Introduction to Bayesian Networks

Alice Gao

Lecture 11b

Readings: RN 13.4. PM 8.3.

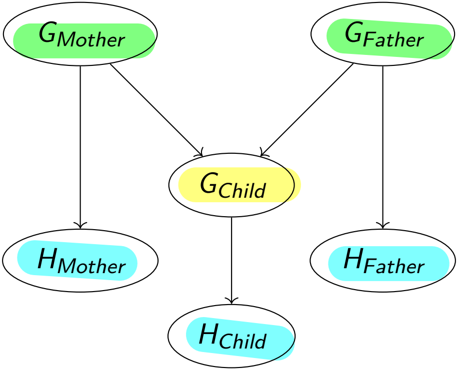
Outline

Learning Goals

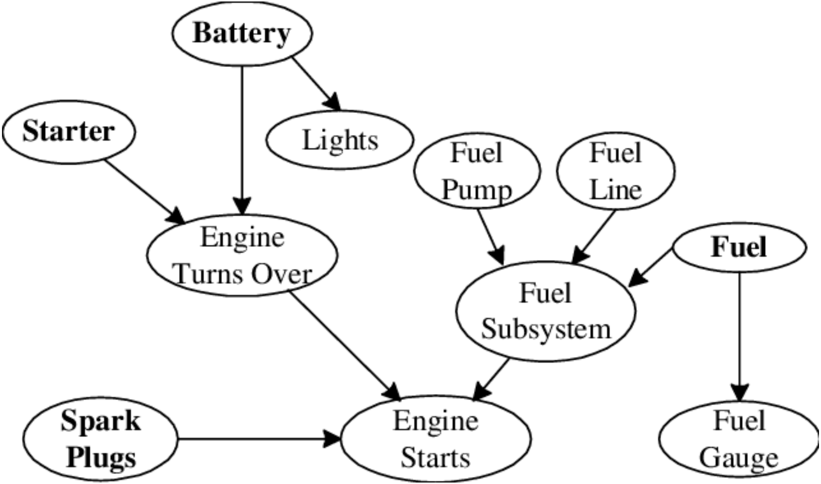
By the end of the lecture, you should be able to

- Describe components of a Bayesian network.
- Compute a joint probability given a Bayesian network.
- Given a Bayesian network, determine if two variables are independent or conditionally independent given a third variable.
- Given a joint probability distribution and an order of the variables, construct a Bayesian network that correctly represents the independent relationships among the variables in the distribution.

