Uncertainty

AIMA Chapter 13

Axioms of Probability

- Total probability of a set of possible worlds is 1

$$\sum_{x \in D_X} p_X(x) = 1$$

- Probability of an event is the sum of probabilities of the worlds in which it holds

$$\Pr[X \subseteq A] \equiv \Pr_X[A] = \sum_{x \in A} p_X(x)$$

- Probability of a disjunction (inclusion-exclusion principle)

$$\Pr[A] + \Pr[B] = \Pr[A \wedge B] + \Pr[A \vee B]$$

We can think about ∩ as ∧ and ∪ as ∨

- Conditional probability/product rule

$$\Pr[A|B] = \frac{\Pr[A \wedge B]}{\Pr[B]}$$

Instead of $X_1 \wedge \cdots \wedge X_k \Rightarrow Y$, we infer the likelihood of Y, given probabilities of other events, $\Pr[Y|X_1, \ldots X_k] = ?$

Inference by Enumeration

- Start with the joint probability distribution:

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- For any proposition (event) X, sum the atomic events y where X holds: $\Pr[X] = \sum_{y \in X} \Pr[X = y]$
- $\Pr[\text{toothache}] = 0.108 + 0.016 + 0.012 + 0.064 = 0.2$

Marginalization / summing out

$\Pr[toothache] = \sum_z \Pr[toothache, z]$ where z is all each possible value of other variables.

Inference by Enumeration

- Start with the joint probability distribution:

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- For any proposition (event) X, sum the atomic events y where X holds: $\Pr[X] = \sum_{y \in X} \Pr[X = y]$
- $\Pr[\neg\text{cavity} \mid \text{toothache}] = \dfrac{\Pr[\neg\text{cavity} \wedge \text{toothache}]}{\Pr[\text{toothache}]}$

$$= \frac{0.016 + 0.064}{0.108 + 0.016 + 0.012 + 0.064} = 0.4$$

Inference by Enumeration

- Start with the joint probability distribution:

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- $\Pr[\text{cavity} \mid \text{toothache}] = \dfrac{\Pr[\text{cavity} \wedge \text{toothache}]}{\Pr[\text{toothache}]}$

Distribution

$$\mathbf{Pr}[\text{cavity}|\text{toothache}] = \alpha\langle 0.12, 0.08\rangle$$

Normalization

$\Pr[\mathit{toothache}]$ is common. Treat it as a constant α (normalization constant)

Posterior/Conditional Probability

$\Pr[A \mid B] = \frac{\Pr[A \wedge B]}{\Pr[B]}$ assuming that $\Pr[B] > 0$

Bayes rule: $\Pr[\, A \mid B \,] = \frac{\Pr[B|A]\,\Pr[A]}{\Pr[B]}$

Chain rule: derived by successive application of Bayes' rule:

$$\Pr[X_1 \wedge X_2 \wedge \cdots \wedge X_k] = \prod_{j=1,\ldots,k} \Pr\left[\, X_j \mid X_1 \wedge \cdots X_{j-1} \,\right]$$

Joint probability

Conditional Independence

Recollect: **Independence**

A and B are independent if $\Pr[A \wedge B] = \Pr[A] \cdot Pr[B]$.
Equivalent to $\Pr[\, A \mid B \,] = \Pr[A]$.

"Knowing B adds no information about A"

Conditional Independence

Suppose that we test for pneumonia using two tests

- Blood Test: B
- Throat Swab: T

- Are they fully independent?
- BUT: B, T independent given knowledge of **underlying cause** $S = \text{sick}$!

$$\Pr[B \wedge T \mid S] = \Pr[B \mid S] \Pr[T \mid S]$$

"Tests were conducted independently, only related by the underlying sickness"

Conditional Independence

- Write out full joint distribution using chain rule:

$$\Pr[\overbrace{\boxed{T_1 \wedge T_2 \wedge \cdots \wedge T_n}}^{\text{Effects}} \wedge \overbrace{S}^{\text{Cause}}]$$
$$= \Pr[\, T_1 \mid S \,] \Pr[\, T_2 \mid S \,] \cdots \Pr[\, T_n \mid S \,] \Pr[S]$$

- Joint distribution of n Boolean RVs: $2^n - 1$ entries.
- Conditional independence: linear!
- Conditional independence is more robust and common than absolute independence

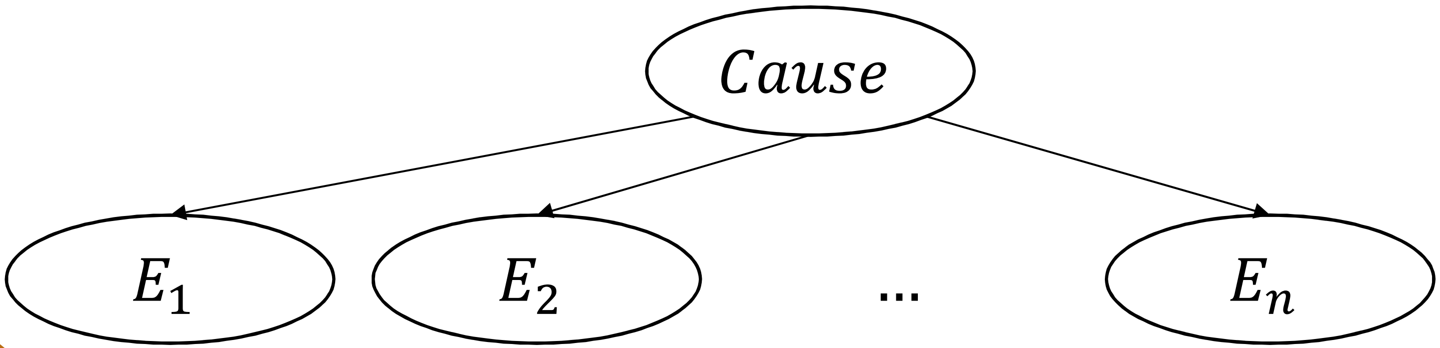
Bayes' Rule and Conditional Independence

A cause (heavy rain) can have several conditionally independent effects (Alice takes umbrella, Bob takes umbrella, Claire takes umbrella…)

$$\Pr[\,Cause \mid E_1 \ldots E_n\,] = \frac{\Pr[Cause]\Pr[\,E_1, \ldots, E_n \mid Cause\,]}{\Pr[E_1, \ldots, E_n]}$$

$$= \alpha \Pr[Cause] \prod_i \Pr[\,E_i \mid Cause\,]$$

- This is an example of a naive Bayes model:



Normalization

- We are trying to diagnose the disease X. 70% of the population is healthy, 20% are carriers, and 10% are sick.
- A blood test will come back positive with the following probability:
  - $\Pr[T = 1 \mid X = healthy] = 0.1$
  - $\Pr[T = 1 \mid X = carrier] = 0.7$
  - $\Pr[T = 1 \mid X = sick] = 0.9$
- We run a test three times (independently) and obtain two positive (on tests 1 and 2) and one negative (on test 3). What is the likeliest value for X?

Normalization

$$\Pr[\,X \mid T_1 = T_2 = 1, T_3 = 0\,] = \frac{\Pr[X]\Pr[T_1 = T_2 = 1, T_3 = 0 \mid X]}{\Pr[T_1 = T_2 = 1, T_3 = 0]}$$

Don't care about $\frac{1}{\Pr[T_1=T_2=1,T_3=0]}$! Set it to α.

$$\alpha \Pr[X] \times \Pr[\,T_1 = 1, T_2 = 1, T_3 = 0 \mid X\,]$$
$$= \alpha \Pr[X] \times \Pr[\,T_1 = 1 \mid X\,] \times \Pr[\,T_2 = 1 \mid X\,] \times \Pr[\,T_3 = 0 \mid X\,]$$

$$\Pr[\,X = healthy \mid A\,] = \alpha \times 0.7 \times 0.1 \times 0.1 \times 0.9 = 0.0063\alpha$$
$$\Pr[\,X = carrier \mid A\,] = \alpha \times 0.2 \times 0.7 \times 0.7 \times 0.3 = 0.0294\alpha$$
$$\Pr[\,X = sick \mid A\,] = \alpha \times 0.1 \times 0.9 \times 0.9 \times 0.1 = 0.0081\alpha$$
$$\mathbf{Pr}[.] = \langle 0.1438, 0.6712, 0.1849 \rangle$$

BAYESIAN NETWORKS

AIMA Chapter 14.1 – 14.2

Bayesian Networks

- A graphical way of writing joint distributions
- Nodes are random variables
- Edge from X to Y: X directly influences Y
- a conditional distribution for each node given its parents:
$$\Pr[\, X \mid Parents(X) \,]$$
- In the simplest case, conditional distribution can be represented as a conditional probability table (CPT): the distribution over X for each combination of parent values

Bayesian Networks

Given $X_1, \ldots, X_n$, write

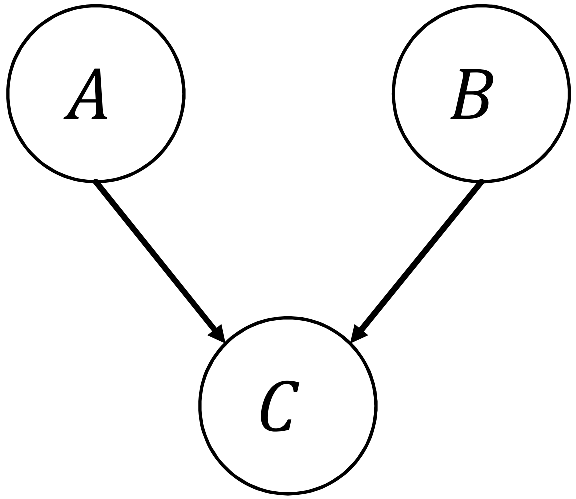
$$\Pr[X_1 \wedge \cdots \wedge X_n] = \prod_i \Pr[\, X_i \mid Parents(X_i) \,]$$

