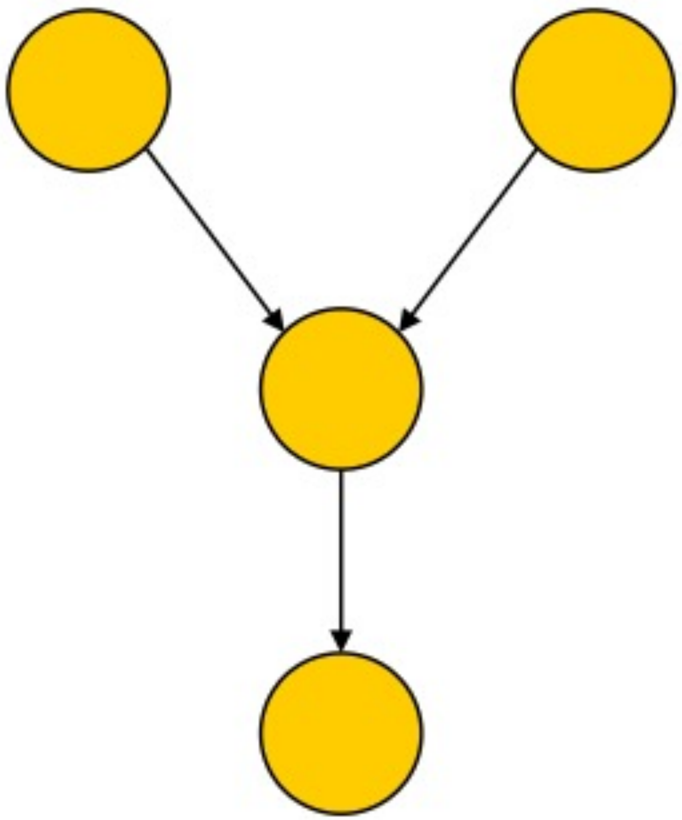
Bayes Net



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:

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

- 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:

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

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

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

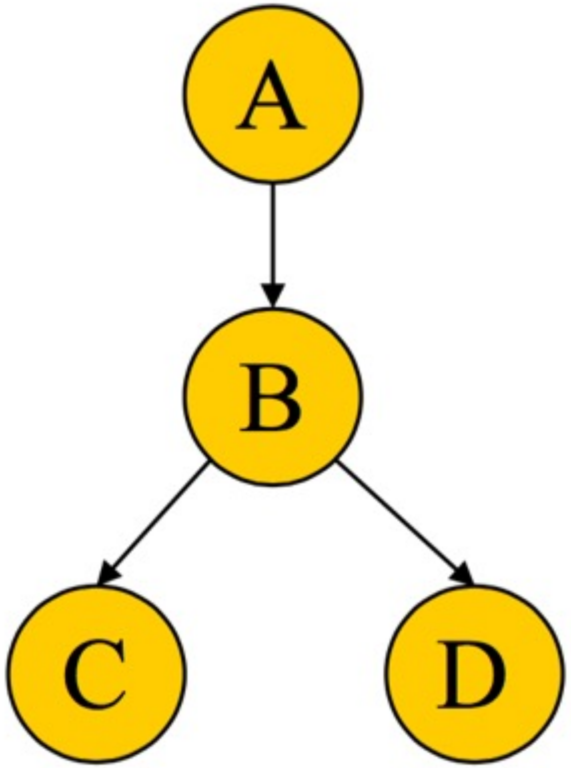
Notation: $A \perp\!\!\!\perp B \mid C$

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

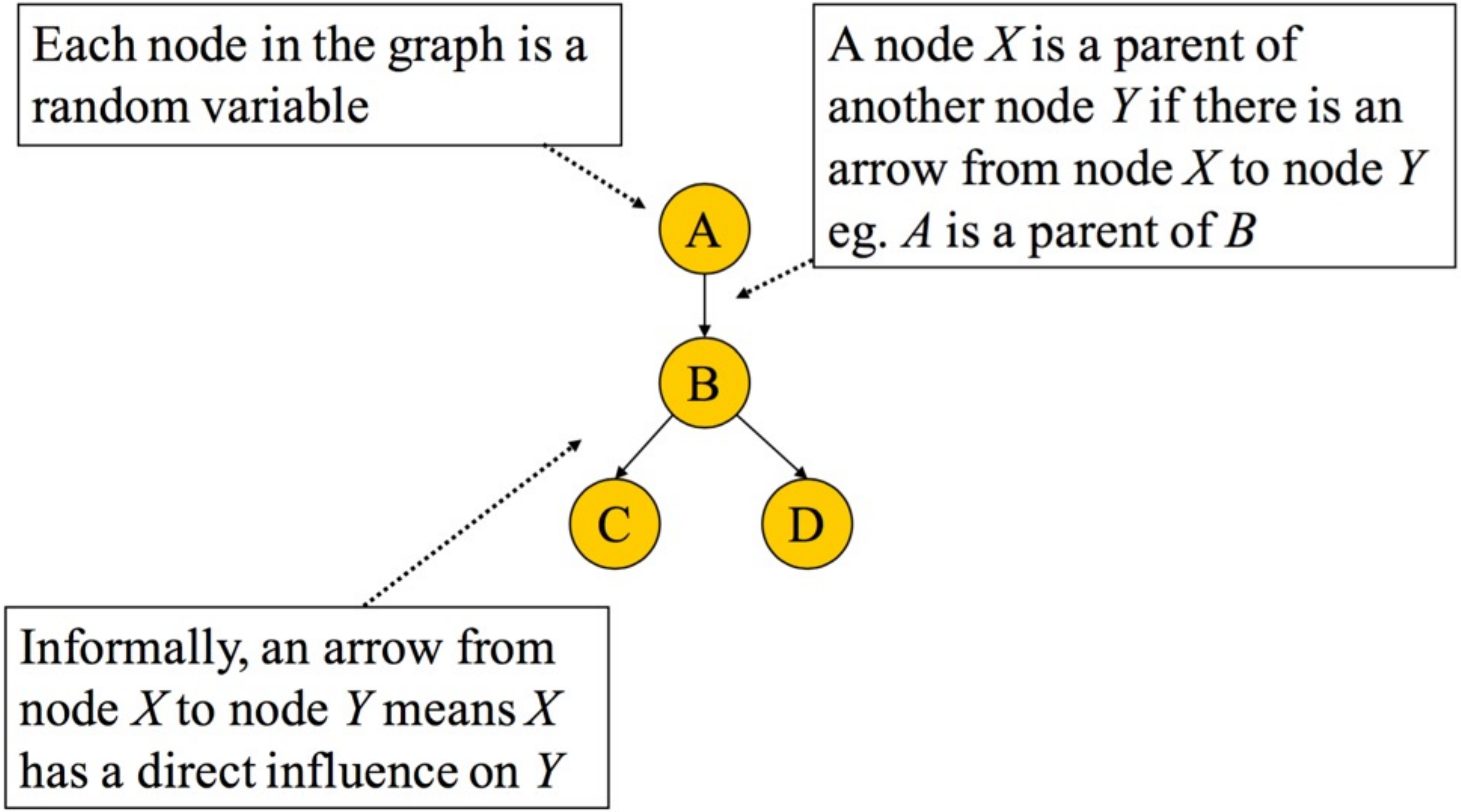
A	P(A)
false	0.6
true	0.4

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
A
B
C
D
Informally, an arrow from node X to node Y means X has a direct influence on Y