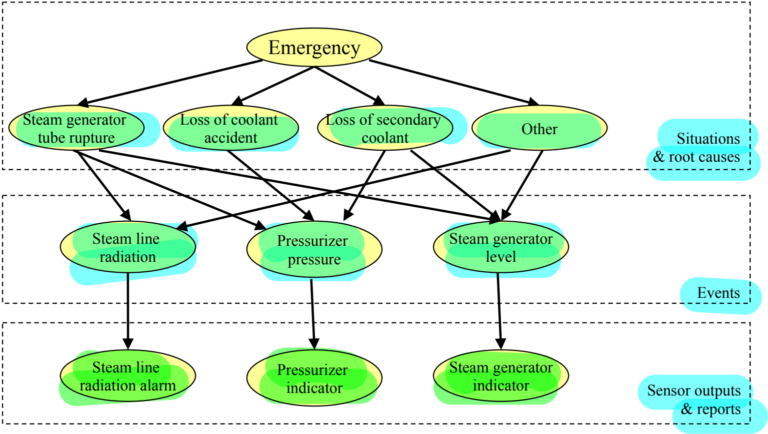
Inheritance of Handedness



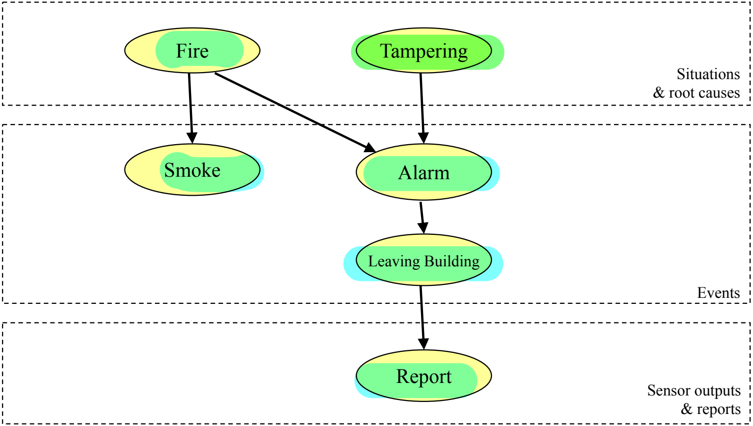
Car Diagnostic Network



Example: Nuclear power plant operations

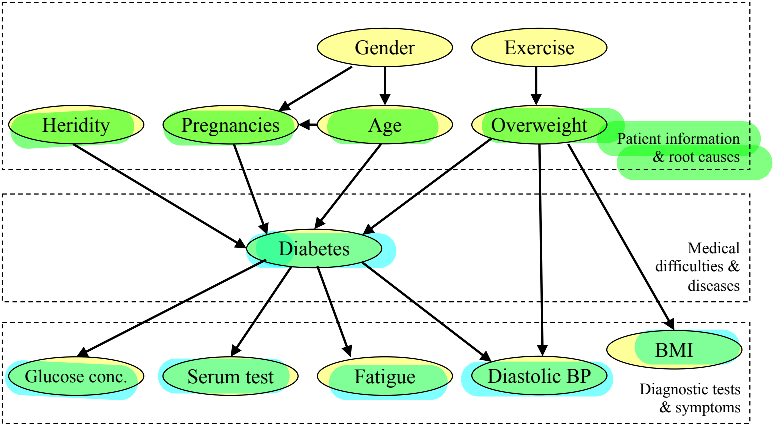


Example: Fire alarms



Report: “report of people leaving building because a fire alarm went off”

Example: Medical diagnosis of diabetes



Why Bayesian Networks?

A probabilistic model of the Holmes scenario:

- The random variables:
  Earthquake, Radio, Burglary, Alarm, Watson, and Gibbon.
- # of probabilities in the joint distribution: $2^6 = 64$.
- For example,
  $P(E \wedge R \wedge B \wedge A \wedge W \wedge G) = ?$
  $P(E \wedge R \wedge B \wedge A \wedge W \wedge \neg G) = ?$
  ... etc ...

We can compute any probability using the joint distribution, but

- Quickly become intractable as the number of variables grows.
- Unnatural and tedious to specify all the probabilities.

Why Bayesian Networks?

A Bayesian Network

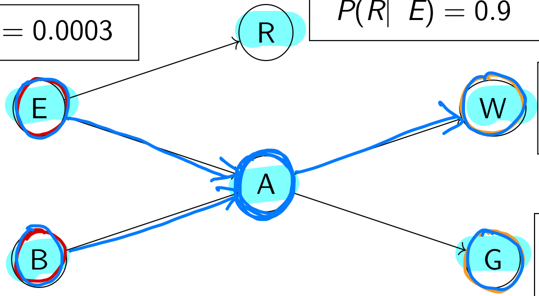
- is a compact version of the joint distribution, and
- takes advantage of the unconditional and conditional independence among the variables.

A Bayesian Network for the Holmes Scenario

$2^6 - 1 = 63 \quad 1+1+2+4+2+2 = 12$

$P(E) = 0.0003$

$P(R|\neg E) = 0.0002$
$P(R|\ E) = 0.9$



$P(W|\neg A) = 0.4$
$P(W|\ A) = 0.8$

$P(G|\neg A) = 0.04$
$P(G|\ A) = 0.4$

$\boxed{P(B) = 0.0001}$

$P(A|\neg B \wedge \neg E) = 0.01$
$P(A|\neg B \wedge\ E) = 0.2$
$P(A|\ B \wedge \neg E) = 0.95$
$P(A|\ B \wedge\ E) = 0.96$

Bayesian Network

A Bayesian Network is a directed acyclic graph.

- Each node corresponds to a random variable.
- X is a parent of Y if there is an arrow from node X to node Y.
- 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 Semantics of Bayesian Networks

Two ways to understand Bayesian Networks:

- A representation of the joint probability distribution
- An encoding of the conditional independence assumptions

Representing the joint distribution

We can compute each joint probability using the following formula.

$$P(X_n \wedge \cdots \wedge X_1) = \prod_{i=1}^{n} P(X_i | Parents(X_i))$$

Representing the joint distribution

Example: What is the probability that

- The alarm has sounded,
- Neither a burglary nor an earthquake has occurred,
- Both Watson and Gibbon call and say they hear the alarm, and
- There is no radio report of an earthquake?