Examples

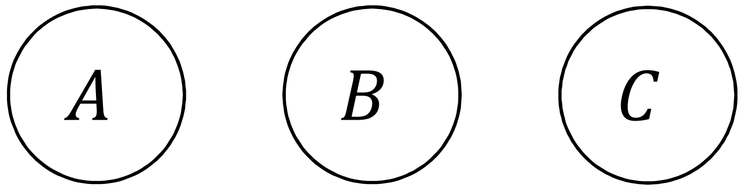
- $\Pr[A \wedge B \wedge C] = \Pr[\, C \mid A, B \,] \Pr[A] \Pr[B]$

Independent causes:
"I can be late either because of rain or because I was sick"

(in logic: $A \vee B \rightarrow C$)



- $\Pr[A \wedge B \wedge C] = \Pr[C] \Pr[A] \Pr[B]$

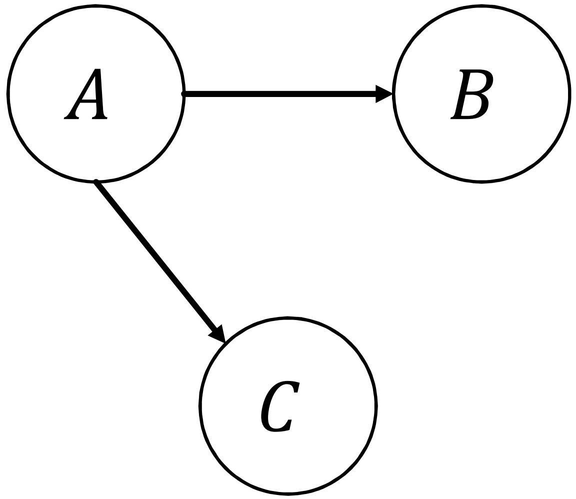


Examples

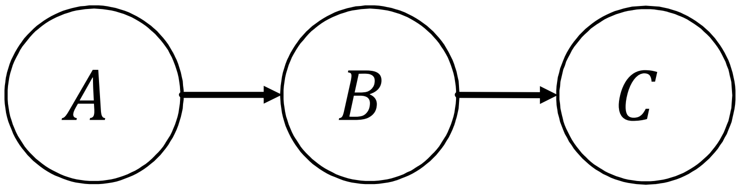
- $\Pr[A \wedge B \wedge C] = \Pr[\,C \mid A\,]\Pr[B \mid A]\Pr[A]$

Conditionally independent effects: "A disease can cause two independent tests to be positive"

(in logic: $A \to B; A \to C$)

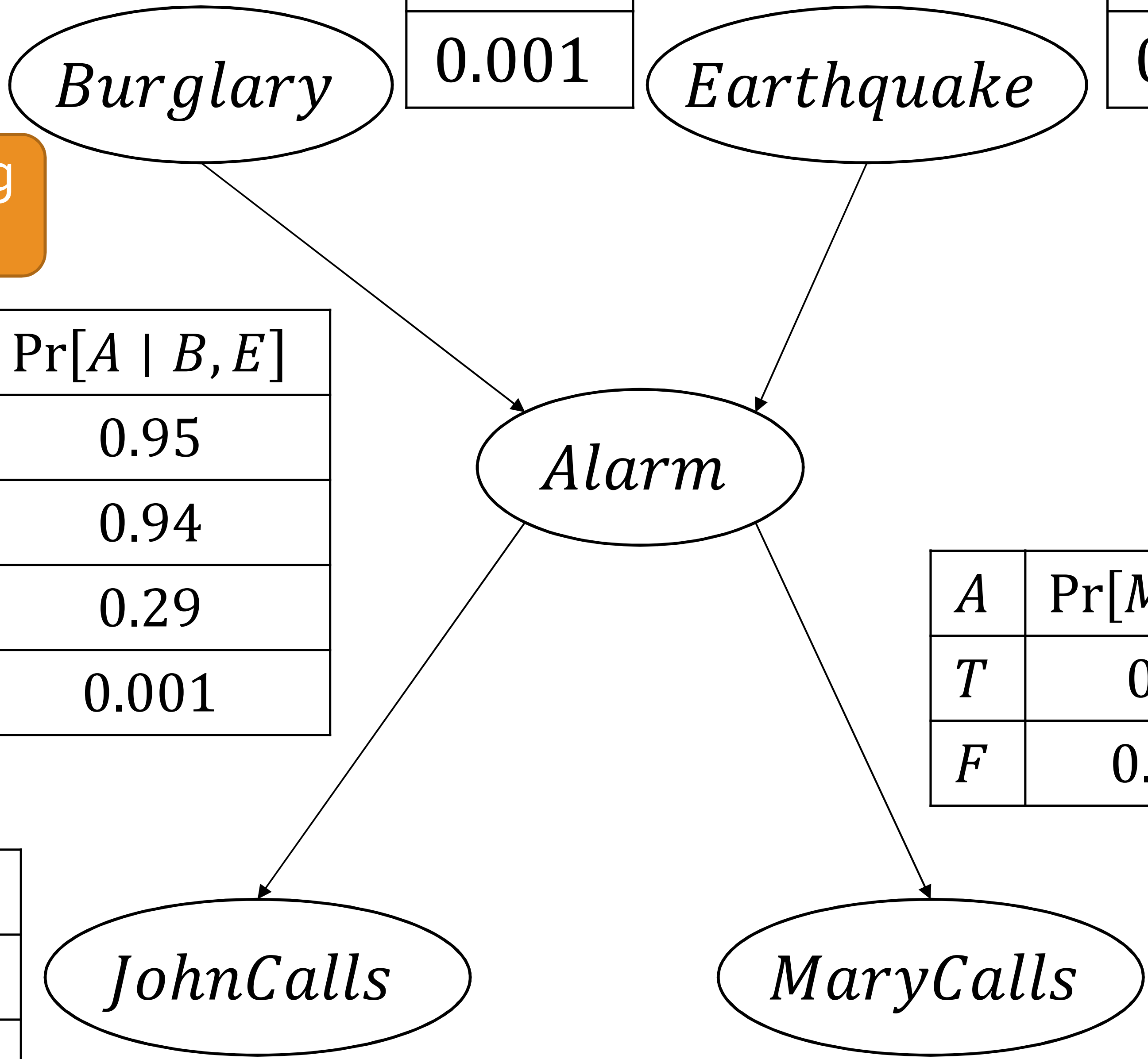


- $\Pr[A \wedge B \wedge C] = \Pr[C \mid B]\Pr[B \mid A]\Pr[A]$



Example With More Variables

- I'm at work
  - neighbor John calls to say my house alarm is ringing
  - neighbor Mary doesn't call
  - Alarm sometimes set off by minor earthquake.
  - Is there a burglar?
- Variables: B, E, A, J, M
- 5 binary variables: joint distribution table size $2^5 - 1$
- Exploit domain knowledge $\rightarrow$ smaller representation.



$\Pr[B]$
0.001

$\Pr[E]$
0.002

Conditioning case

B	E	$\Pr[A \mid B, E]$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	$\Pr[M \mid A]$
T	0.7
F	0.01

A	$\Pr[J \mid A]$
T	0.9
F	0.05

Assignment 3: How to verify?

- Use Genie modeller: https://download.bayesfusion.com/files.html?category=Academia#GeNIe
- Setup your network and assign probability values.
- Set evidence and run updates
- You can find the probability values after running the inference

Bayesian Networks – Compactly Representing Joint Distributions

- Conditional probability table for Boolean X with k Boolean parents has 2^k rows: **all** possible parent values
- Each row requires one number p for $X = True$
- If each variable has $\leq k$ parents, network representation requires $\mathcal{O}(n2^k)$ values, vs. $\mathcal{O}(2^n)$ for full joint distribution.
- For burglary network, $1 + 1 + 2 + 2 + 4 = 10$ numbers as compared to $2^5 - 1 = 31$ numbers for full joint distribution

Inference in Bayesian Networks

A Bayesian Network represents the full joint distribution; can infer any query.

$$\Pr[\,B = 1 \mid J = 1, M = 0\,] = \frac{\Pr[B = 1, J = 1, M = 0]}{\Pr[J = 1, M = 0]} = ?$$

$$\Pr[J, M, A, B, E] = \Pr[\,J \mid A\,]\Pr[\,M \mid A\,]\Pr[\,A \mid B, E\,]\Pr[B]\Pr[E]$$

e.g.

$$\Pr[B = 1, J = 1, M = 0, A = 1, E = 0]$$
$$= \Pr[\,j \mid a\,]\Pr[\neg m \mid a]\Pr[\,a \mid b, \neg e\,]\Pr[b]\Pr[\neg e]$$
$$= 0.9 \times 0.3 \times 0.94 \times 0.001 \times 0.998 \simeq 0.000253$$

Need to compute the cases $A = 0, E = 0$; $A = 1, E = 1$; $A = 0, E = 1$.