A Set of Tables for Each Node

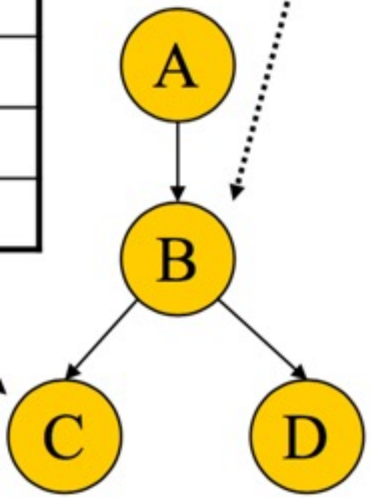
A	P(A)
false	0.6
true	0.4

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

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

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

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

A Set of Tables for Each Node

Conditional Probability
Distribution for C given B

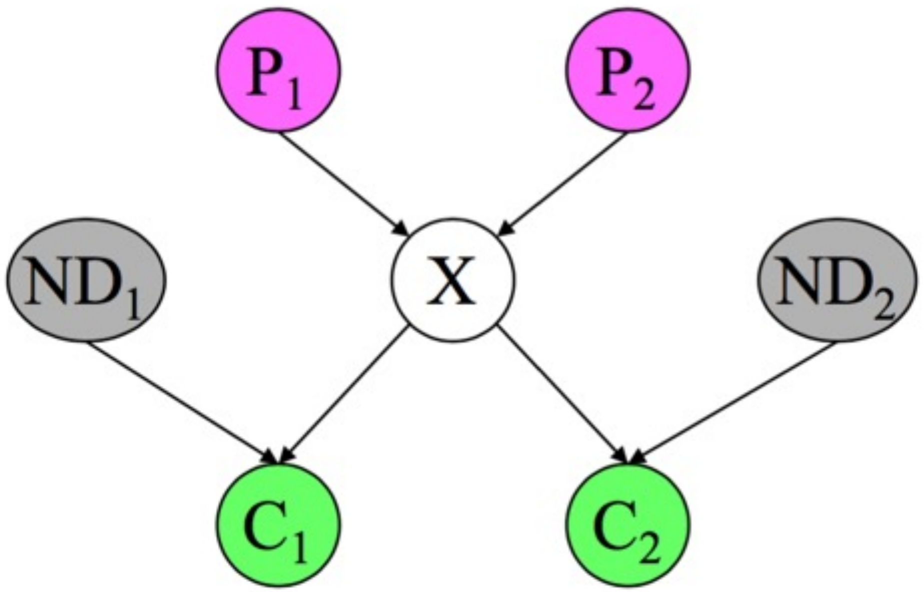
B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

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, …, 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:

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

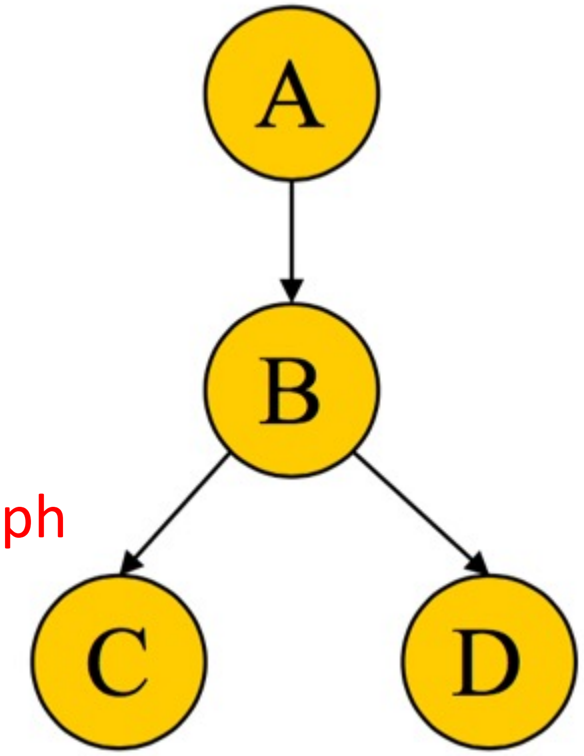
  $= P(A = true) * P(B = true \mid A = true) * P(C = true \mid B = true) * P(D = true \mid B = true)$

  This knowledge comes from graph structure

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

  This knowledge comes from CPT

  $= 0.0114$



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

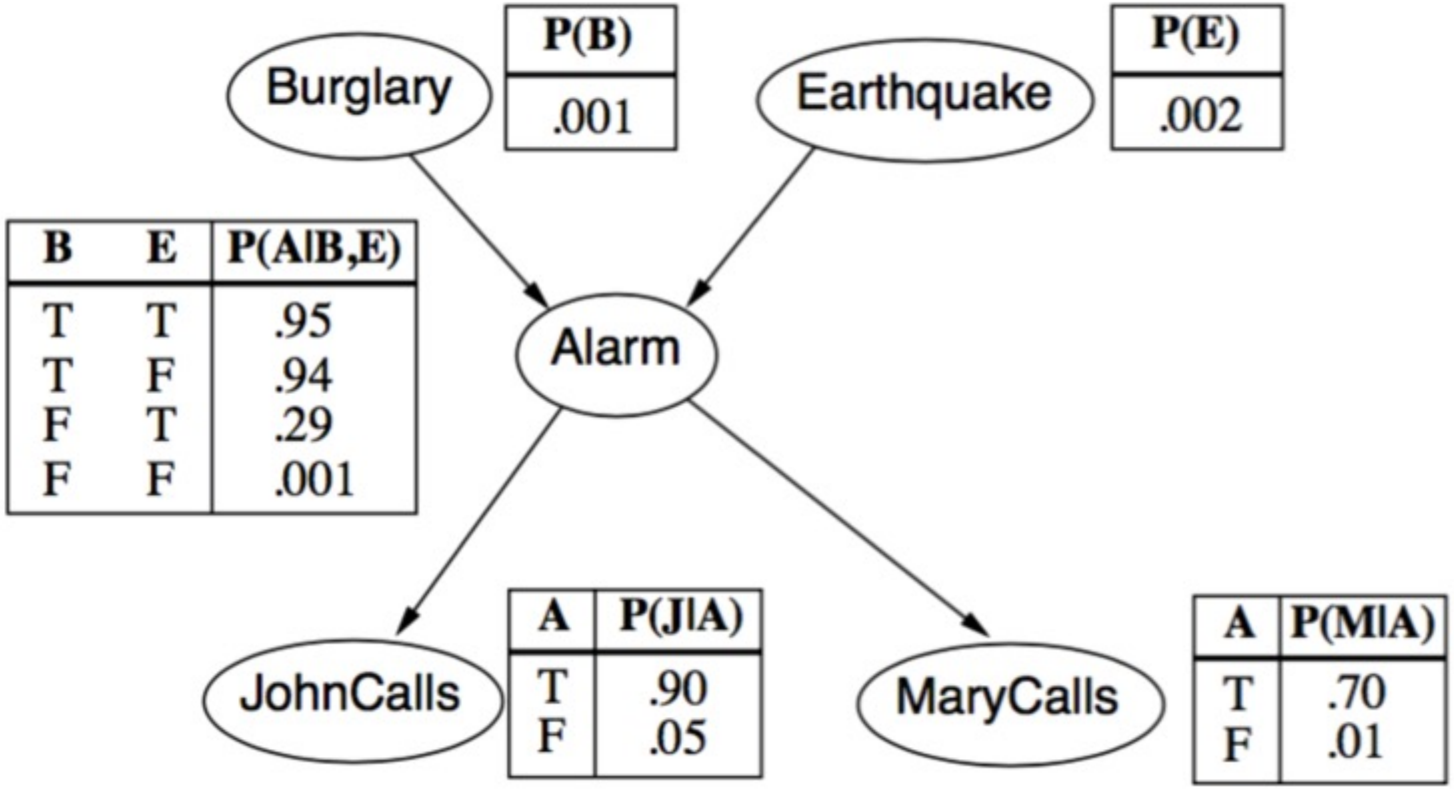
B	C	P(C|B)
false	false	0.4
false	true	0.6
true	false	0.9
true	true	0.1

