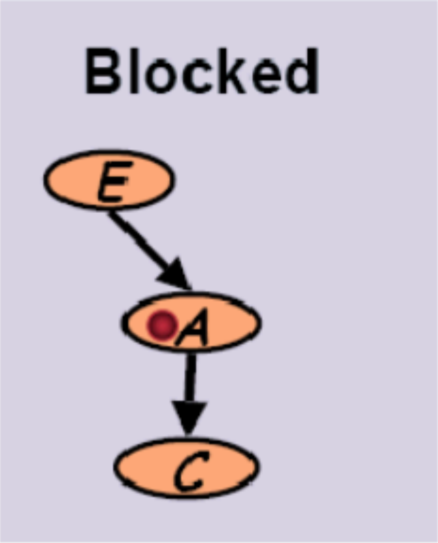
  (1) Z has one incoming and one outgoing arrow and Z is in X2.
  (Serial Connection or Causal Chain)
  (2) Z has two outgoing arrows and Z is in X2.
  (Diverging Connection or Common Cause)
  (3) Z has two incoming arrows and neither Z nor any of its descendants is in X2.
  (Converging Connection)

A serial connection

$$X_1 \rightarrow X_2 \rightarrow X_3$$

- In a serial connection from *X1 to X3 via X2, evidence from X1 to X3 is **blocked** only when we have* hard evidence about *X2.*
- Intermediate cause.





Proof of Serial Connection (Causal Chain)

- 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?

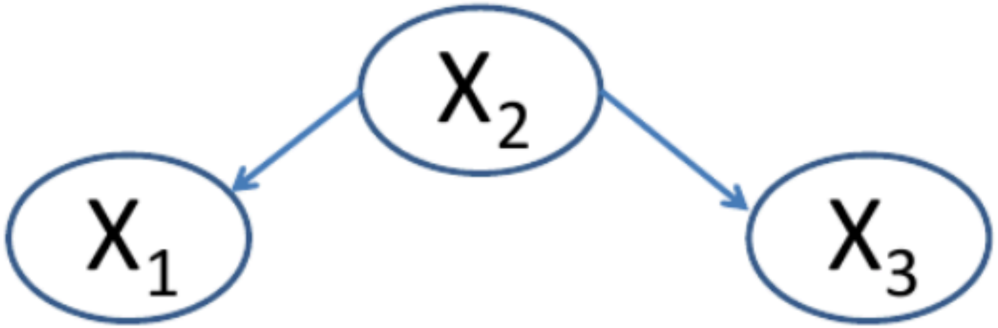
$$P(z|x, y) = \frac{P(x, y, z)}{P(x, y)} = \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)}$$

$$= P(z|y)$$

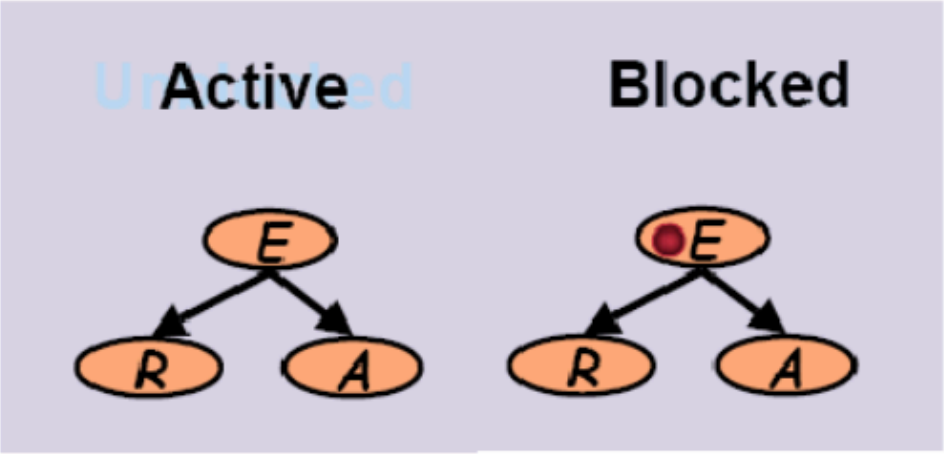
Yes!

- Evidence along the chain "blocks" the influence

A diverging connection

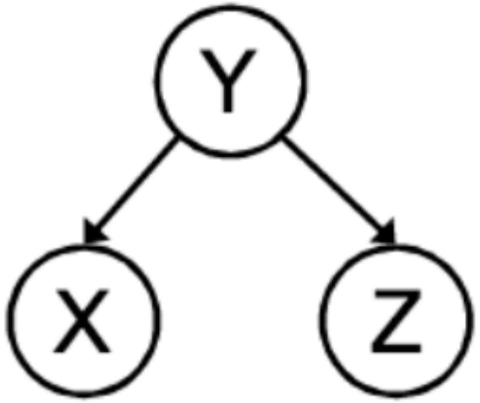


- In a diverging connection where *X1 and X3 have the common parent X2, evidence from X1 to X3 is* blocked only when we have hard evidence about *X2.*
- Common cause.



Proof of Diverging Connection (Common Cause)

- 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: Newsgroup busy

Z: Lab full

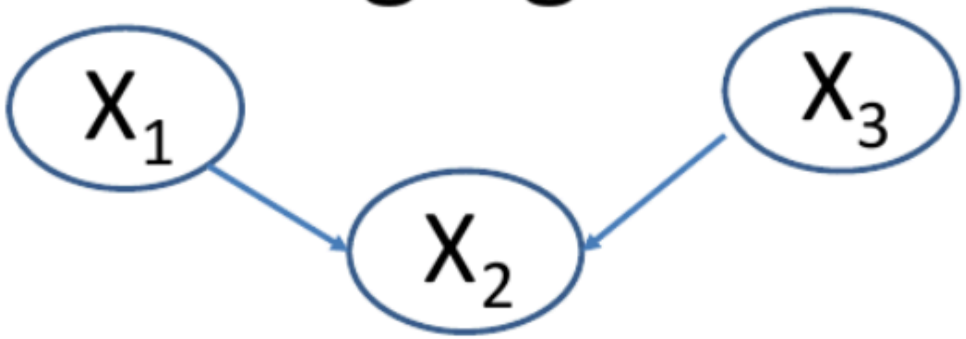
$$P(z|x,y) = \frac{P(x,y,z)}{P(x,y)} = \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}$$