$P(\neg B \wedge \neg E \wedge A \wedge \neg R \wedge W \wedge G)$ $\quad * P(G|A)$

$= P(\neg B) * P(\neg E) * P(A | \neg B \wedge \neg E) * P(\neg R | \neg E) * P(W|A)$

$= (1 - 0.0001)(1 - 0.0003)(0.01)(1 - 0.0002)(0.4)(0.8)$

$= 0.0032$

CQ: Calculating the joint probability

CQ: What is the probability that

- NEITHER a burglary NOR an earthquake has occurred,
- The alarm has NOT sounded,
- NEITHER of Watson and Gibbon is calling, and
- There is NO radio report of an earthquake?

(A) 0.5699

(B) 0.6699

(C) 0.7699

(D) 0.8699

(E) 0.9699

$(1-0.0001)*(1-0.0003)*(1-0.01)*(1-0.4)$

$*(1-0.04)*(1-0.0002) = 0.5699.$

Nothing "exciting" is happening in the world.

"No news is good news!"

Burglary, Alarm and Watson

CQ Unconditional Independence

CQ: Are Burglary and Watson independent?

Burglary → Alarm → Watson

(A) Yes

(B) No

If we learned B, would our belief of W change?

If Burglary is happening, then Alarm is likely going off, and Watson is likely calling Holmes.

CQ: Conditional Independence

CQ: Are Burglary and Watson conditionally independent given Alarm?

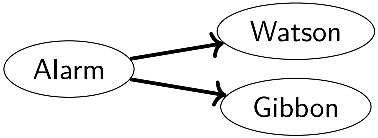
(A) Yes

(B) No

Watson does not observe Burglary directly. Burglary could only influence Watson through Alarm.

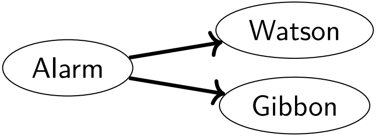
If Alarm is known, we have "cut the chain" in the middle. Burglary and Watson cannot affect each other anymore.

Alarm, Watson and Gibbon



CQ Unconditional Independence

CQ: Are Watson and Gibbon independent?



(A) Yes

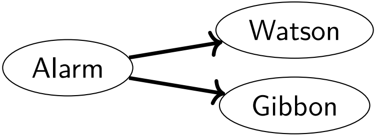
(B) No

If we learned the value of Watson, does this influence our belief about Gibbon?

If Watson is calling, then it's more likely that the Alarm is going off, which means that it is more likely that Gibbon is calling.

CQ Conditional Independence

CQ: Are Watson and Gibbon conditionally independent given Alarm?

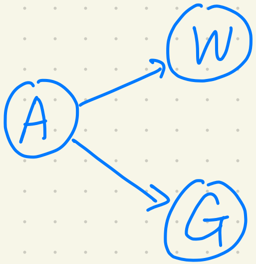


(A) Yes

(B) No

Alarm as an event.
Watson and Gibbon are noisy sensors for the event. The value of each sensor depends entirely on the event.

If we know whether the event is happening or not, the two sensors can no longer affect each other.



$P(A) = 0.1$

$P(W|A) = 0.8$ $\quad$ $P(G|A) = 0.4$

$P(W|\neg A) = 0.4$ $\quad$ $P(G|\neg A) = 0.1$

$$P(W|A \wedge G) = P(W|A \wedge \neg G) = P(W|A)$$

$$P(W|\neg A \wedge G) = P(W|\neg A \wedge \neg G) = P(W|\neg A)$$

$$P(W \wedge A \wedge G) = P(A)\,P(W|A)\,P(G|A) = 0.1 * 0.8 * 0.4 = 0.032$$

$$P(\neg W \wedge A \wedge G) = P(A)\,P(\neg W|A)\,P(G|A) = 0.1 * 0.2 * 0.4 = 0.008$$

$$P(W|A \wedge G) = \frac{0.032}{0.032 + 0.008} = 0.8 \qquad P(W|A) = \frac{0.08}{0.08 + 0.02} = 0.8$$

$$P(W \wedge A) = P(W \wedge A \wedge G) + P(W \wedge A \wedge \neg G) = 0.032 + 0.048 = 0.08$$

$$P(W \wedge A \wedge \neg G) = P(A)\,P(W|A)\,P(\neg G|A) = 0.1 * 0.8 * 0.6 = 0.048$$

$$P(\neg W \wedge A) = P(\neg W \wedge A \wedge G) + P(\neg W \wedge A \wedge \neg G) = 0.008 + 0.012 = 0.02$$