Constructing Bayesian Networks

1. Choose an ordering of variables $X_1, \ldots, X_n$
2. For $i = 1, \ldots, n$:
   - Add node X_i to the network
   - Select minimal set of parents from $X_1, \ldots, X_{i-1}$ such that $\Pr[\, X_i \mid Parents(X_i)\,] = \Pr[X_i \mid X_1, \ldots, X_{i-1}]$
   - Link every parent to X_i
   - Write down CPT for $\Pr[\, X_i \mid Parents(X_i)\,]$

Where does this come from?

Constructing Bayesian Networks

Consequence of chain rule, generally true!

This construction guarantees

$$\Pr[X_1, \dots, X_n] = \prod_i \Pr[X_i \mid X_1, \dots, X_{i-1}]$$

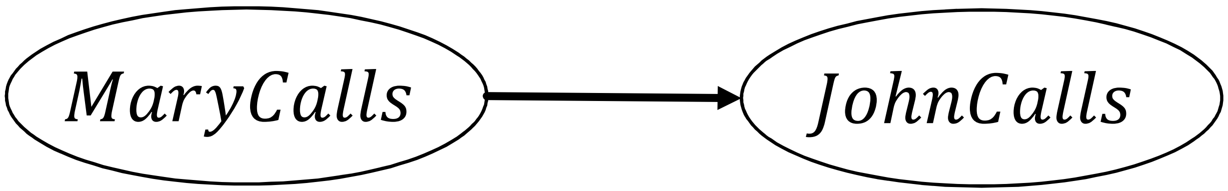
$$= \prod_i \Pr[\, X_i \mid Parents(X_i)\,]$$

By choice of parents

Network is acyclic (why??), and has no redundancies

Variable Order Matters

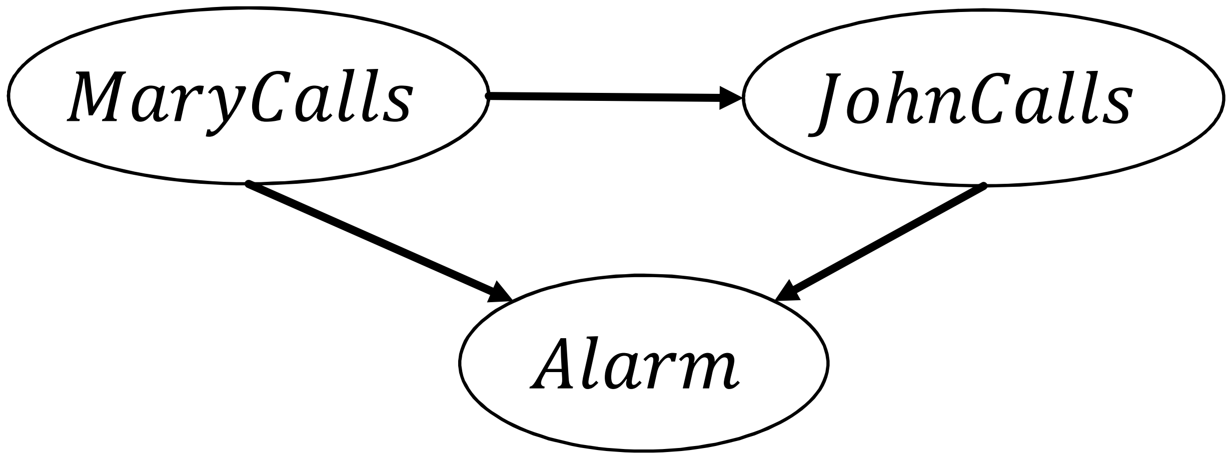
We choose the ordering M, J, A, B, E
(originally was B, E, A, M, J)



Is it true that $\Pr[J \mid M] = \Pr[J]$?

Variable Order Matters

We choose the ordering M, J, A, B, E

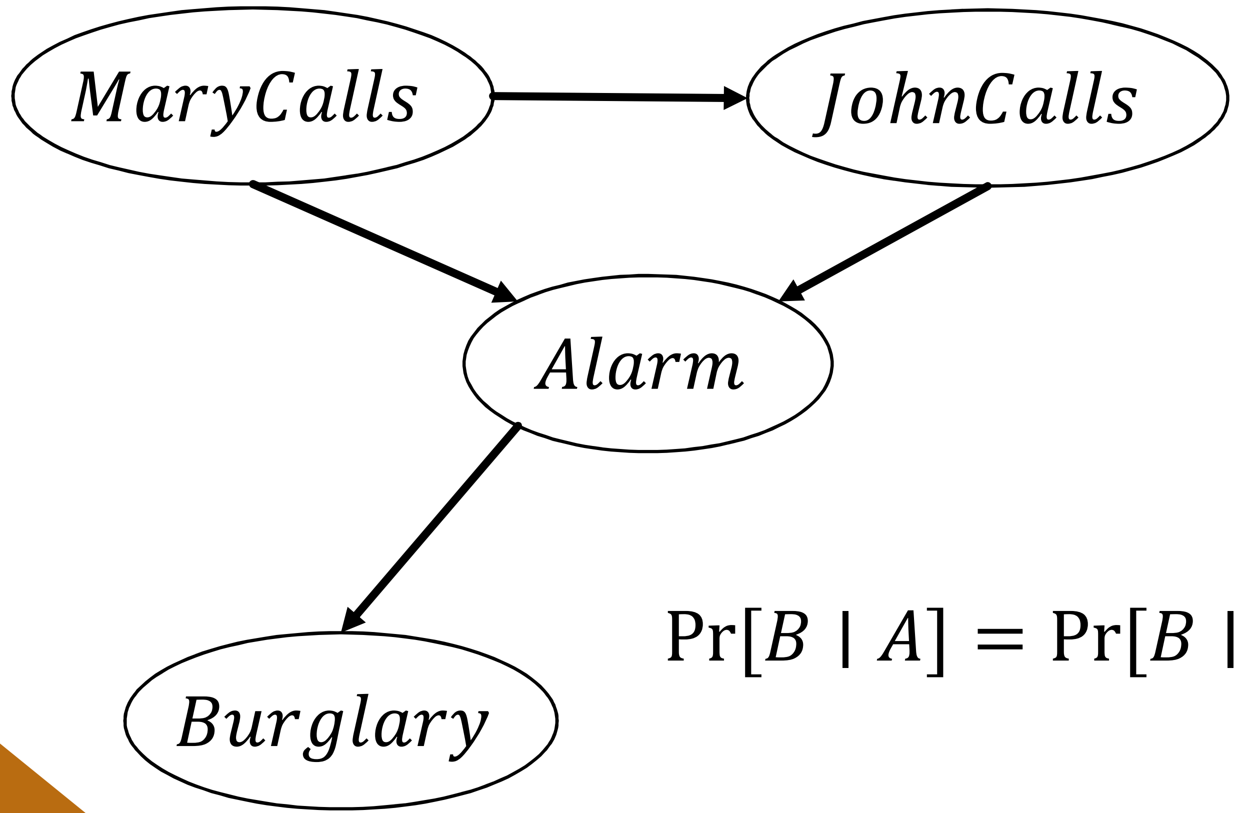


Is it true that

$$\Pr[A \mid M, J] = \Pr[A]$$
$$\Pr[A \mid M, J] = \Pr[A \mid J]$$
$$\Pr[A \mid M, J] = \Pr[A \mid M]$$