$$= P(z|y)$$

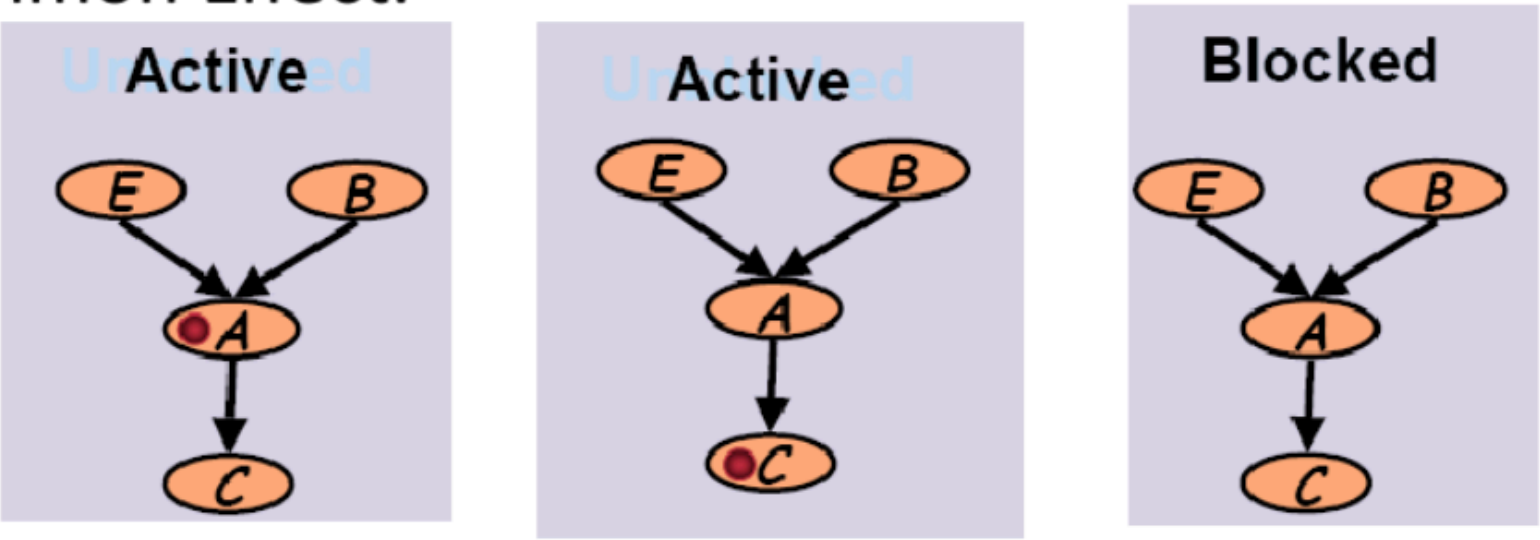
Yes!

  - Observing the cause blocks influence between effects.

A Converging connection

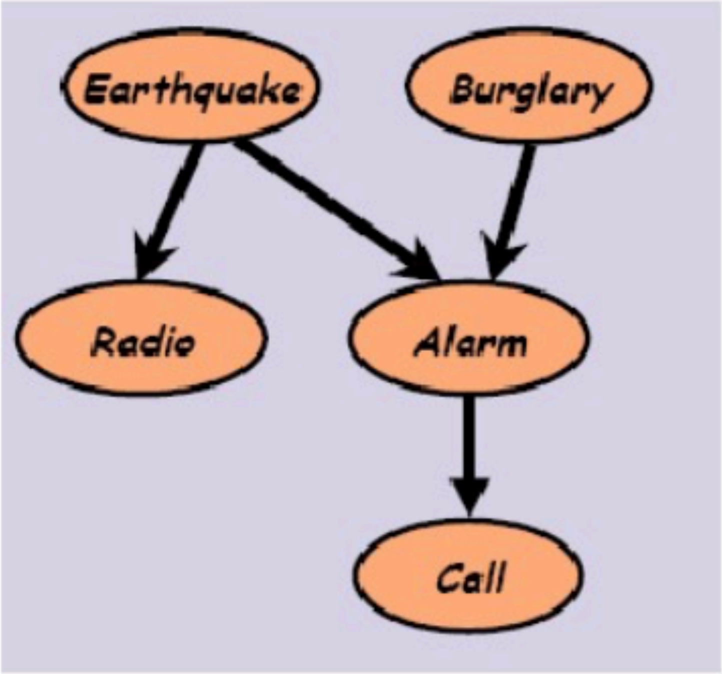


- In a converging connection where *X2 has parents X1 and X3*, *any evidence about X2 results in* evidence transmitted between *X1 and X3*.
- Common Effect.



Example 1

- d-sep(R,B)?

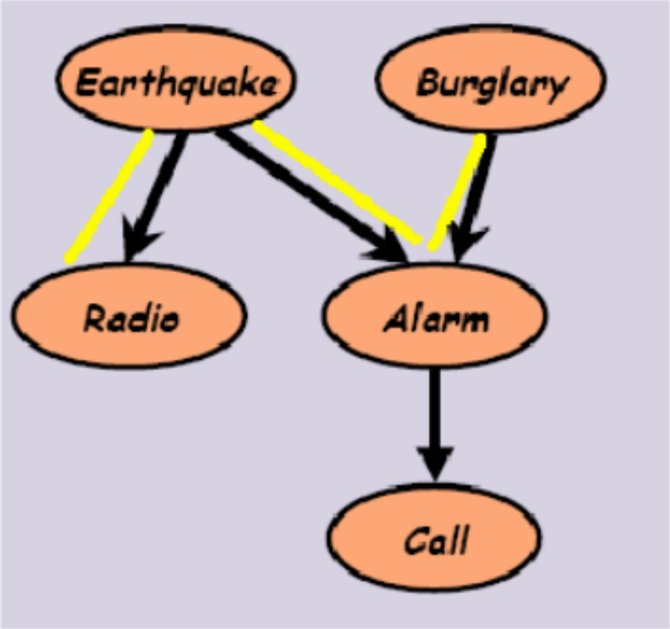


Example 1

- d-sep(X1, X3|X2)
- d-sep(R,B)?
  - X1 = {R}, X3={B}, X2 ={}
  - Find all the path between R, B
  - Check the node:
    - Earthquake.