Three steps to follow to calculate any probability:

① To calculate a conditional probability, convert it into a joint/prior probability using the product rule.

$$P(B|A) = \frac{P(B \wedge A)}{P(A)}$$

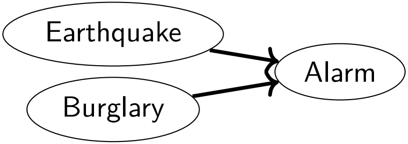
② To calculate a joint/prior probability not involving all the variables, introduce all the other variables by using the sum rule in reverse.

$$P(B \wedge A) = P(A \wedge B \wedge C) + P(A \wedge B \wedge \neg C)$$

③ Calculate every joint probability using the chain rule or using a Bayesian network.

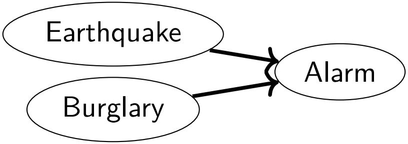
$$P(A \wedge B \wedge C) = P(A)\, P(B \mid A)\, P(C \mid A \wedge B)$$

Earthquake, Burglary, and Alarm



CQ Unconditional Independence

CQ: Are Earthquake and Burglary independent?



(A) Yes

(B) No

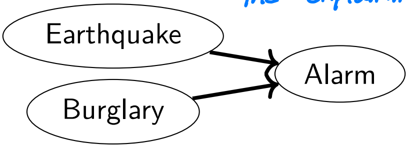
If we learned whether an earthquake is happening or not, this does NOT change our belief about Burglary, and vice versa.

Let's assume that looting is not more frequent during an earthquake. $\Downarrow$

CQ: Conditional Independence

CQ: Are Earthquake and Burglary conditionally independent given Alarm?

the "explaining away" effect



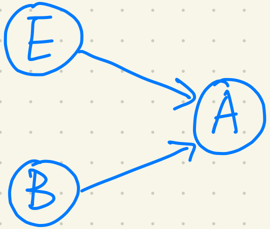
Suppose that the Alarm is going off.

If an earthquake is happening, then it is less likely that the Alarm is caused by a Burglary.

(A) Yes

(B) No

If an earthquake is not happening, then it is more likely that a Burglary is happening and causing the alarm to go off.



$P(E) = 0.1$ $\quad$ $P(A \mid E \wedge B) = 0.9$ $\quad$ $P(A \mid \neg E \wedge B) = 0.7$

$P(B) = 0.2$ $\quad$ $P(A \mid E \wedge \neg B) = 0.5$ $\quad$ $P(A \mid \neg E \wedge \neg B) = 0.1$

Prove that $P(E \mid A \wedge B) \neq P(E \mid A)$.

$$P(E \wedge A \wedge B) = \overset{P(E)\,\underline{P(B|E)}\,P(A|E \wedge B)}{P(E)\,\underline{P(B)}\,P(A \mid E \wedge B)} = 0.1 * 0.2 * 0.9 = 0.018.$$

$$P(\neg E \wedge A \wedge B) = P(\neg E)\,P(B)\,P(A \mid \neg E \wedge B) = 0.9 * 0.2 * 0.7 = 0.126.$$

$$\frac{P(E \wedge A \wedge B)}{P(A \wedge B)} = P(E \mid A \wedge B) = \frac{0.018}{0.018 + 0.126} = 0.125$$

$$P(E \wedge A \wedge \neg B) = P(E)\,P(\neg B)\,P(A \mid E \wedge \neg B) = 0.1 * 0.8 * 0.5 = 0.04$$

$$P(E \wedge A) = P(E \wedge A \wedge B) + P(E \wedge A \wedge \neg B) = 0.018 + 0.04 = 0.058$$