Variable Order Matters

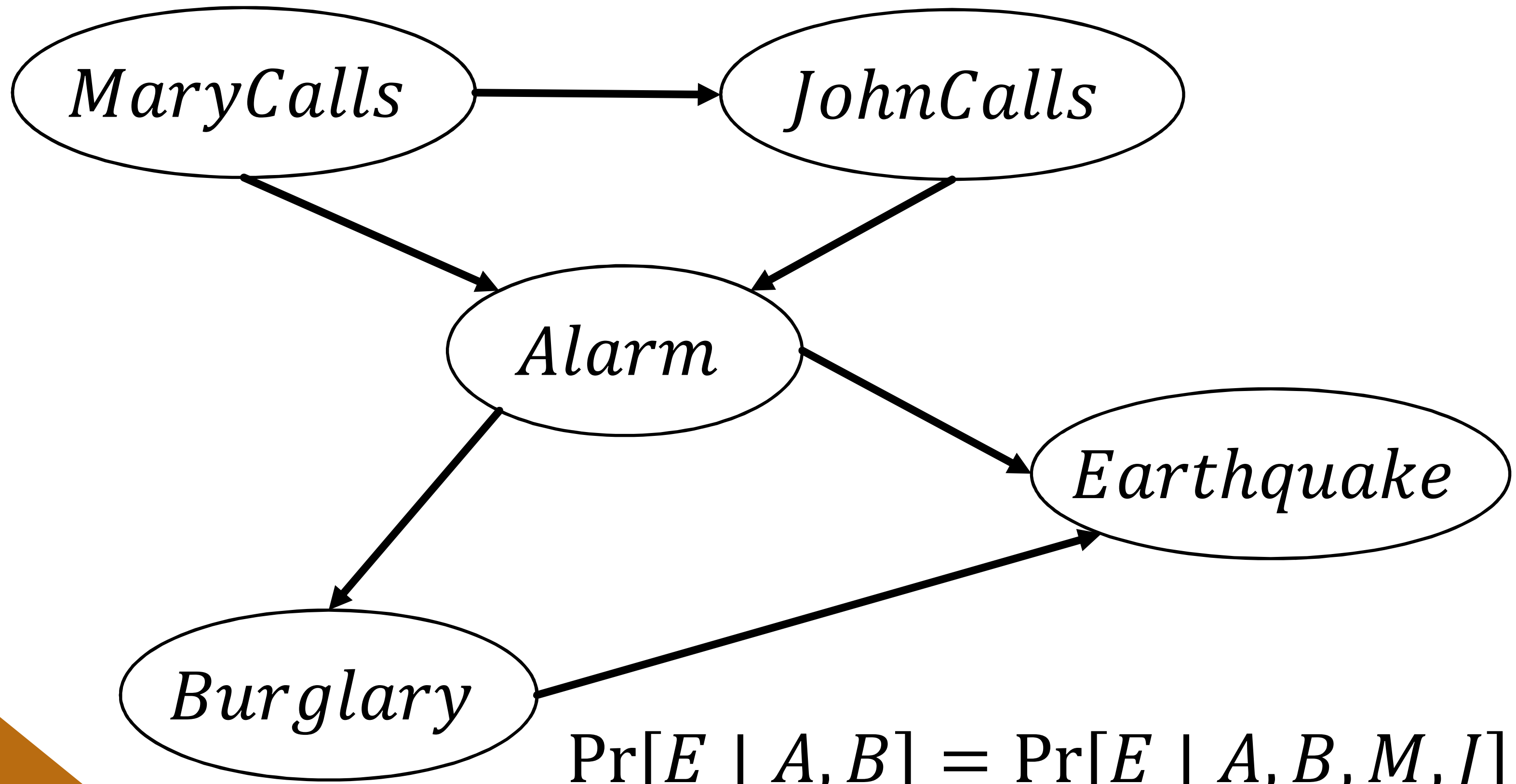
We choose the ordering M, J, A, B, E



$\Pr[B \mid A] = \Pr[B \mid A, M, J]$?

Variable Order Matters

We choose the ordering M, J, A, B, E

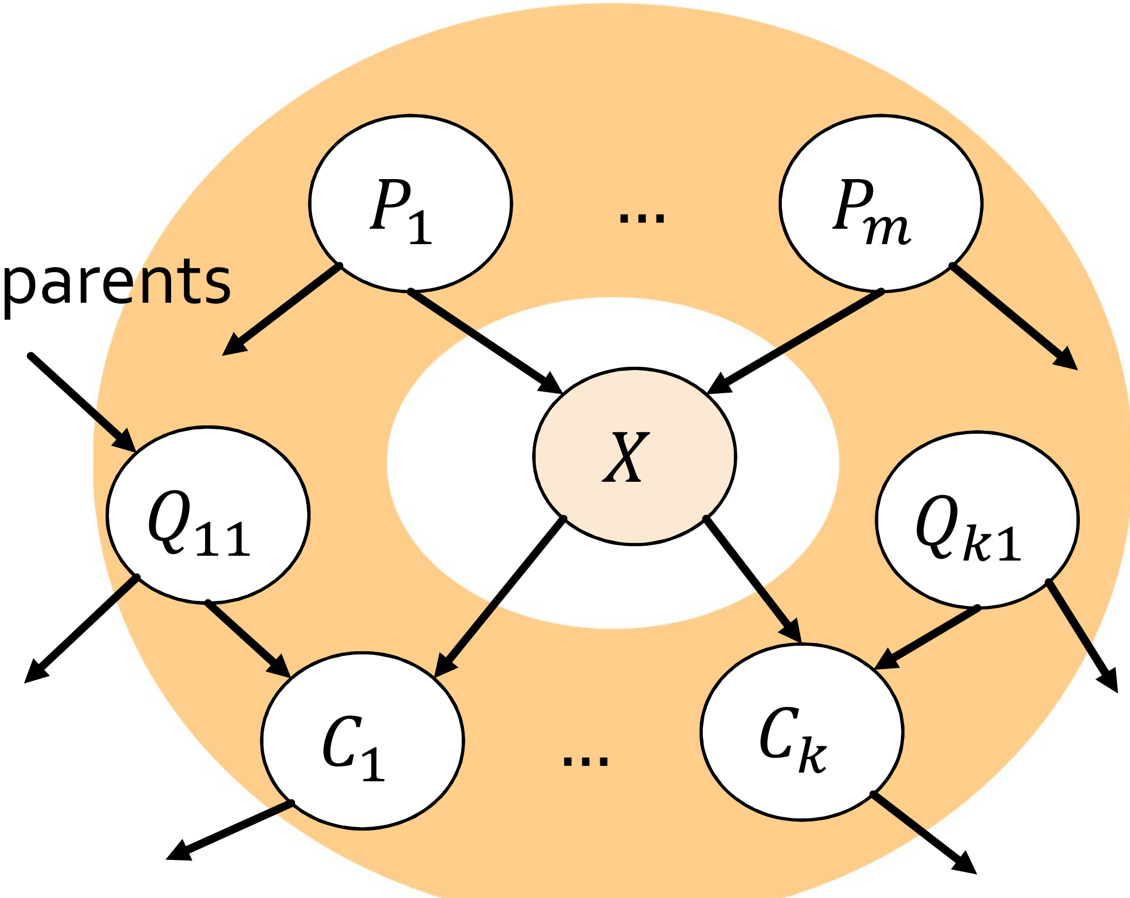


$\Pr[E \mid A, B] = \Pr[E \mid A, B, M, J]$?

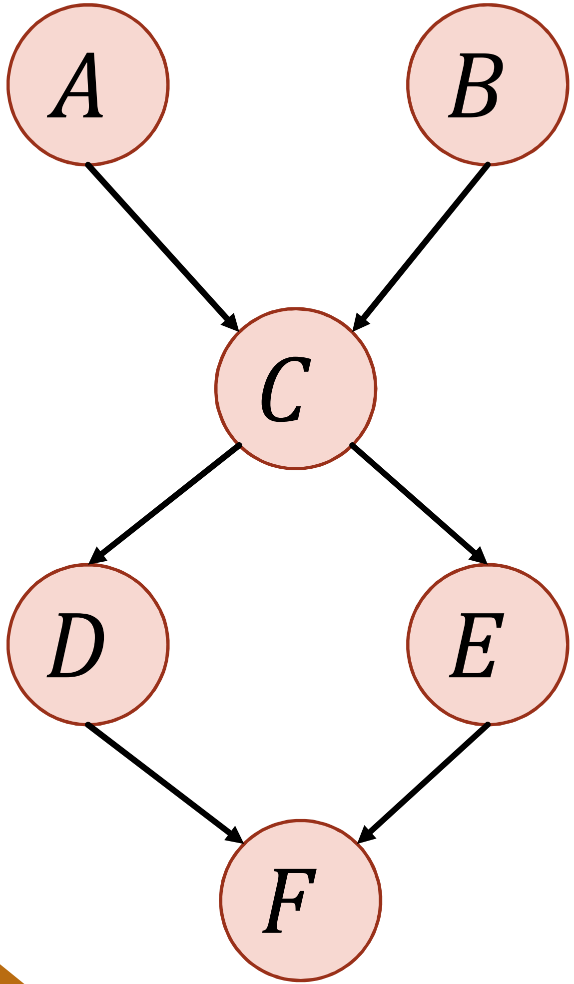
The Markov Blanket

A node is conditionally independent of everything else **given the values** of its:

- parents
- children
- its children's parents



Conditional Independence in BN



Given variables X, Y and **known variables**

$$\mathcal{E} = \{E_1, \ldots, E_k\}$$

are X and Y independent given knowledge of $\mathcal{E}$?

Conditional Independence in BN

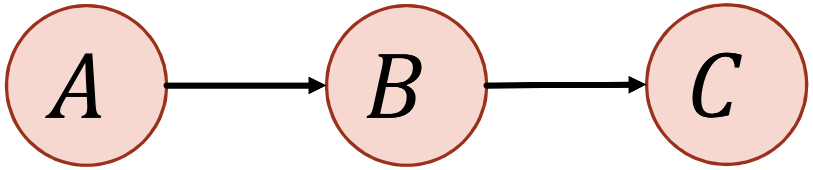
Given variables X, Y and **known variables**
$$\mathcal{E} = \{E_1, \dots, E_k\}$$
are X and Y independent given knowledge of $\mathcal{E}$?

Can be shown

- using algebra (annoying and tedious):
$$\Pr[\, X \mid \mathcal{E} \,] = \cdots = \Pr[\, X \mid \mathcal{E}, Y \,]$$
- via counterexample (computing via the CPTs)

Can we show that two nodes are **necessarily independent?**

Causal Chains



"Rain (A) causes traffic (B) which causes me to be late (C)"

Question: are A and C **necessarily independent**?

Question: are A and C **conditionally independent**, given B?

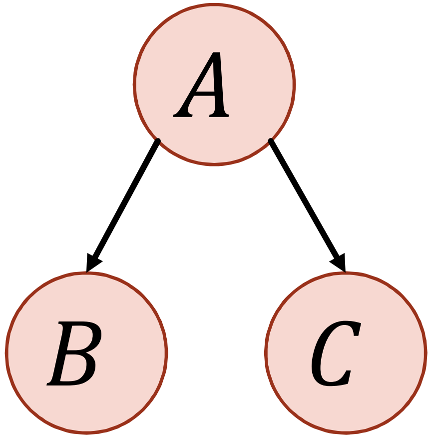
Causal Chains

$$A \rightarrow B \rightarrow C$$

$$\Pr[\,C \mid A, B\,] = \frac{\Pr[A \wedge B \wedge C]}{\Pr[A \wedge B]} = \frac{\Pr[A]\Pr[\,B \mid A\,]\Pr[\,C \mid B\,]}{\Pr[A]\Pr[\,B \mid A\,]} = \Pr[\,C \mid B\,]$$

$\Pr[\,C \mid A, B\,] = \Pr[\,C \mid B\,]$: given B, knowing A does not update my beliefs on C!