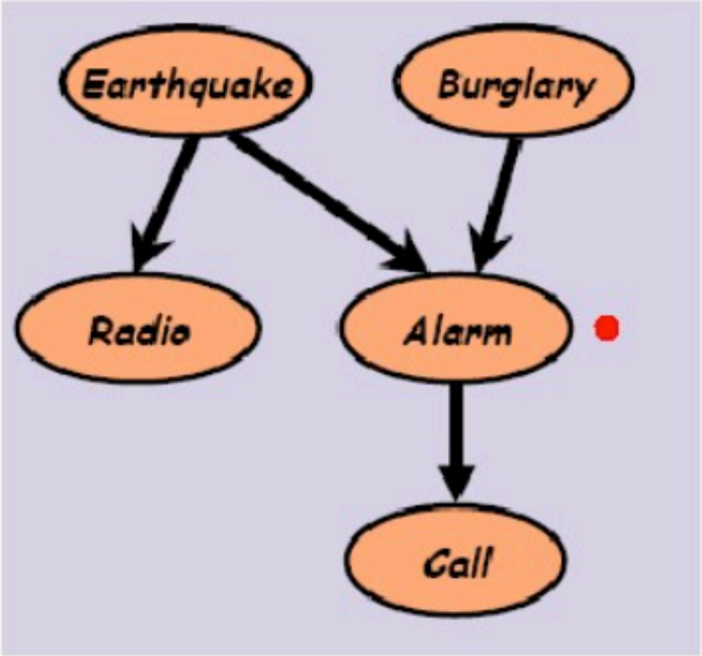
      (diverging, not in X2). Not blocking.
    - Alarm

      (Converging, A or C are not in X2). Block!



Example 2

- d-sep(R,B|A)?

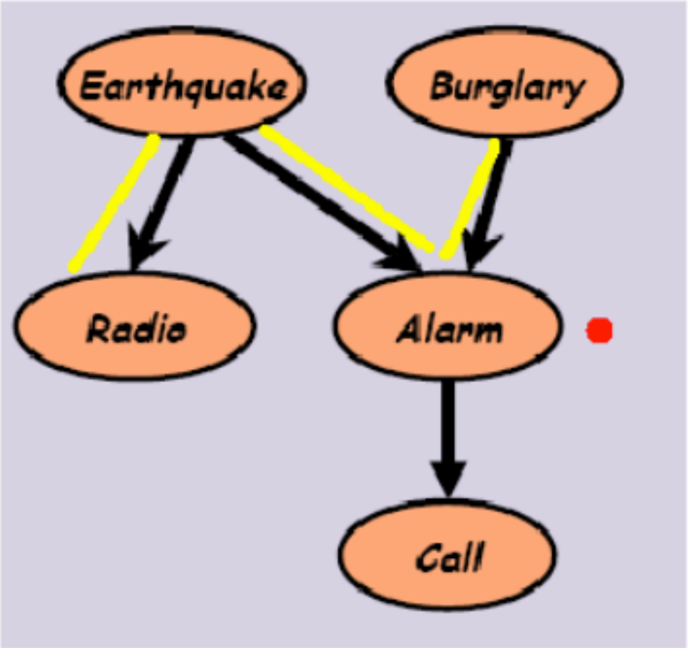


Example 2

- d-sep(X1, X3|X2)
- d-sep(R,B|A)?
  - X1 = {R}, X3={B}, X2 ={A}
  - Find all the path between R, B
  - Check the node:
    - Earthquake.

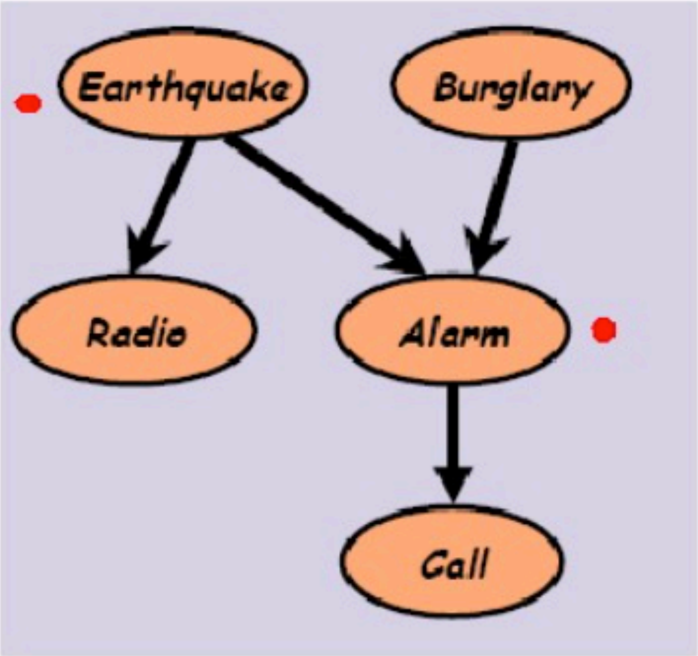
      (diverging, not in X2). Not blocking.
    - Alarm

      (Converging, A or C are **IN** X2). Not blocking!



Example 3

- d-sep(R,B|E,A)?

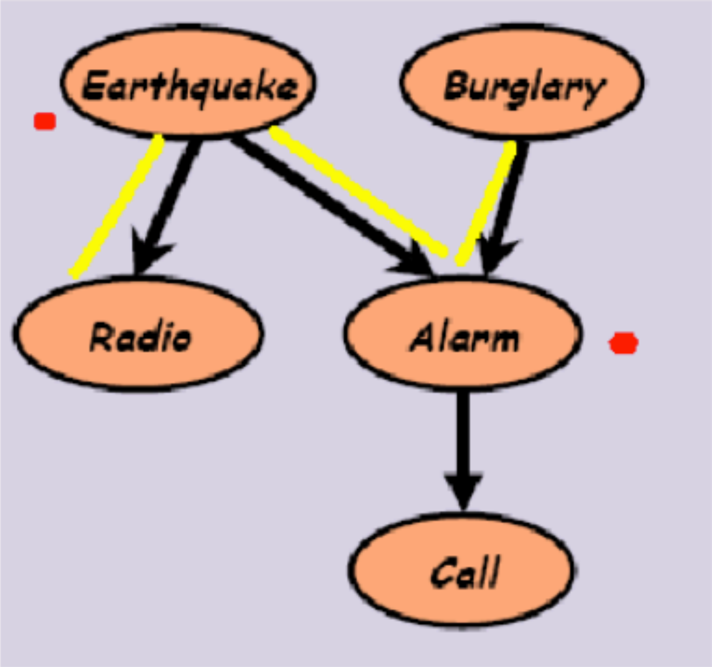


Example 3

- d-sep(X1, X3|X2)
- d-sep(R,B|E,A)?
  – X1 = {R}, X3={B}, X2 ={E,A}
  – Find all the path between R, B
  – Check the node:
    - Earthquake.