Example

- Consider the Bayes Net:

- How many parameters?
10

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

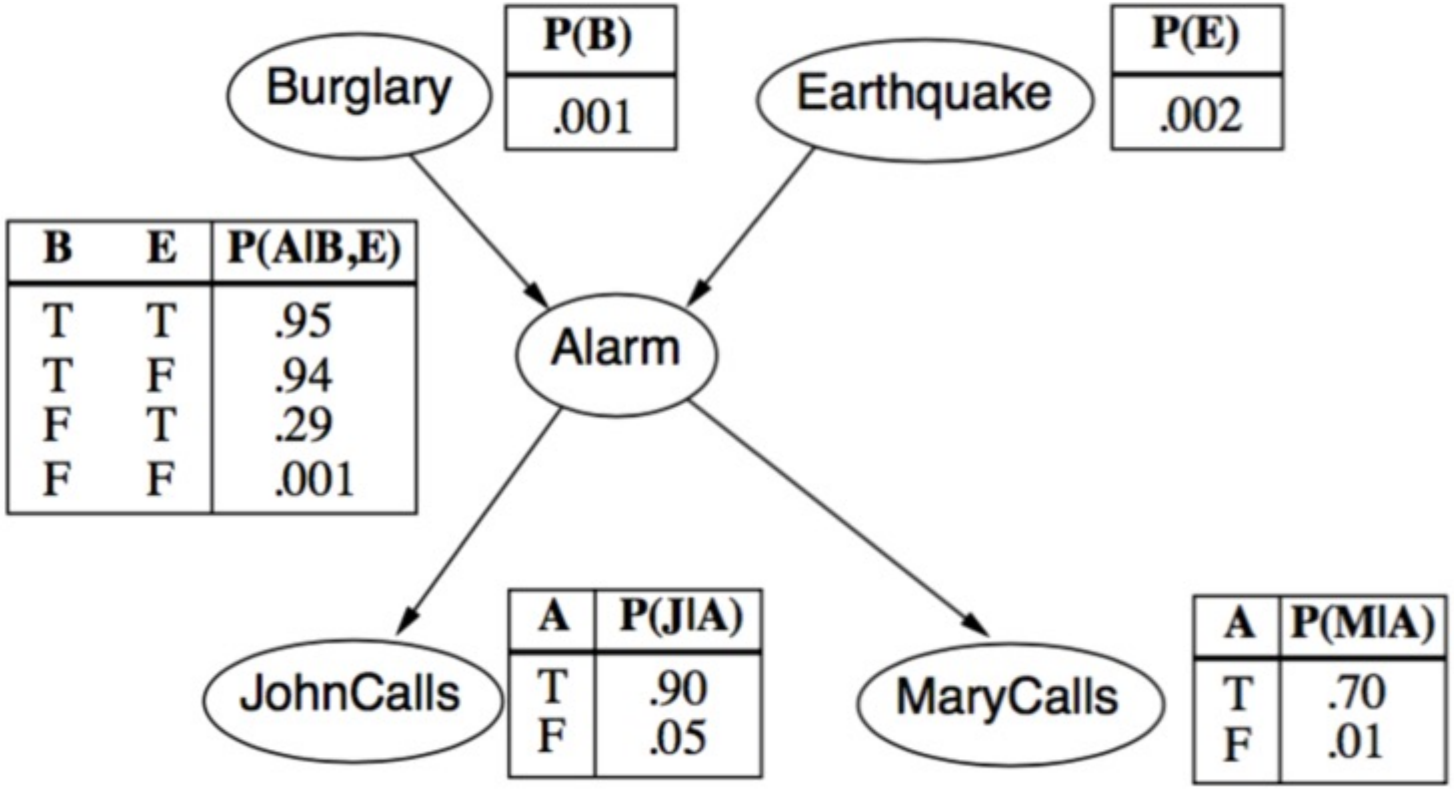


Example

- 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:

- Use it to evaluate:

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

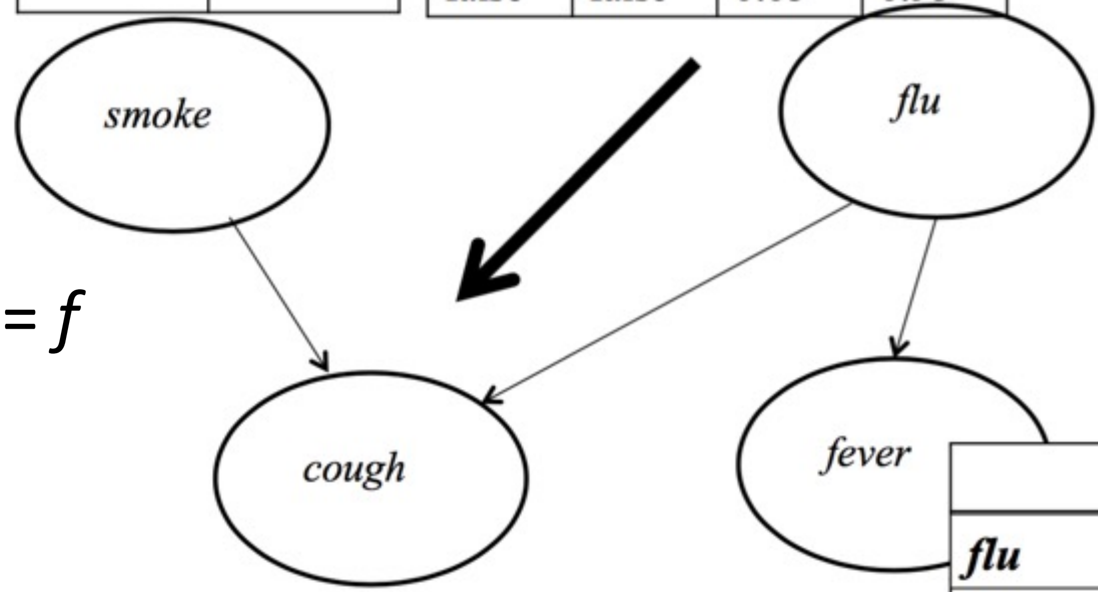
Example:

Conditional probability tables for each node

	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

	flu
true	0.01
false	0.99



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

Example:

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

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

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

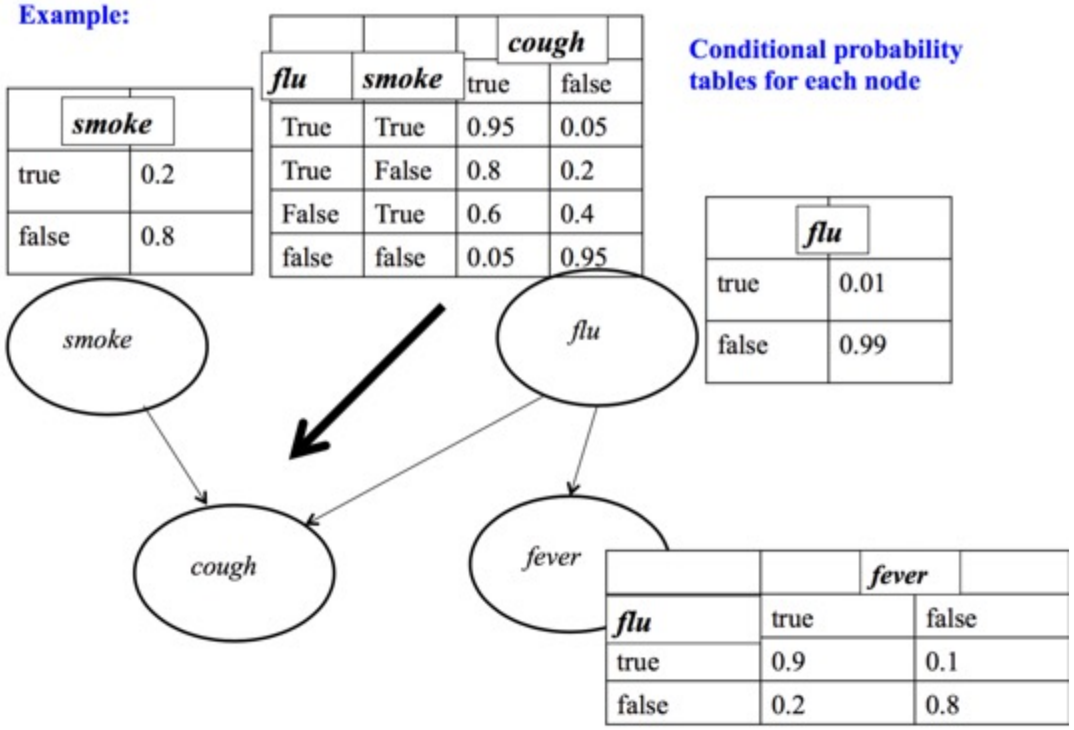
$$\times P(fever = f \mid flu = f)$$

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

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

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

$$= .032$$



	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

	flu
true	0.01
false	0.99

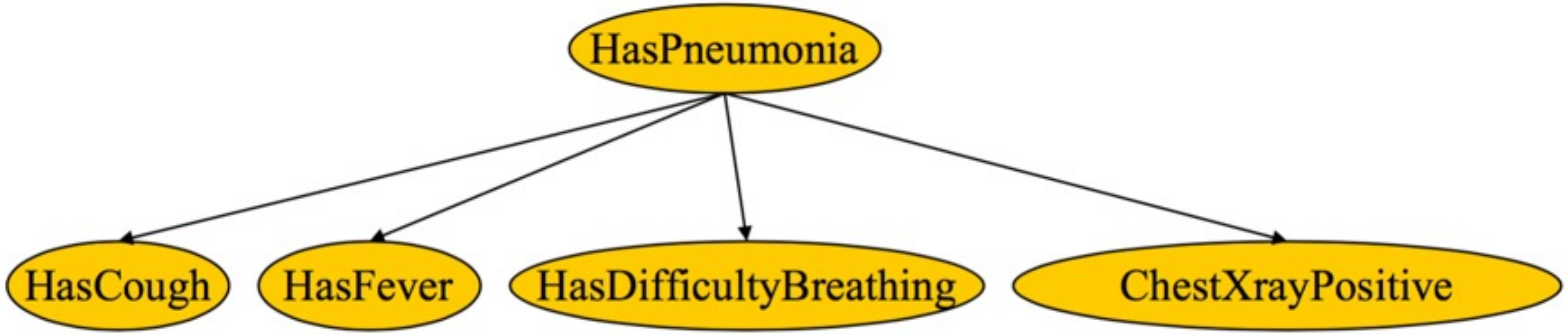
flu	fever true	fever false
true	0.9	0.1
false	0.2	0.8

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

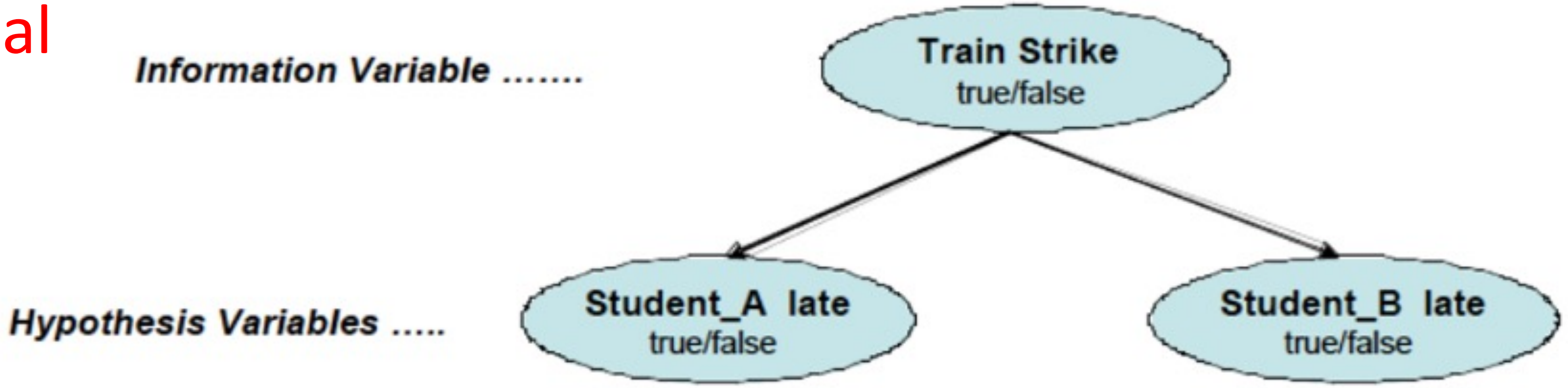
Another Bayesian Belief Network (BBN)

- 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



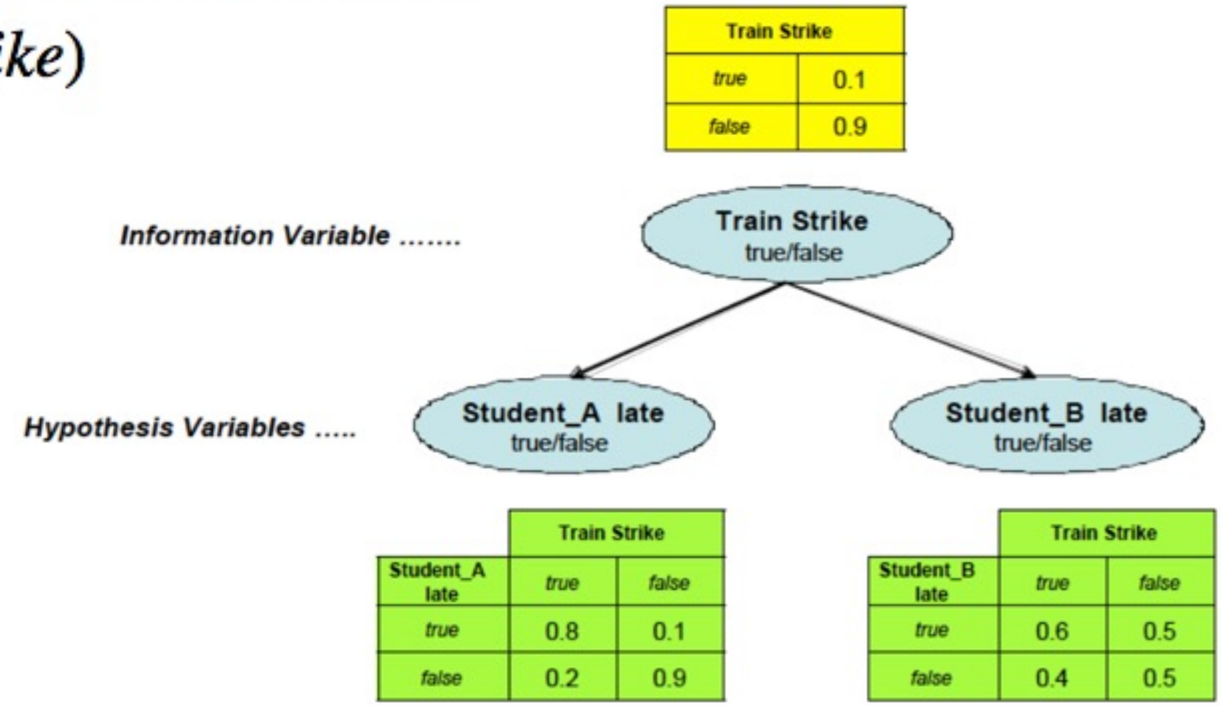
	Train Strike	
Student_A late	*true*	*false*
true	0.8	0.1
false	0.2	0.9