Common Cause

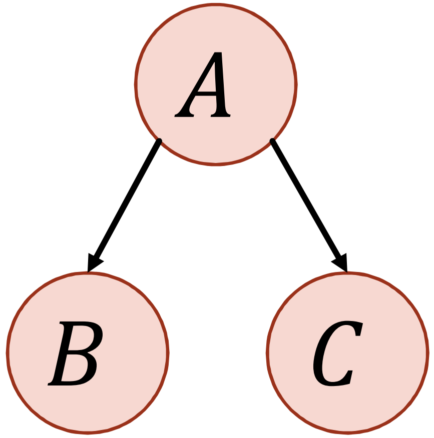


"Batman (A) catches the Joker (B) and Bane (C)"

Question: are B and C **necessarily independent**?

Question: are B and C **conditionally independent**, given A?

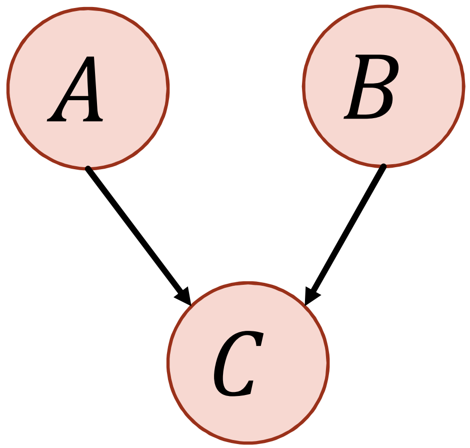
Common Cause



$$\Pr[\, B \mid A, C \,] = \frac{\Pr[A \wedge B \wedge C]}{\Pr[A \wedge C]} = \frac{\Pr[A] \Pr[\, B \mid A \,] \Pr[\, C \mid A \,]}{\Pr[A] \Pr[\, C \mid A \,]} = \Pr[\, B \mid A \,]$$

$\Pr[\, B \mid A, C \,] = \Pr[\, B \mid A \,]$: given A, knowing C does not update my beliefs on B!

Common Effect

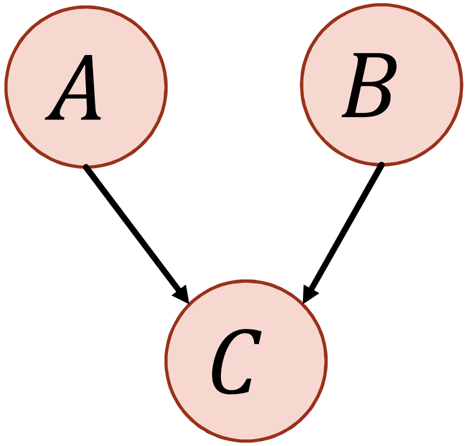


"The Joker (A) and Bane (B) could both rob the bank (C)"

Question: are A and B **necessarily independent**?

Question: are A and B **conditionally independent**, given C?

Common Effect



Observing an effect makes two causes dependent

- I know that the bank was robbed ($C = 1$)
- It could be either the Joker or Bane.
- If I know the Joker didn't do it –my belief about Bane doing it is higher!

$$\Pr[\, A \mid C, B \,] \neq \Pr[\, A \mid C \,]$$

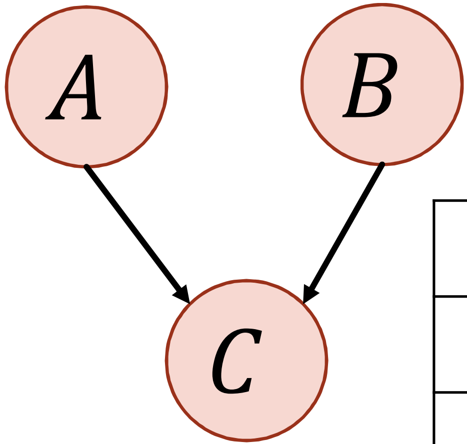
but

$$\Pr[\, A \mid B \,] = \Pr[A]$$

It’s All About the CPTs

$$\Pr[A] = 0.5$$

$$\Pr[B] = 0.5$$



$A =$	$B =$	$\Pr[\, C \mid A, B \,] =$
1	1	1
1	0	1
0	1	1
0	0	0

$$\Pr[A = 1] = \Pr[\, A = 1 \mid B = 0 \,] = 0.5$$

but

$$\Pr[\, A = 1 \mid B = 0, C = 1 \,] = 1$$

General Case – d Separation

Given variables X, Y and **known variables**
$\mathcal{E} = \{E_1, \dots, E_k\}$
are X and Y **surely** independent given $\mathcal{E}$?

Idea: any general graph can be broken down into the three cases described above, to determine conditional independence of X, Y given knowledge of $\mathcal{E}$.

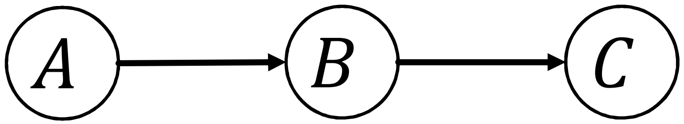
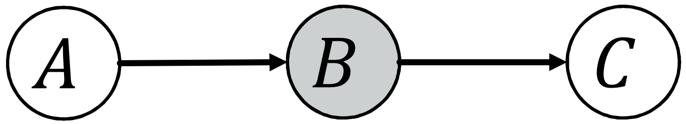
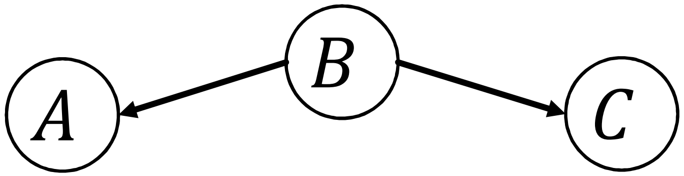
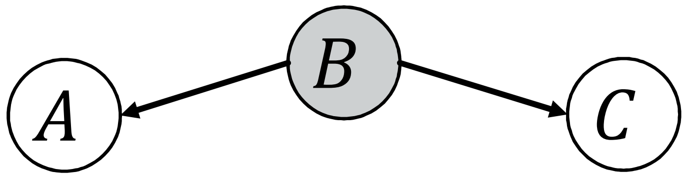
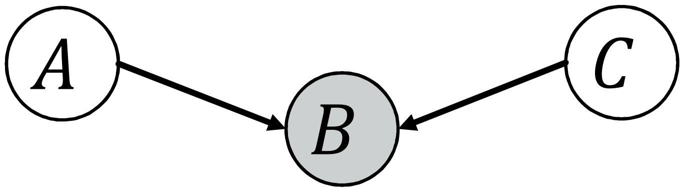
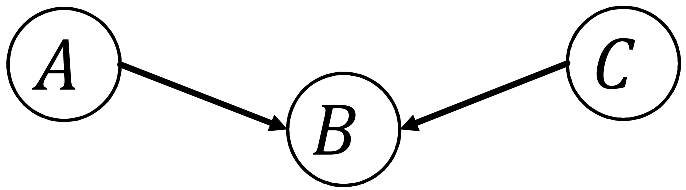
General Case – d Separation

Given variables X, Y and **known variables**
$$\mathcal{E} = \{E_1, \ldots, E_k\}$$
are X and Y **surely** independent given $\mathcal{E}$?

- Check every **undirected** path between X and Y (ignore direction of arcs).
- If all paths are not **active** then X and Y are independent given $\mathcal{E}$.

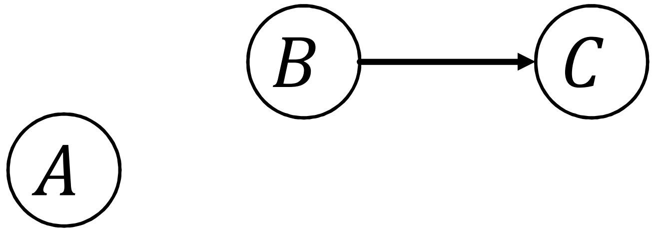
General Case – d Separation

- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!