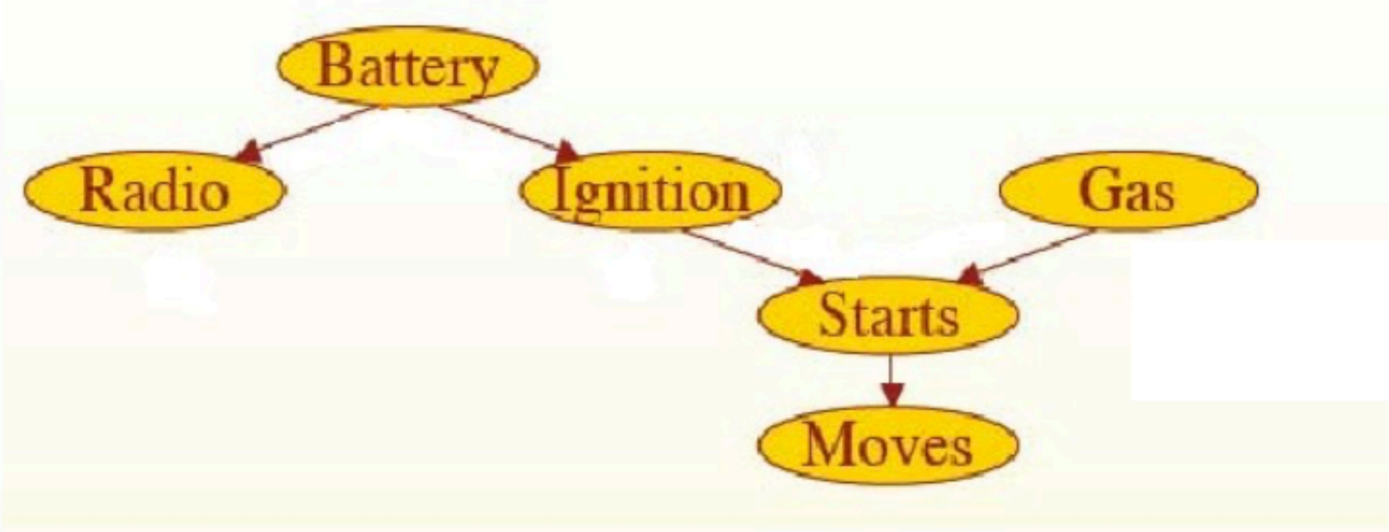
      (diverging, **IN** X2). Blocking!
    - Alarm

      (Converging, A or C are **IN** X2). Not blocking.



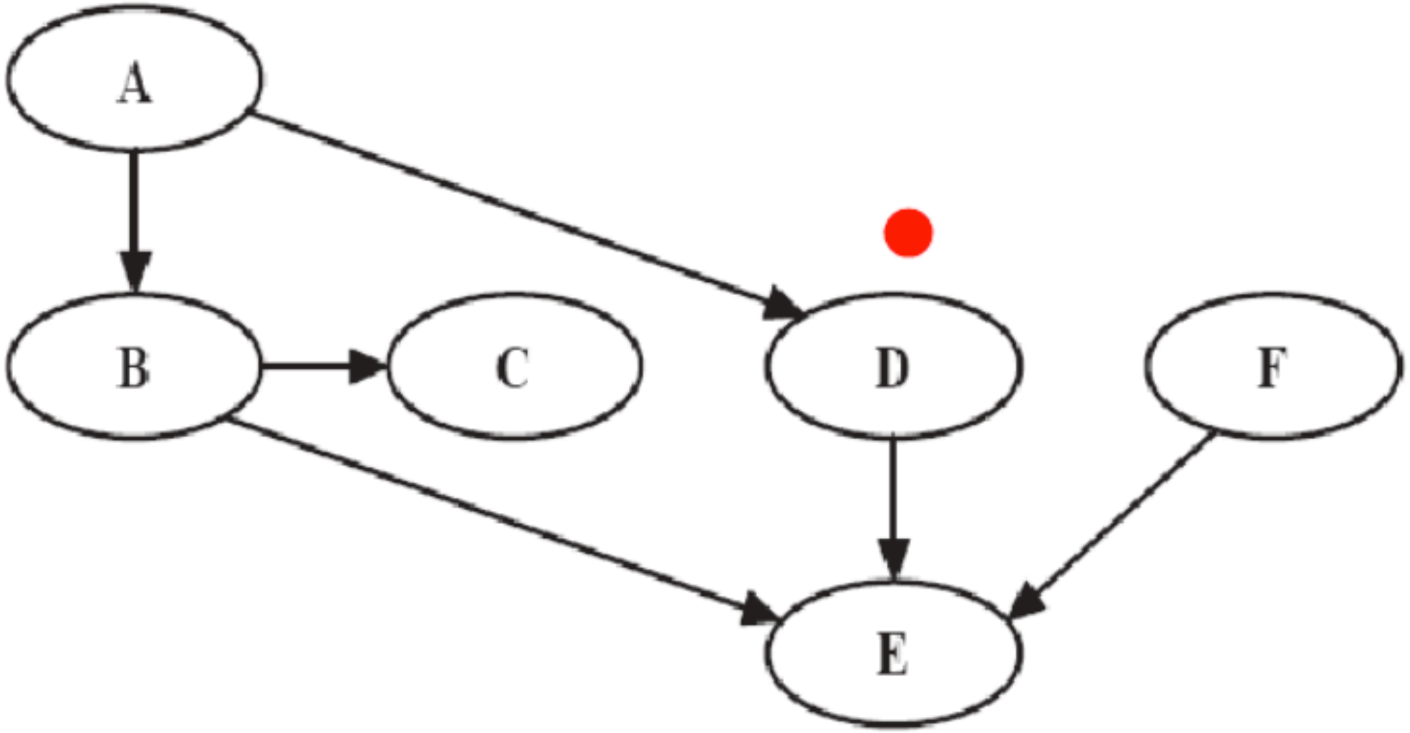
Example 4

- d-sep(Radio,Gas|Moves)?