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

Marginal Probability

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

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



Evidence About Parent Given

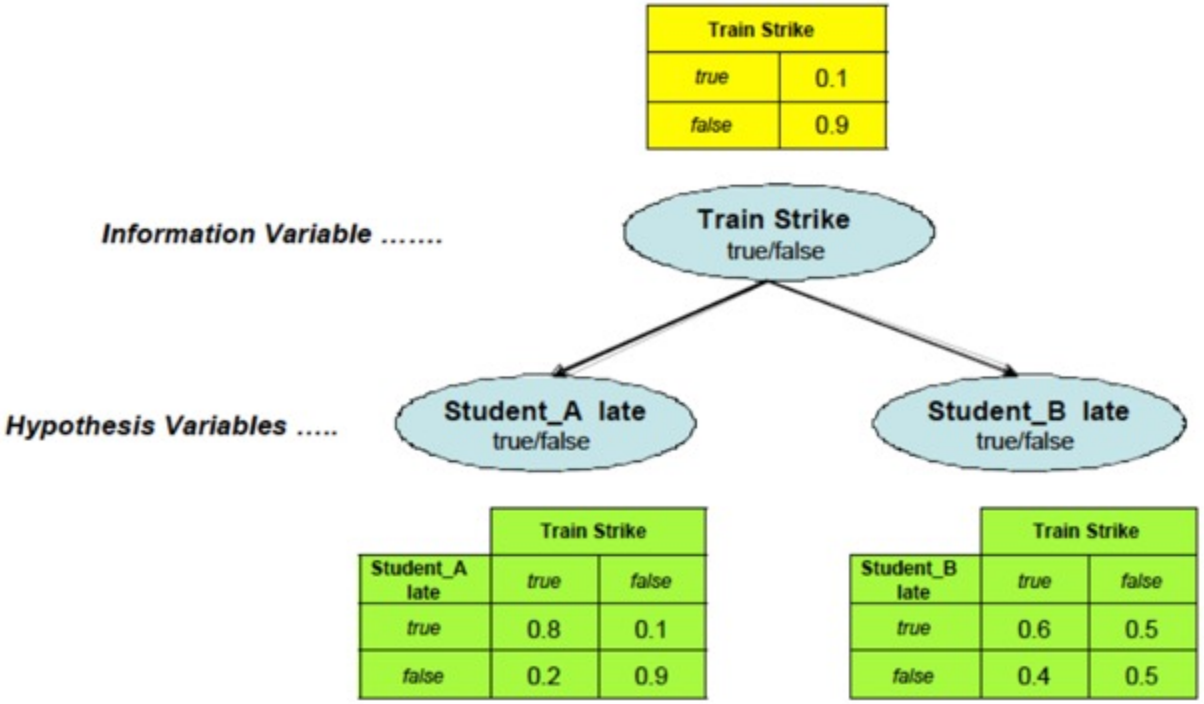
- Given that there is a train strike, what's the probability that A is late. P(Student_A late | Train Strike)

- Simple – just look up the table

Evidence: There is a train strike.

$$P(StudentALate) = 0.8$$

$$P(StudentBLate) = 0.6$$

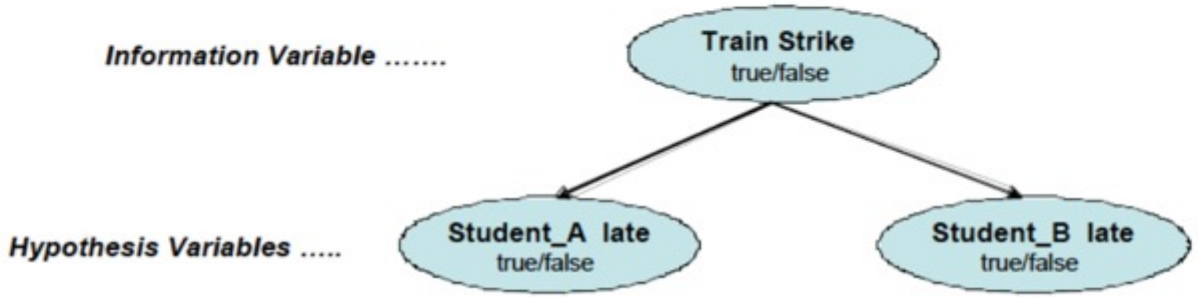


Evidence About Child Node Given

- 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

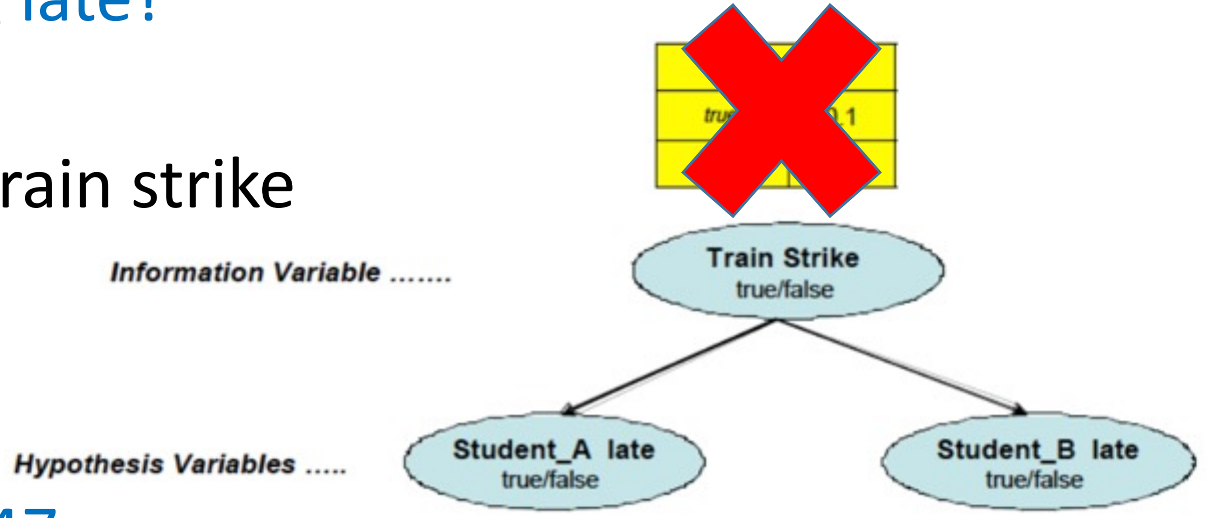
$$P(TrainStrike \mid StudentALate) = \frac{P(StudentALate \mid TrainStrike)P(TrainStrike)}{P(StudentALate)} \text{ by Bayes Theorem}$$