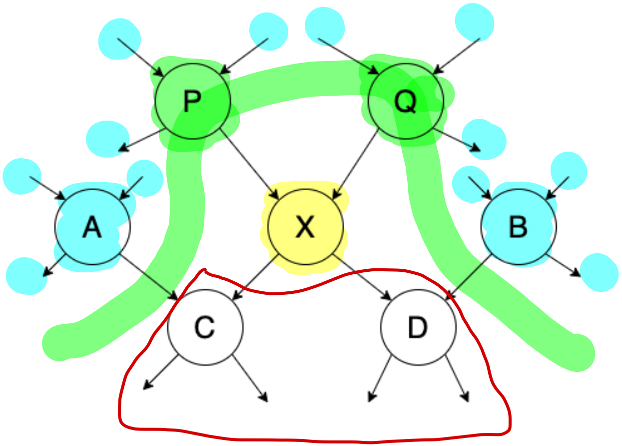
$$P(\neg E \wedge A \wedge \neg B) = P(\neg E)\,P(\neg B)\,P(A \mid \neg E \wedge \neg B) = 0.9 * 0.8 * 0.1 = 0.072$$

$$P(\neg E \wedge A) = P(\neg E \wedge A \wedge B) + P(\neg E \wedge A \wedge \neg B) = 0.126 + 0.072 = 0.198$$

$$P(E \mid A) = \frac{0.058}{0.058 + 0.198} = 0.227$$

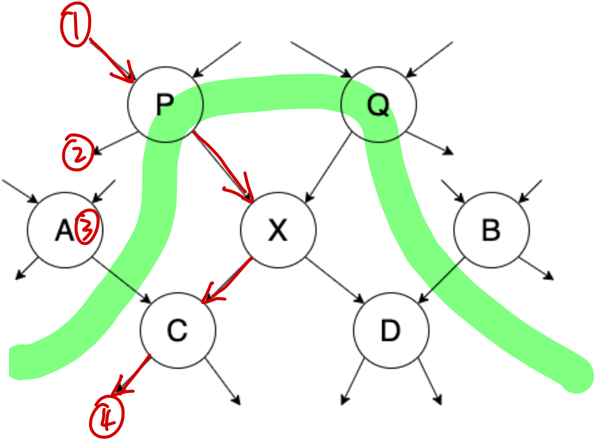
An Independence Relationship in a Bayesian Network

A node is conditionally independent of its non-descendants given its parents.



An Independence Relationship in a Bayesian Network

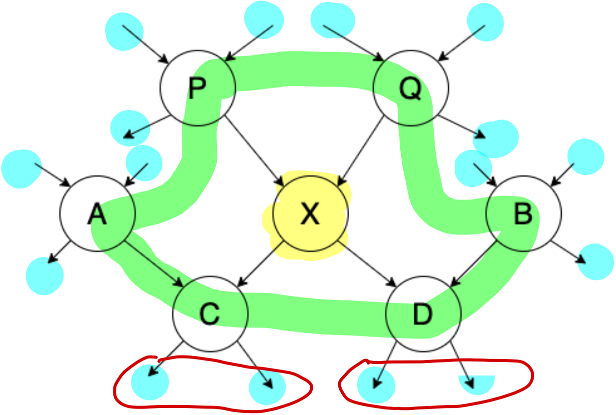
A node is conditionally independent of its non-descendants given its parents.



Another Independence Relationship — The Markov Blanket

A node is conditionally independent of all other nodes given its Markov blanket.

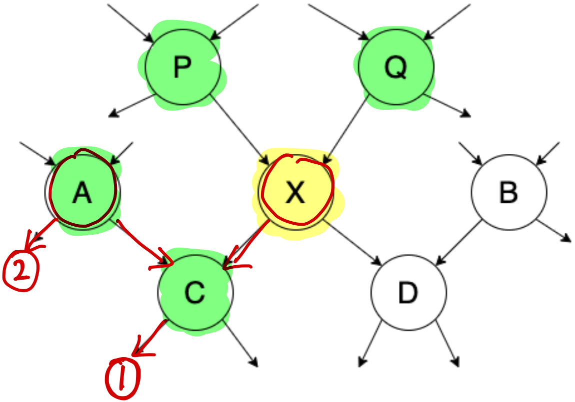
The Markov blanket of a node consists of its parents, its children, and its children's parents.



Another Independence Relationship — The Markov Blanket

A node is conditionally independent of all other nodes given its Markov blanket.

The Markov blanket of a node consists of its parents, its children, and its children's parents.