Active	Inactive
	
	
	

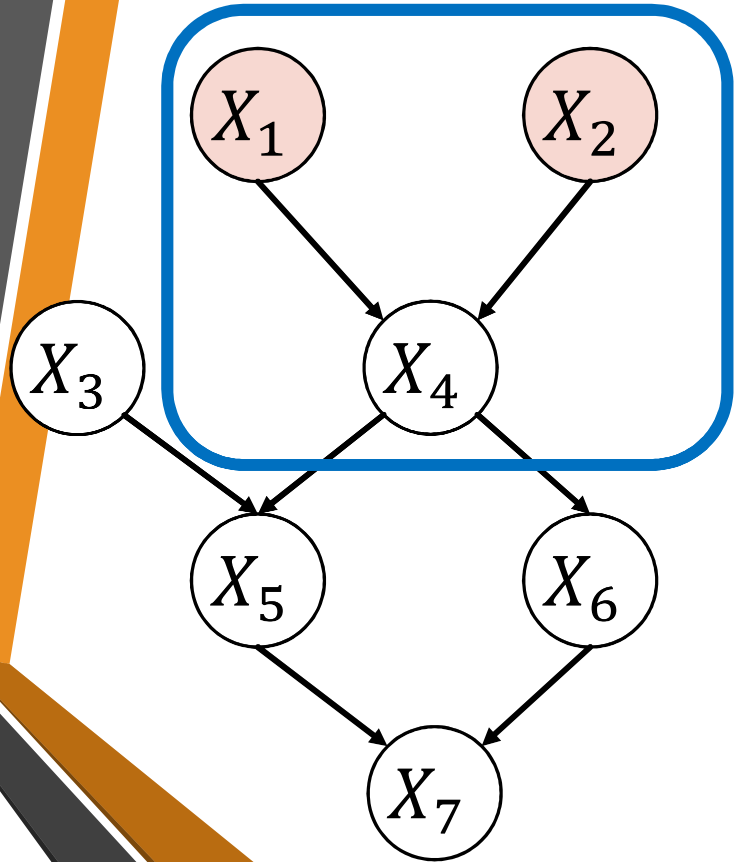
General Case – **Analyze the Graph**

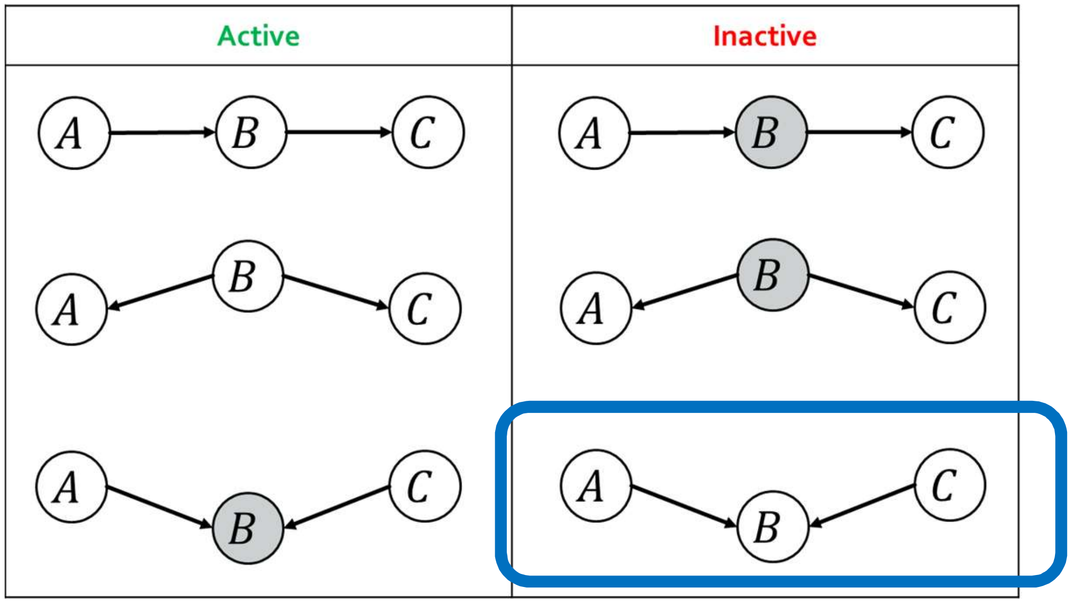
- Degenerate cases:
  - Disconnected variables: always independent.
  - Directly connected variables: never **surely** independent.



General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!



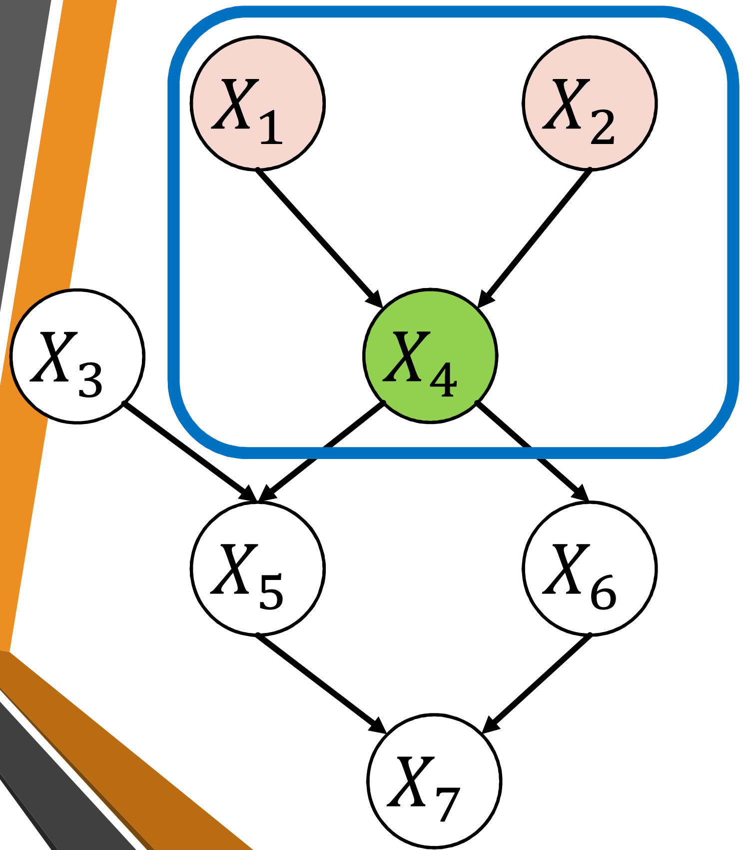


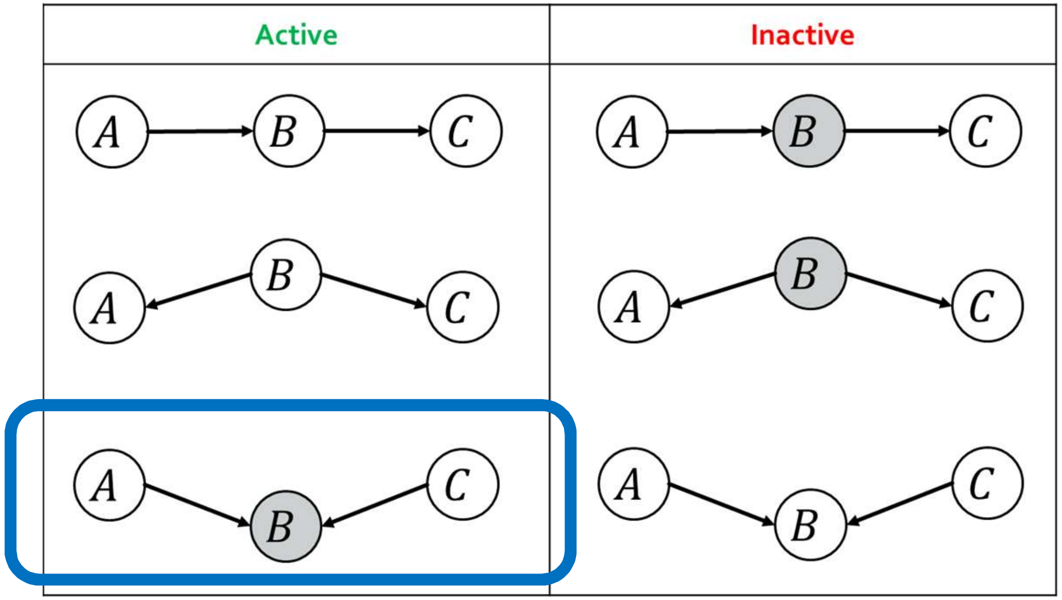
Is X_1 independent of X_2?

Yes!

General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!