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

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



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

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

$$P(StudentBLate) = P(StudentBLate \mid TrainStrike)P(TrainStrike)$$

$$+P(StudentBLate \mid \neg TrainStrike)P(\neg TrainStrike)$$

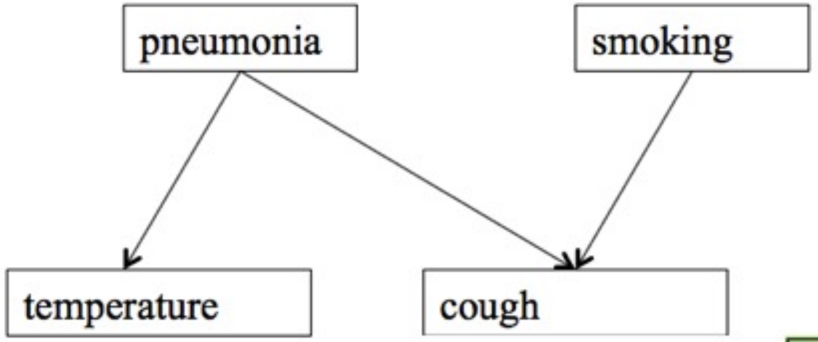
$$= 0.6 \times 0.47 + 0.5 \times 0.53 = 0.55$$

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:

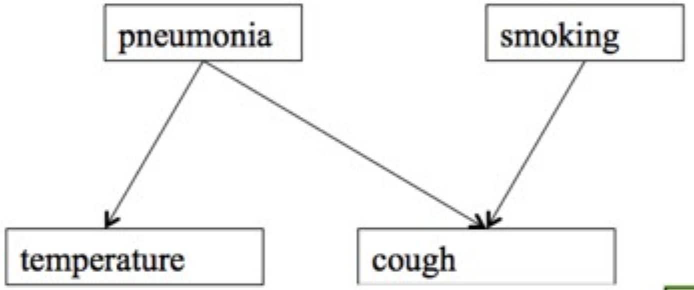
  P(Cough | Pneumonia = F)

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

  P(Cough)

pneumonia	
true	0.1
false	0.9

smoking	
yes	0.2
no	0.8



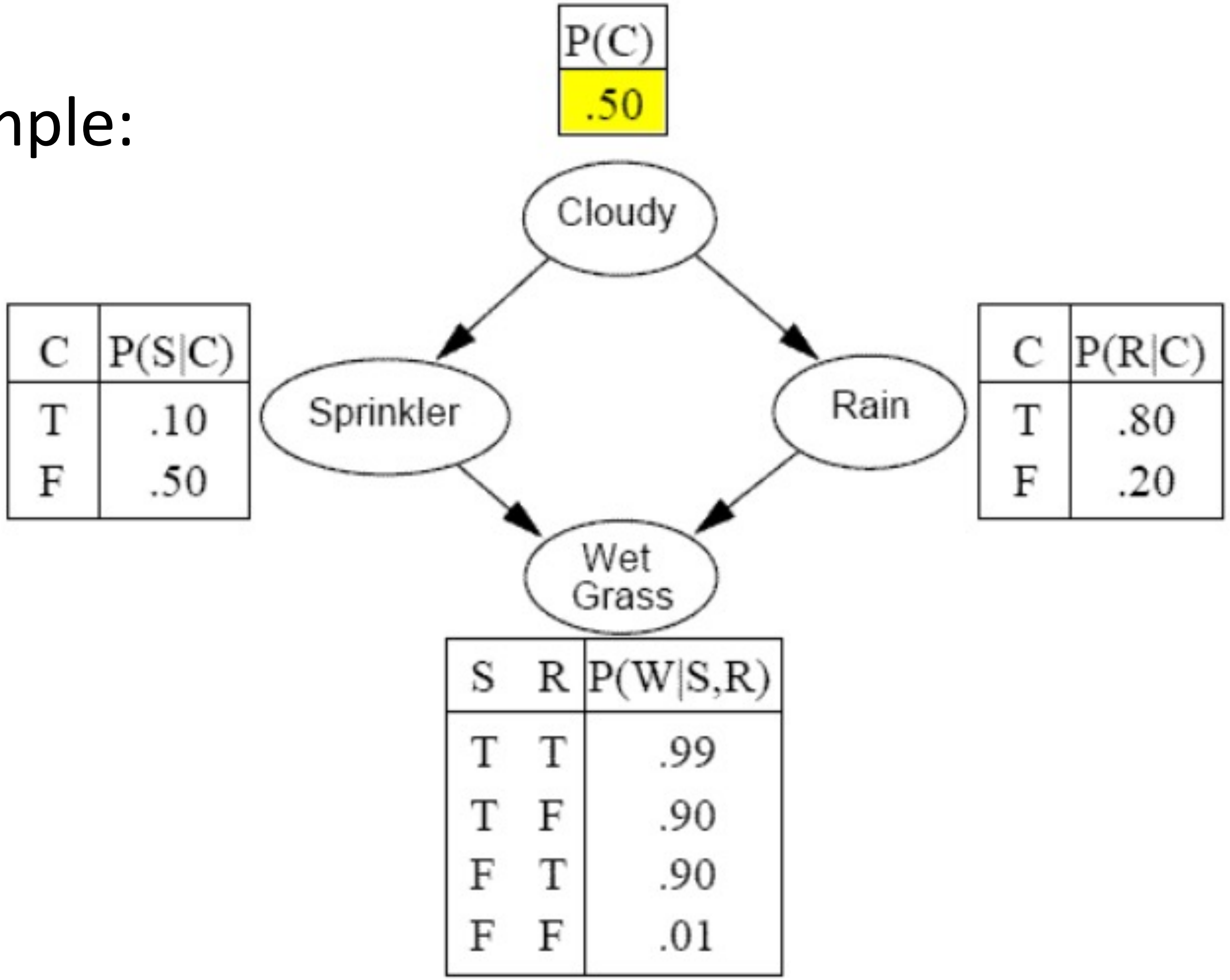
	temperature	
pneumonia	yes	no
yes	0.9	0.1
no	0.2	0.8