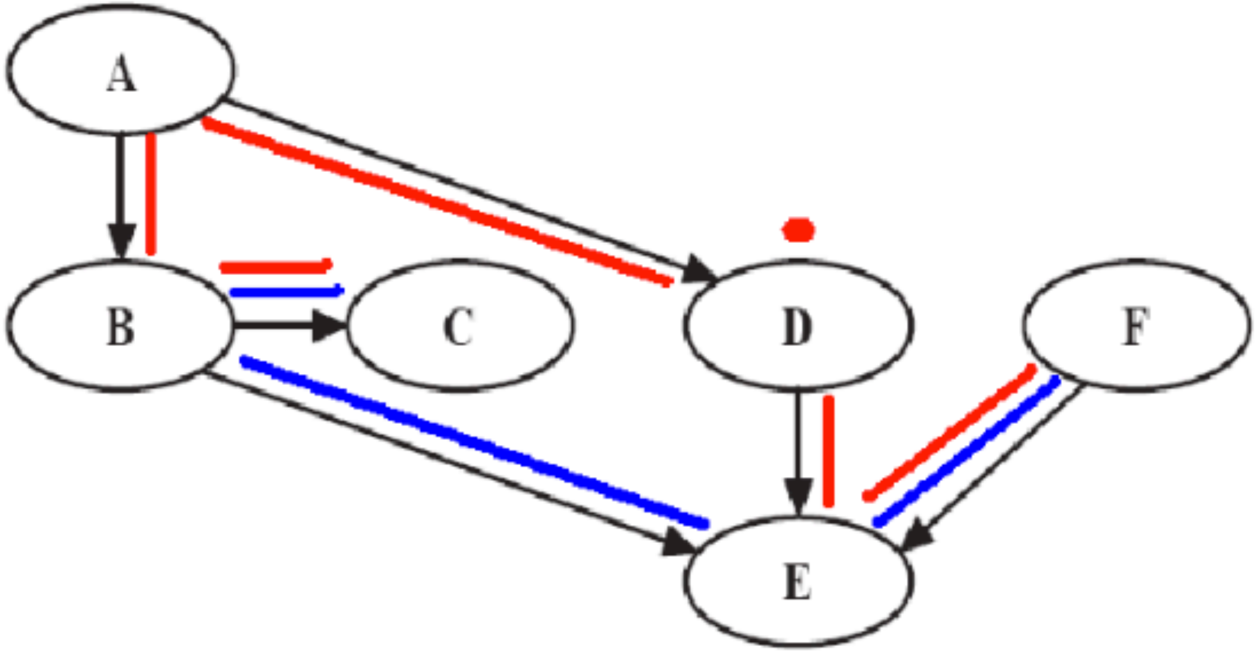
D-seperation: Multiple Paths

- d-sep({C}, {F}|{D})?



D-seperation: Multiple Paths

- d-sep({C}, {F}|{D})?

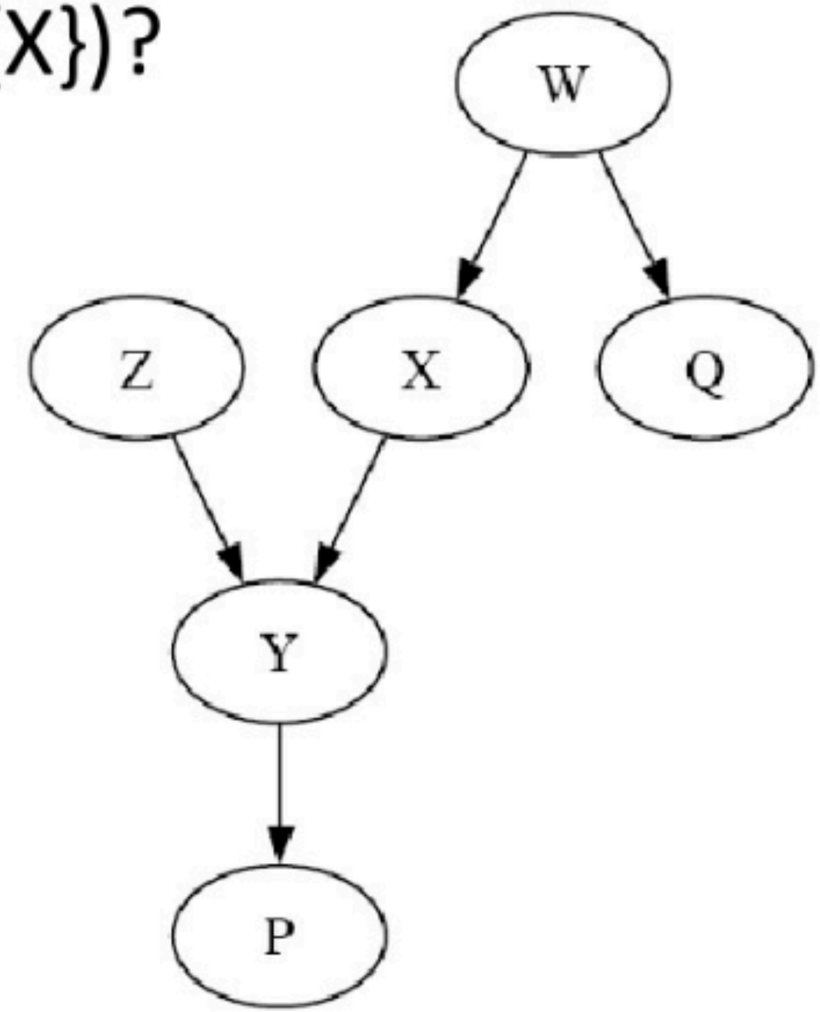


Red path is blocked by D.

Blue path is blocked by E not in evidence.

D-seperation on Sets

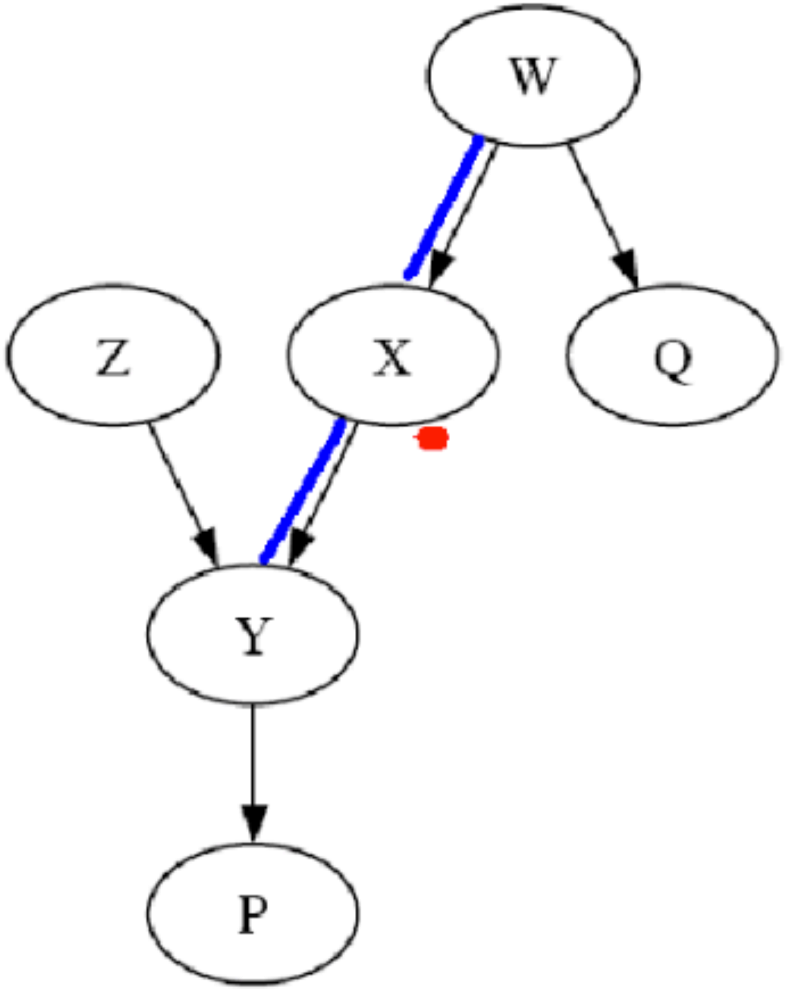
- d-sep({Z, Y, P}, {W, Q}|{X})?