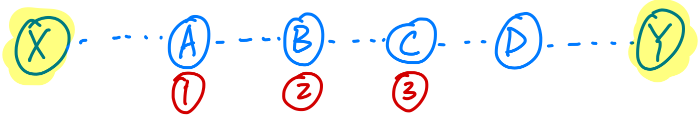
Testing Independence

Given a Bayesian network, how do we determine
if two variables X and Y are independent
if we observe the values of a set of variables E?

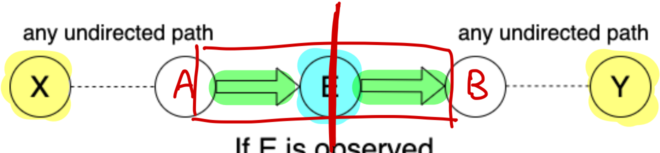
d-separation: A set of variables E d-separates X and Y
if E blocks every undirected path between X and Y in the network.

If E d-separates X and Y,
then X and Y are conditionally independent given E.

What does "block" mean?

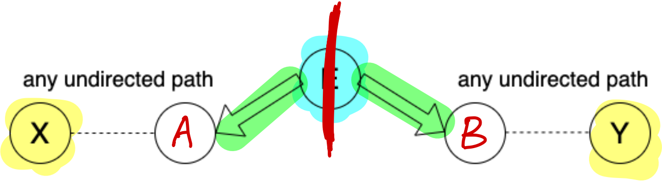


Blocked Paths 1/3



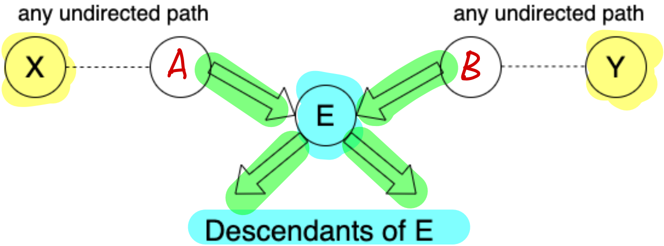
If E is observed,
then it blocks the path between X and Y.

Blocked Paths 2/3



If E is observed,
then it blocks the path between X and Y.

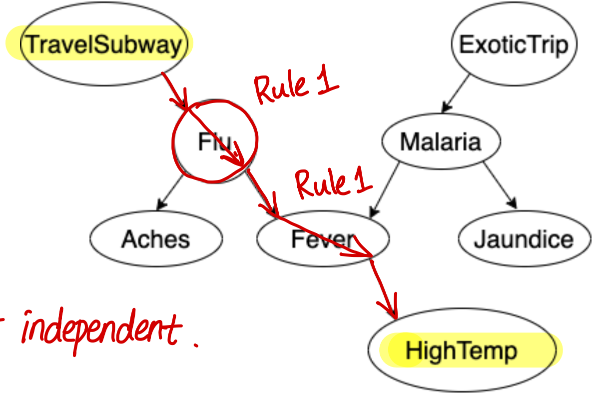
Blocked Paths 3/3



If E and E's descendants are NOT observed,
then they block the path between X and Y.

CQ: Applying D-separation

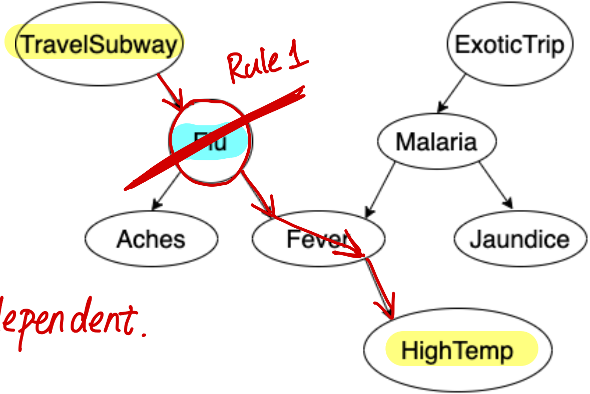
CQ 1a: Are **TravelSubway** and **HighTemp** independent?



Not independent.