		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



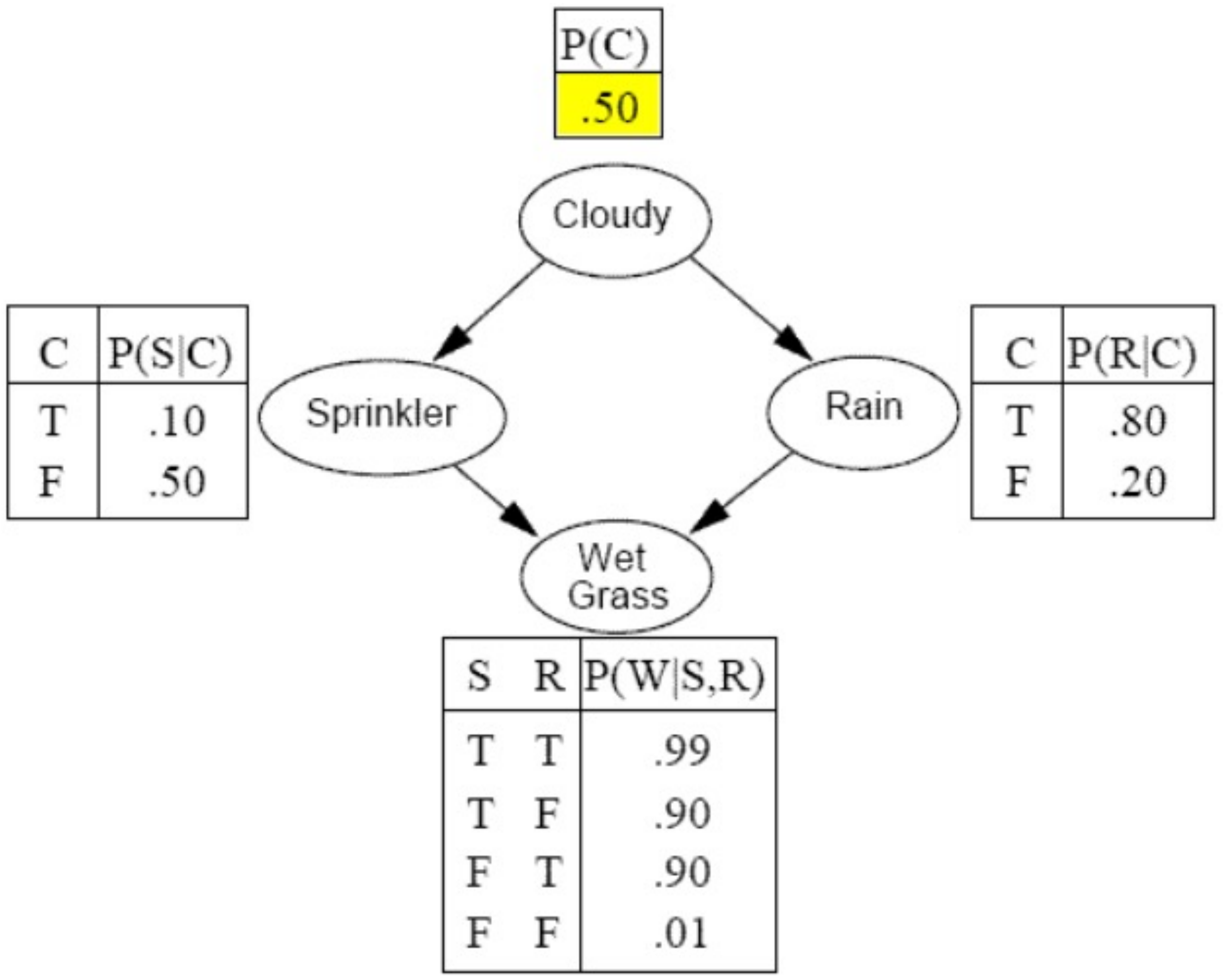
More Exercises:

- Now, some complex queries:

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

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

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



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*)?

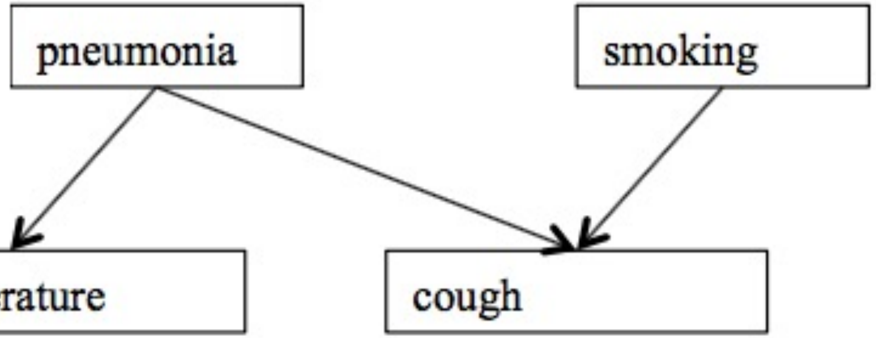
- **Inter-causal inference:** "Explain away" potentially competing causes of a shared effect.

– 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

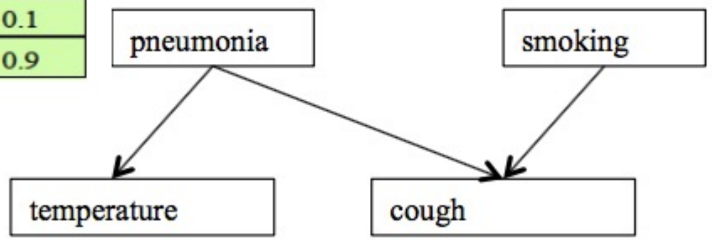
	temperature	
pneumonia	yes	no
yes	0.9	0.1
no	0.2	0.8

		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 $P(pneumonia \mid cough)$?

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

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

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

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

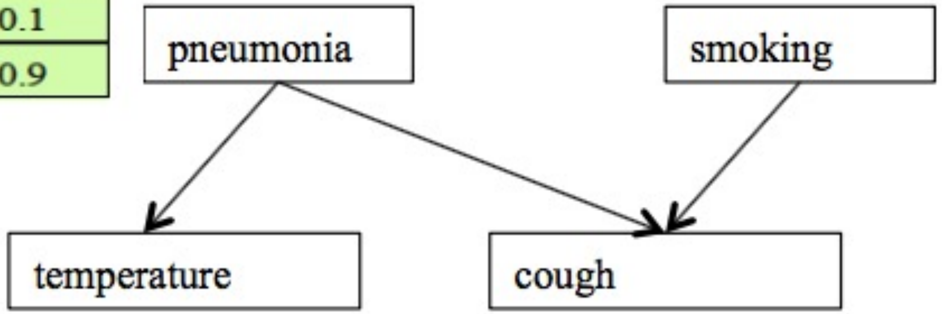
$$= \frac{.083}{P(cough)} = \frac{.083}{.227} = .366$$

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

Causal Queries

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

pneumonia	
true	0.1
false	0.9



smoking	
yes	0.2
no	0.8

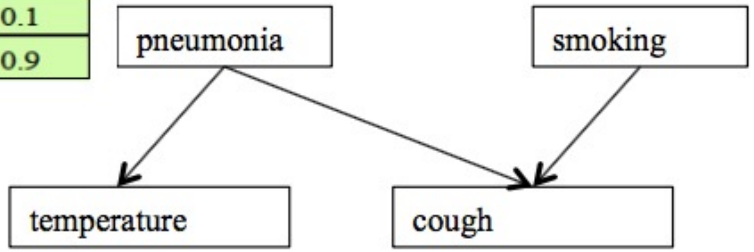
	temperature	
pneumonia	yes	no
yes	0.9	0.1
no	0.2	0.8

		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 $P(cough \mid pneumonia)$?