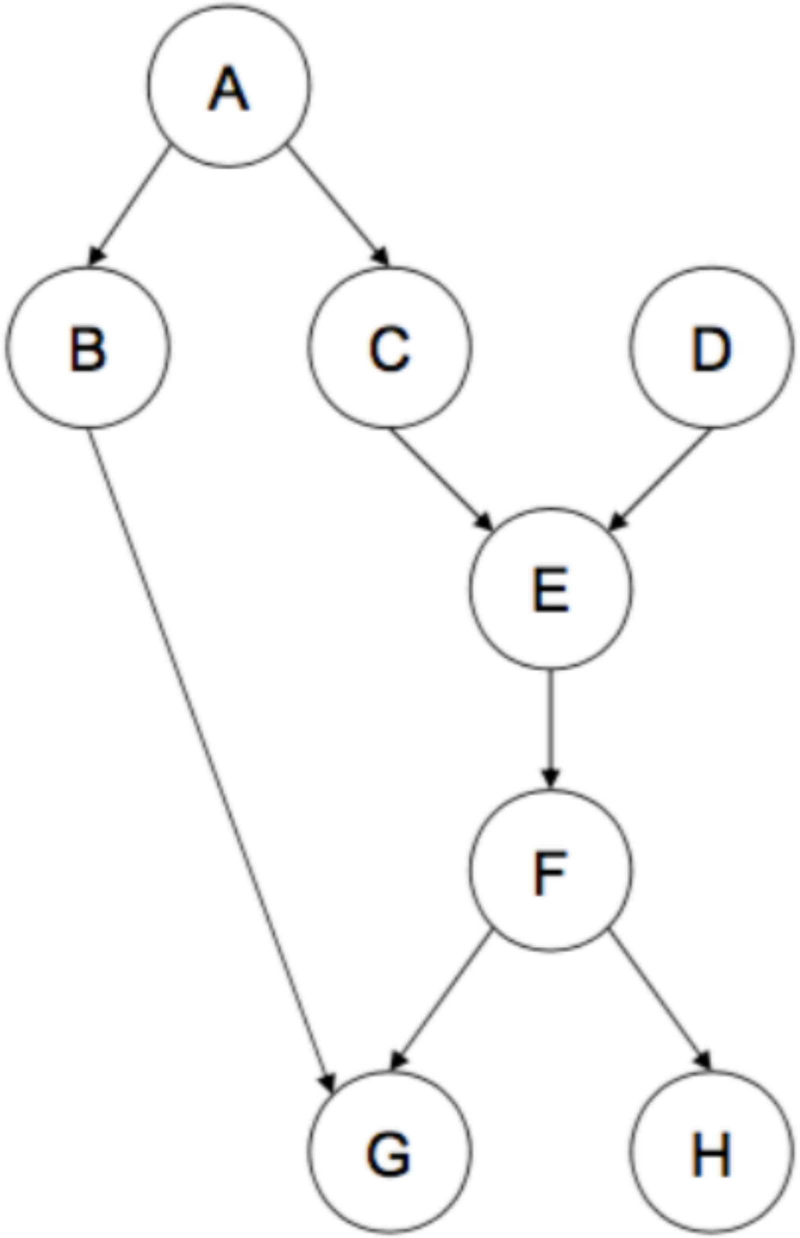
D-seperation on Sets

- d-sep({Z, Y, P}, {W, Q}|{X})? YES

Blue path is a closed sequential path since we condition on X.



D-separation Examples

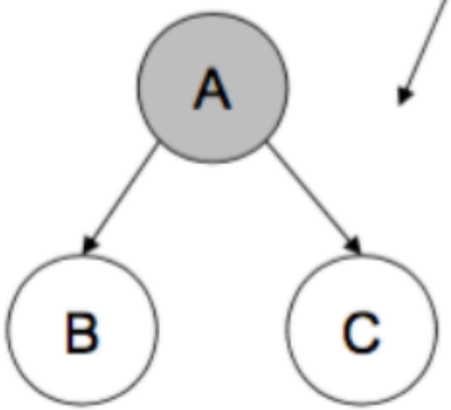


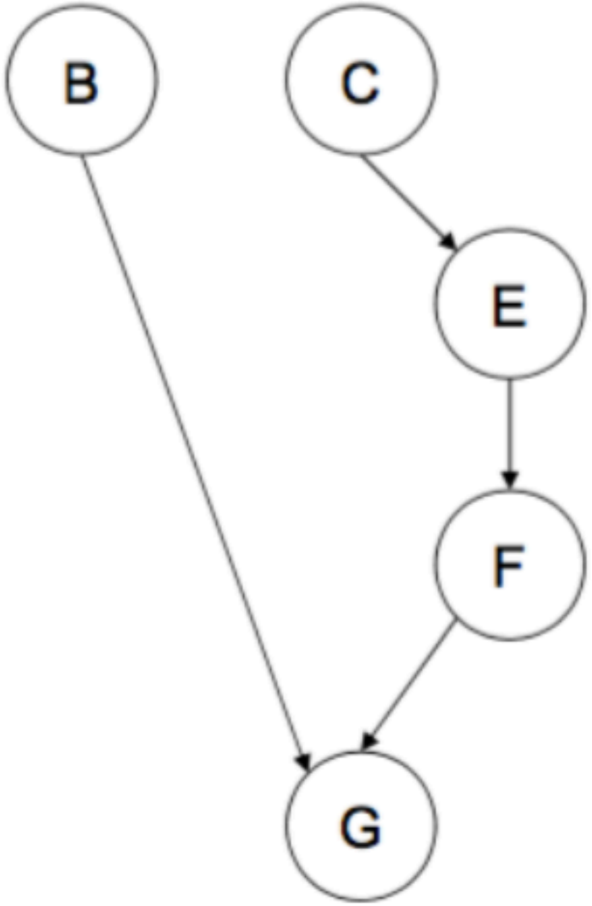
$(B \perp C \mid A)$?