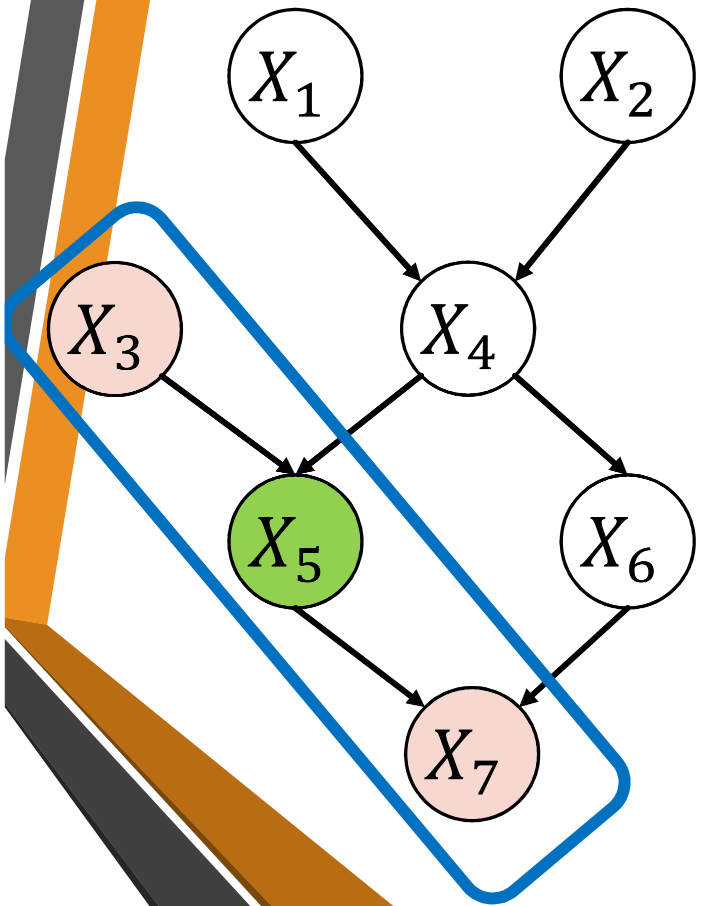


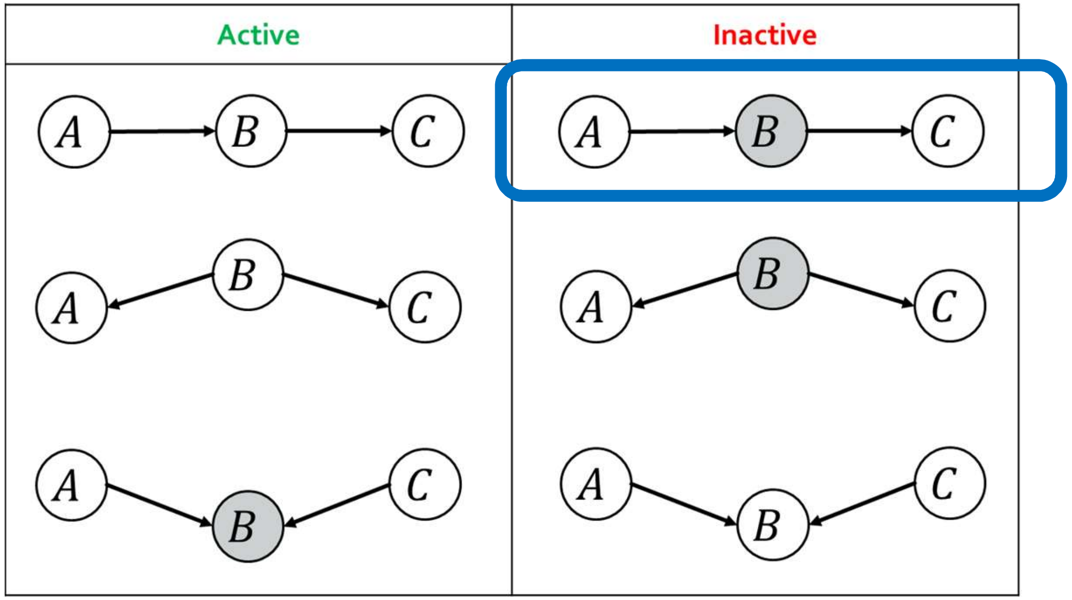
Is X_1 independent of X_2 given X_4?

No!

General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!

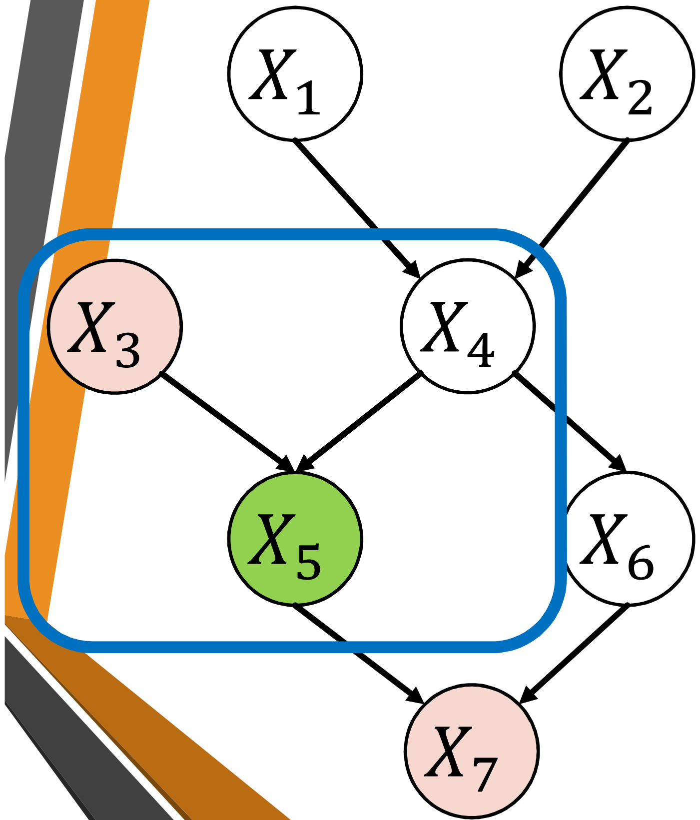


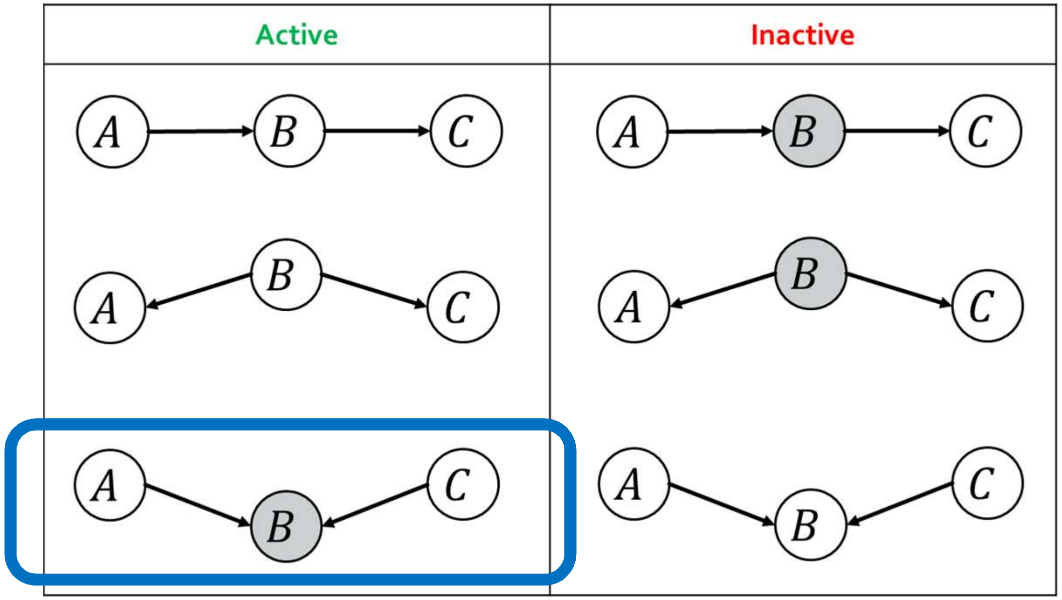


Is X_3 independent of X_7 given X_5?

General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!

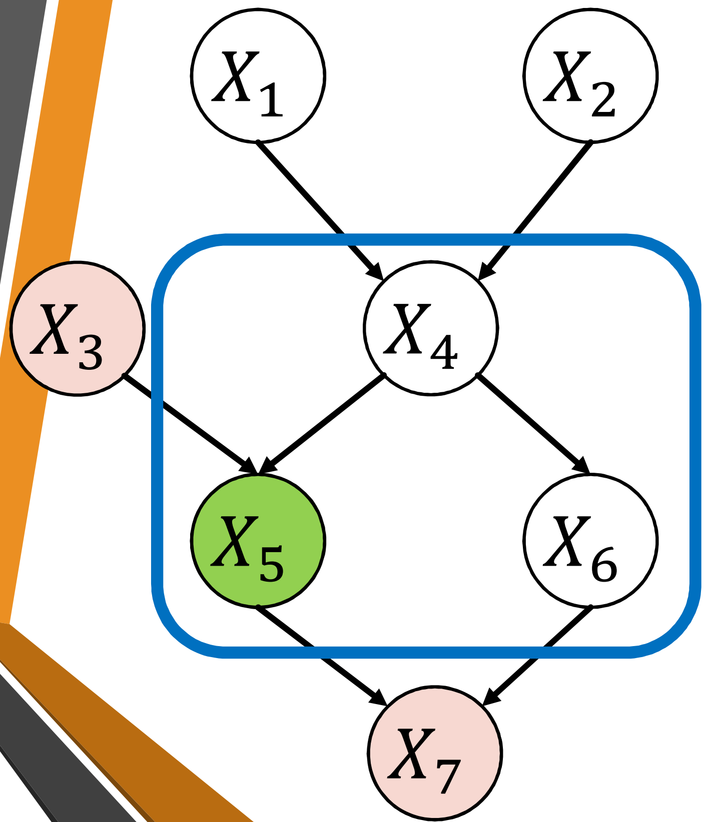


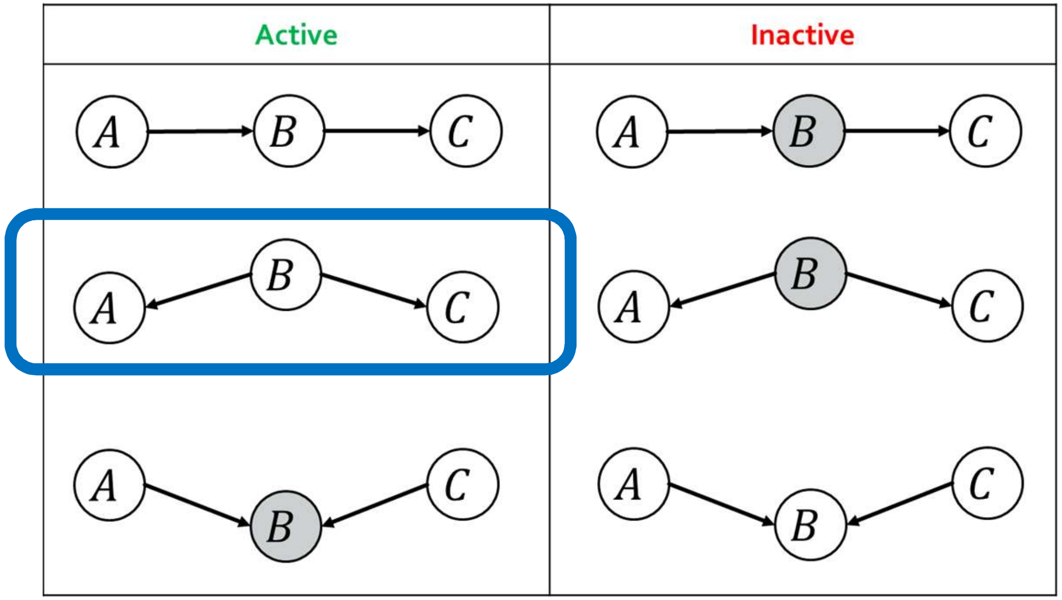


Is X_3 independent of X_7 given X_5?