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

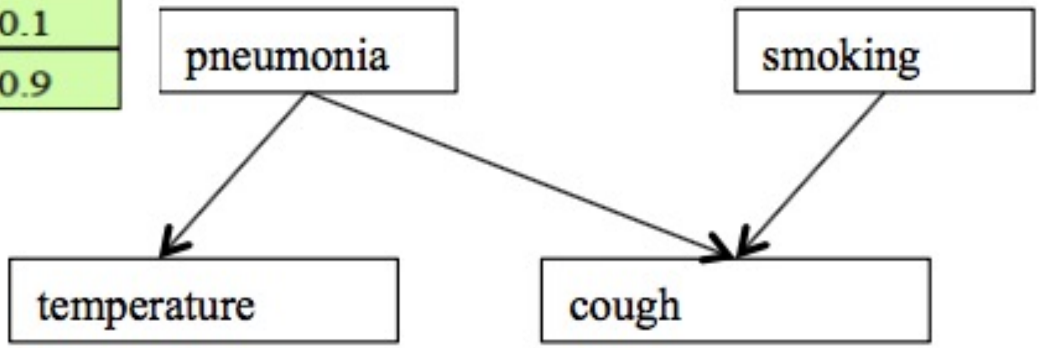
$$P(cough \mid pneumonia) = P(cough \mid pneumonia, smoking)P(smoking)$$
$$+P(cough \mid pneumonia, \neg smoking)P(\neg smoking)$$
$$= [(.95)(.2) + (.8)(.8)] = .83$$

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"

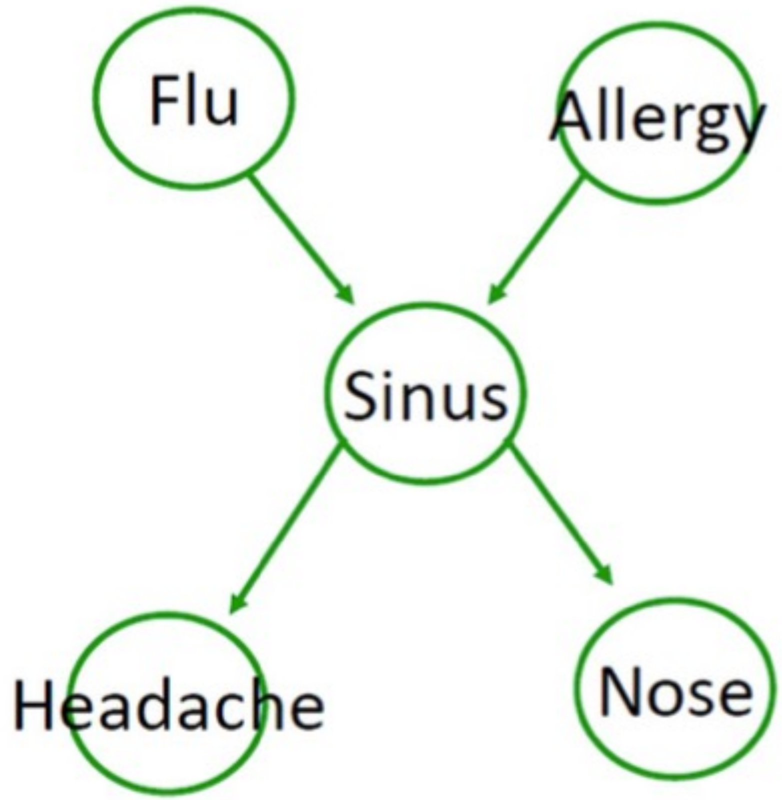
Dependencies and Conditional Independence in a BN

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)

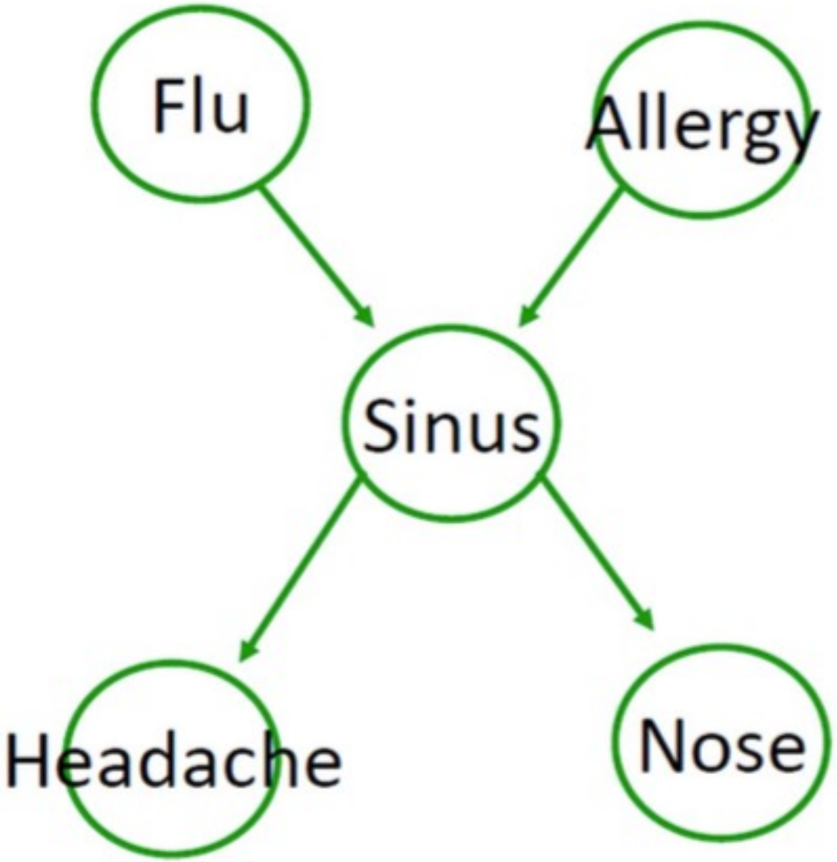
$= P(F)\ P(F|A)\ P(S|F,A)\ P(H|S,F,A)\ P(N|S,F,A,H)$

Chain rule

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

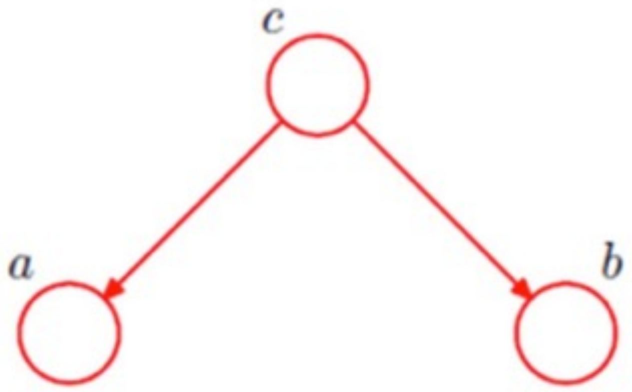
Markov Assumption

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



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?

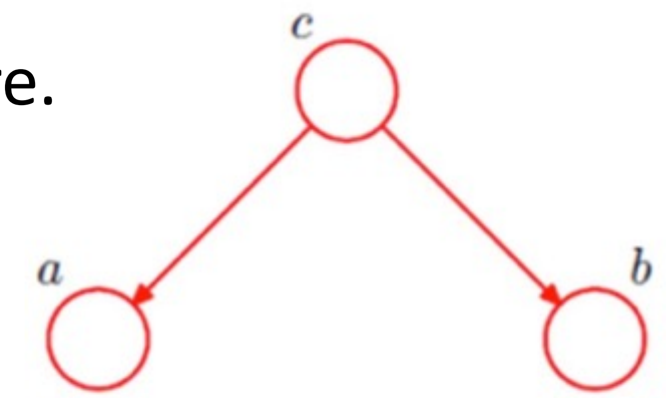


Show that $a \perp\!\!\!\perp b \mid c$

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

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*