CQ: Applying D-separation

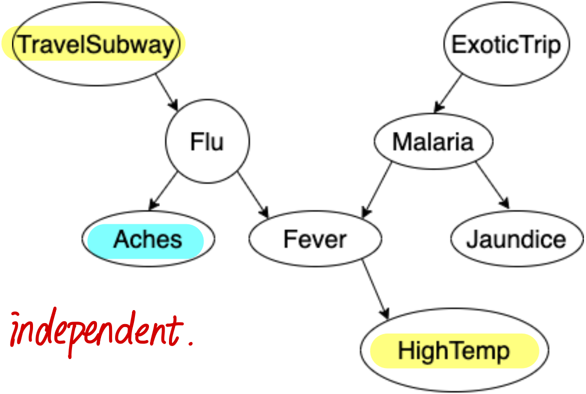
CQ 1b: Are **TravelSubway** and **HighTemp** conditionally independent given **Flu**?



Independent.

CQ: Applying D-separation

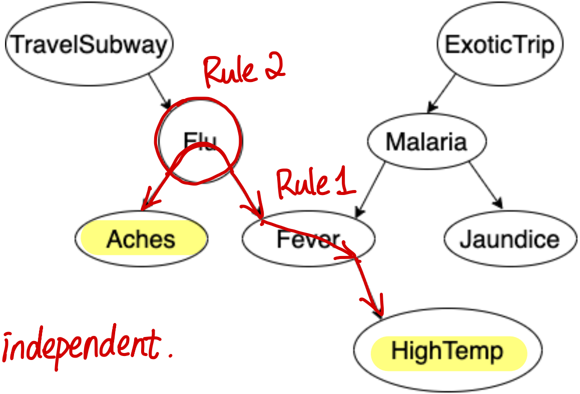
CQ 1c: Are **TravelSubway** and **HighTemp** conditionally independent given **Aches**?



Not independent.

CQ: Applying D-separation

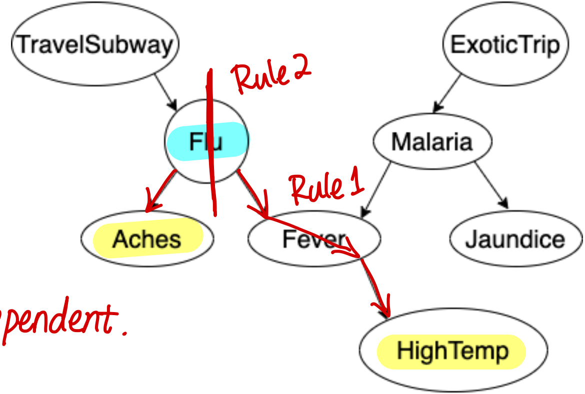
CQ 2a: Are **Aches** and **HighTemp** independent?



Not independent.

CQ: Applying D-separation

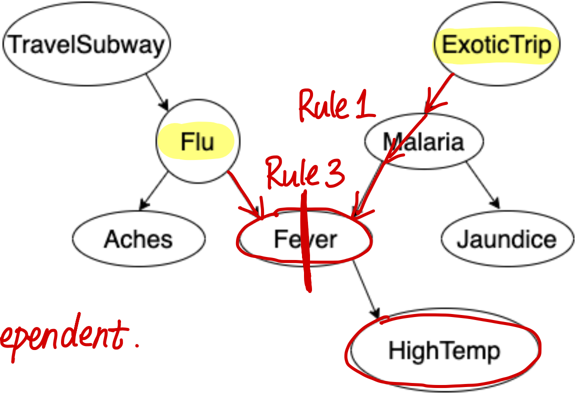
CQ 2b: Are **Aches** and **HighTemp** conditionally independent given **Flu**?



Independent.

CQ: Applying D-separation

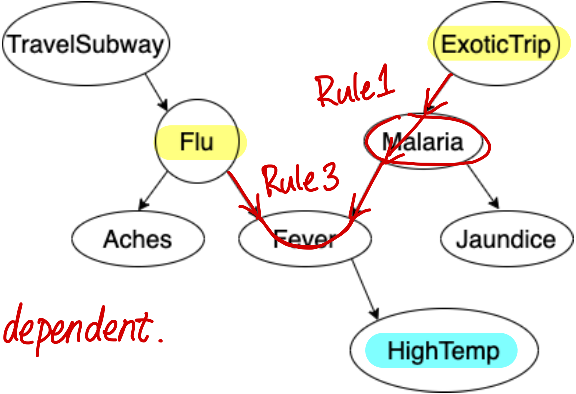
CQ 3a: Are **Flu** and **ExoticTrip** ~~conditionally~~ independent?



Independent.