General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!

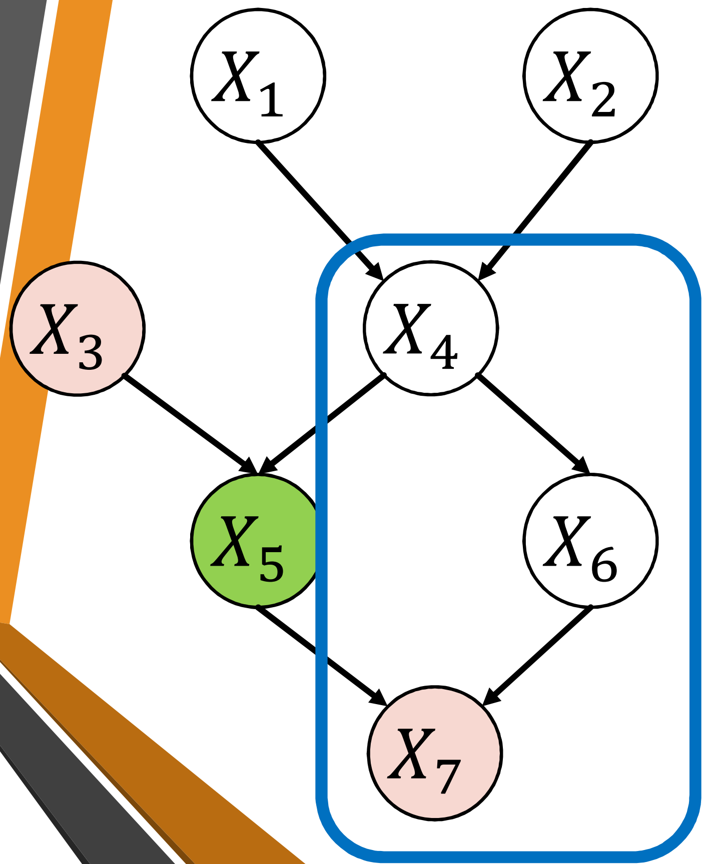


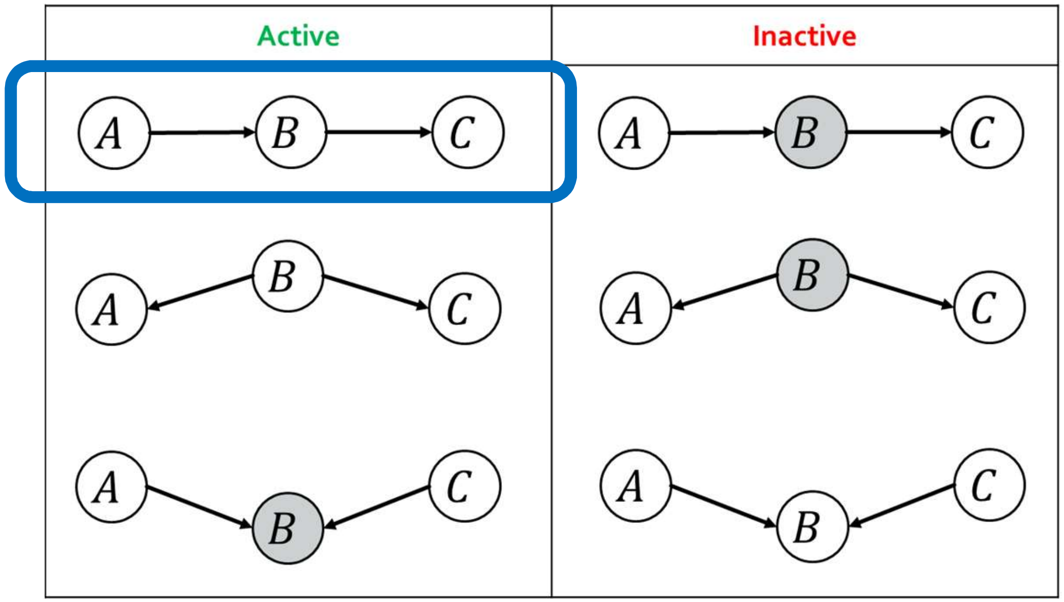


Is X_3 independent of X_7 given X_5?

General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!



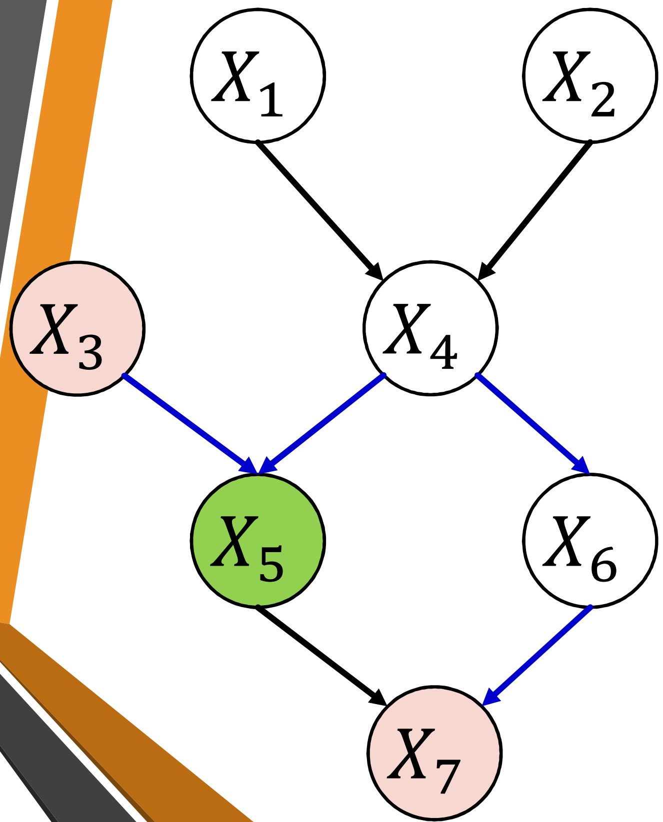


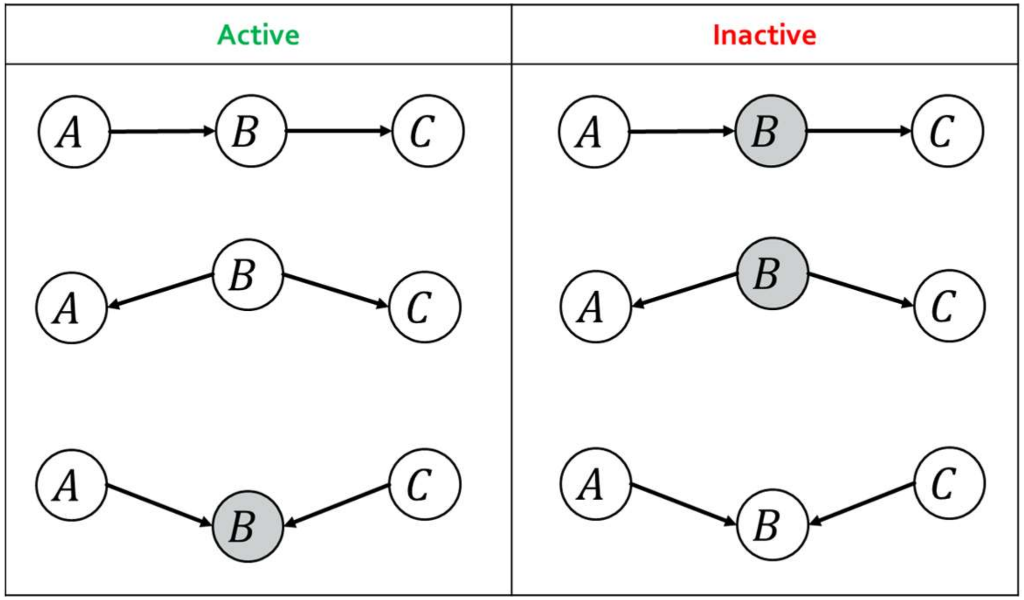
Is X_3 independent of X_7 given X_5?

No!

General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!





Is X_3 independent of X_7 given X_5?

No!

X_3, X_5, X_4, X_6, X_7 form an **active path**