D-separation Examples

(B ⊥ C | A)? Yes. Notice the two (undirected) paths between B and C

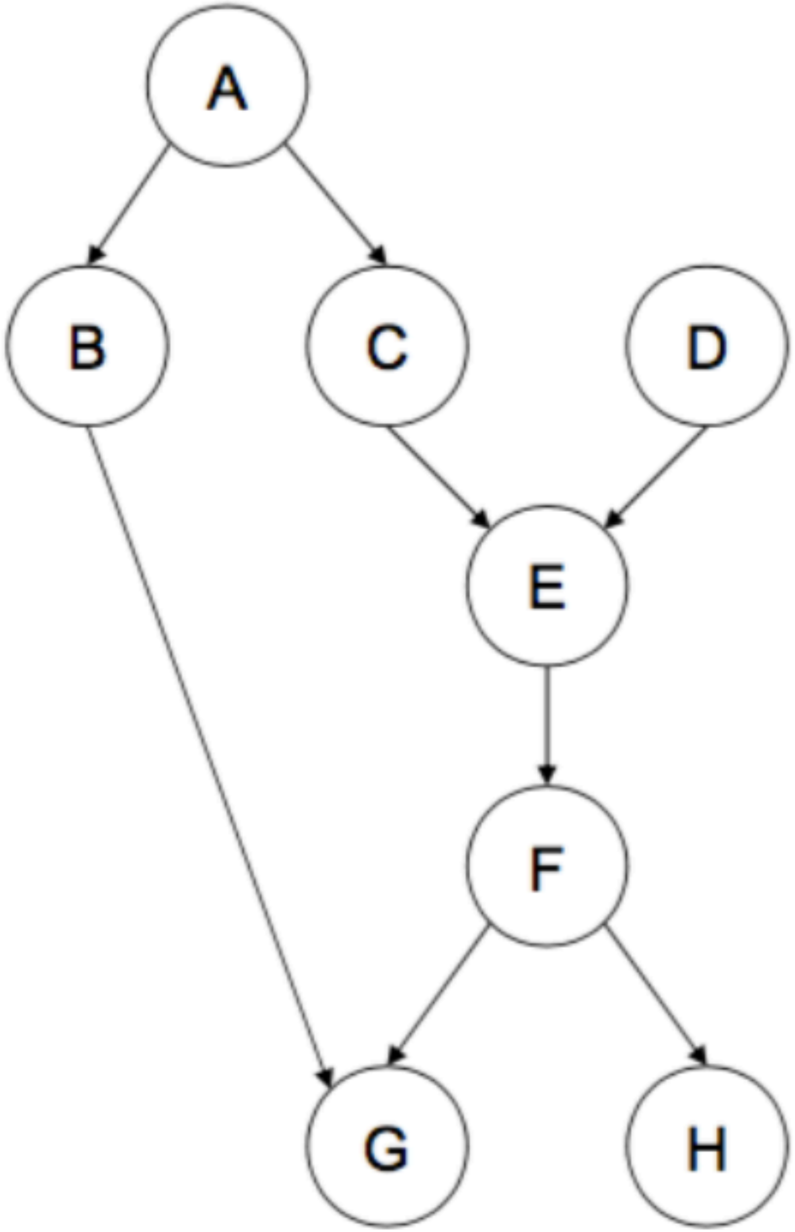
This path from B to C is blocked by A (Case 1)





This path from B to C is blocked by G, which is not in the evidence set (Case 3)

D-separation Examples

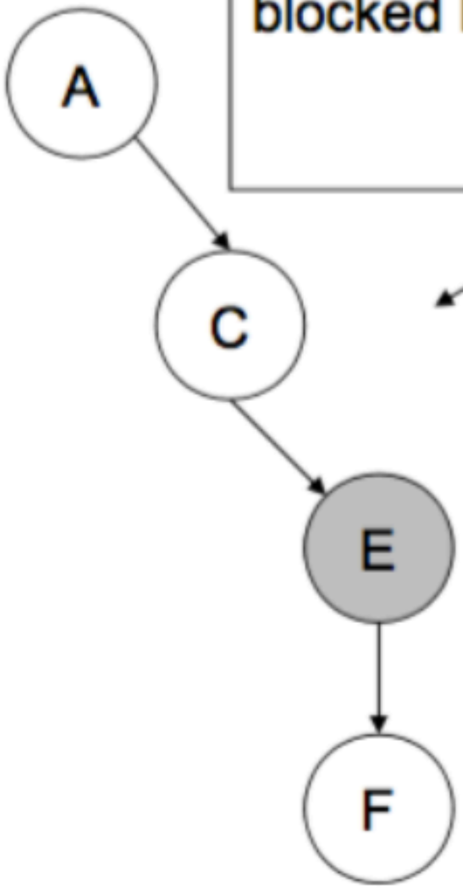


$(A \perp F \mid E)$?

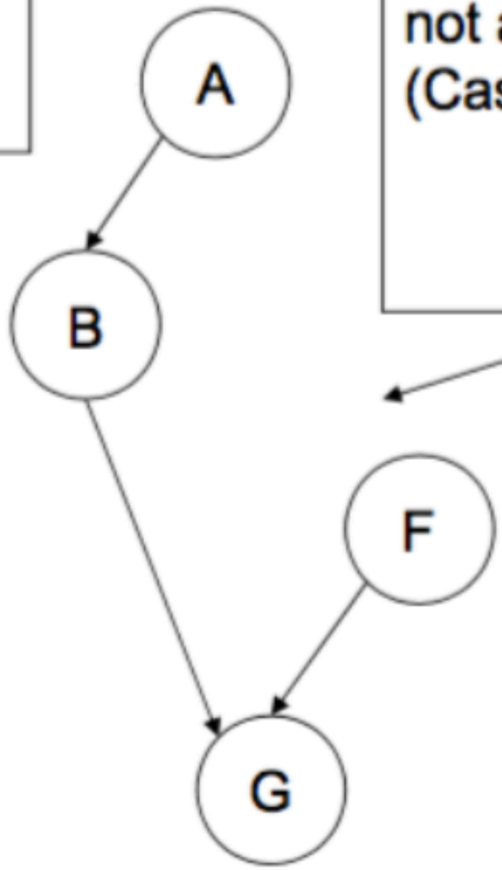
D-separation Examples

$(A \perp F \mid E)$? Yes

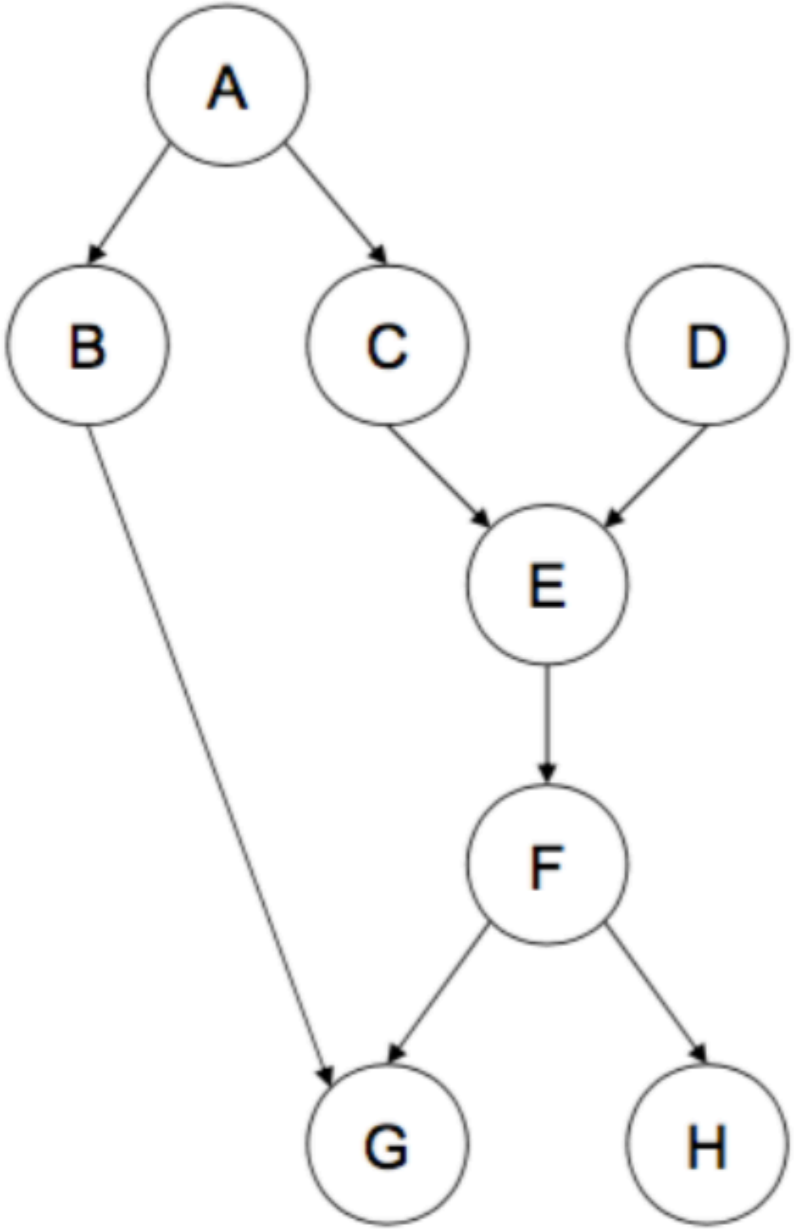
This path from A to F is blocked by E (Case 2)



This path from A to F is blocked by G, which is not an evidence node (Case 3)



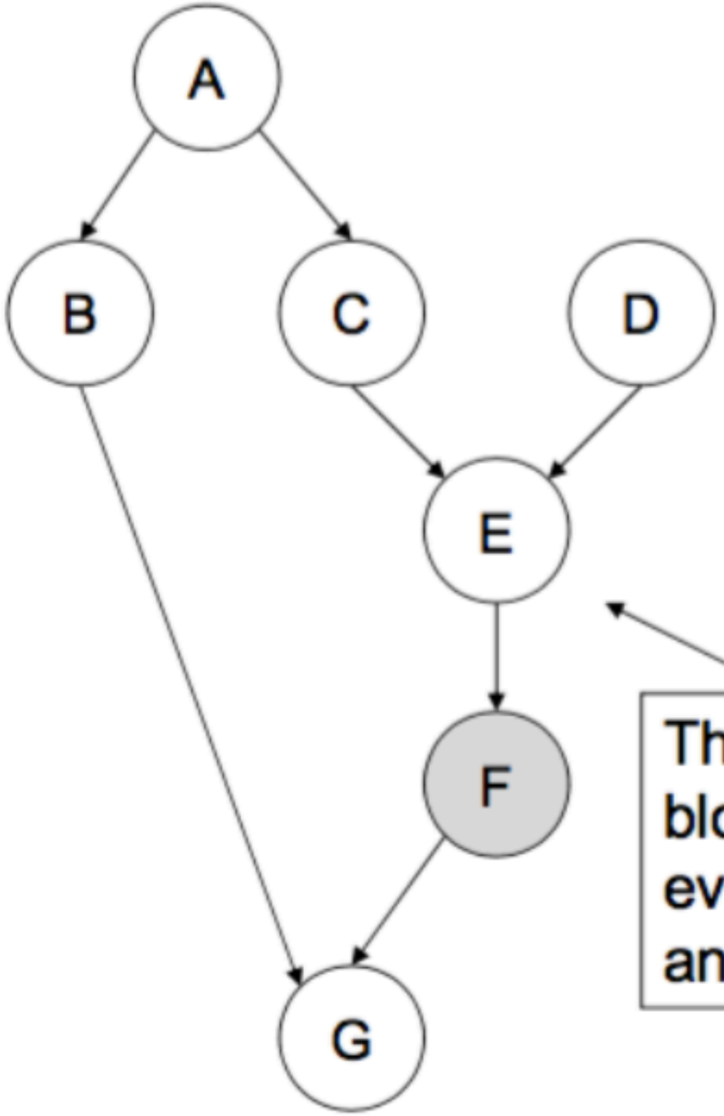
D-separation Examples



$(C \perp D \mid F)$?

D-separation Examples

$(C \perp D \mid F)$? No



This path from C to D is blocked by G (not in evidence set) (Case 3) and by F (Case 2)

But this path from C to D is **not** blocked. This is because F (which is a descendant of E) is in the evidence set (Case 3)