CQ: Applying D-separation

CQ 3b: Are **Flu** and **ExoticTrip** conditionally independent given **HighTemp**?



Not independent.

Constructing Bayesian Networks

For a joint probability distribution,
there are multiple correct Bayesian networks.

A Bayesian network is correct if every independence relationship the network represents is correct. $\rightsquigarrow$ exists in the joint distribution.

We prefer one Bayesian network over another one if the former requires fewer probabilities.

Constructing a Correct Bayesian Network

1. Determine the set of variables for the domain.
2. Order the variables $\{X_1, \ldots, X_n\}$.
3. For each variable X_i in the ordering,
   3.1 **Choose the node's parents:**
   Choose the smallest set of parents from $\{X_1, \ldots, X_{i-1}\}$ such that given $Parents(X_i)$ X_i is independent of all the nodes in $\{X_1, \ldots, X_{i-1}\} - Parents(X_i)$. Formally,

$$P(X_i | Parents(X_i)) = P(X_i | X_{i-1} \wedge \cdots \wedge X_1).$$

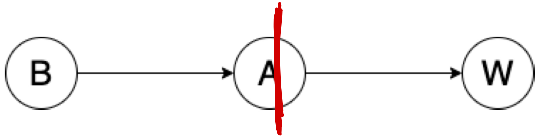
   3.2 Create a link from each parent of X_i to the node X_i.
   3.3 Write down the conditional probability table $P(X_i | Parents(X_i))$.

Example 1: Construct a Bayes Net

Consider the Bayesian network.

Construct a correct Bayesian network
based on the variable ordering: W, A, and B.

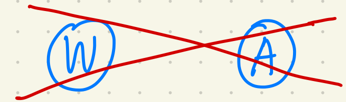
Consider the Bayesian network.

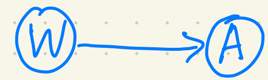


Construct a correct Bayesian network
based on the variable ordering: W, A, and B.

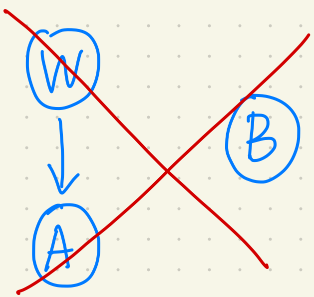
step 1: W

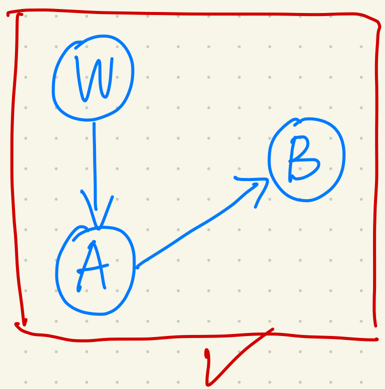
step 2: add A to the network.

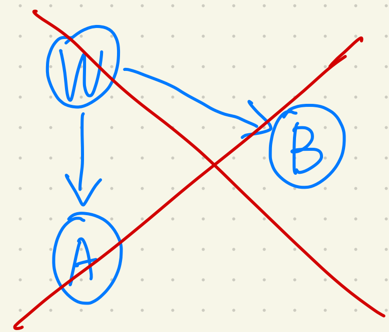




step 3: add B to the network.

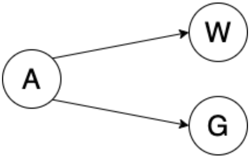




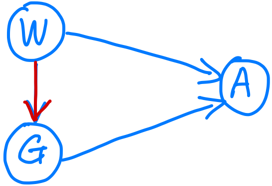


Example 2: Construct a Bayes Net

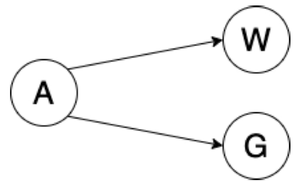
Consider the Bayesian network:



Construct a correct Bayesian network
based on the variable ordering: W, G, and A.



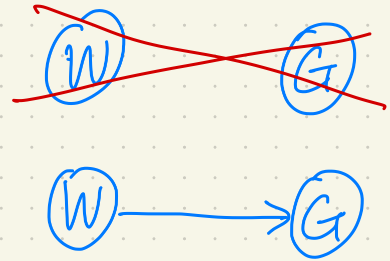
Consider the Bayesian network:



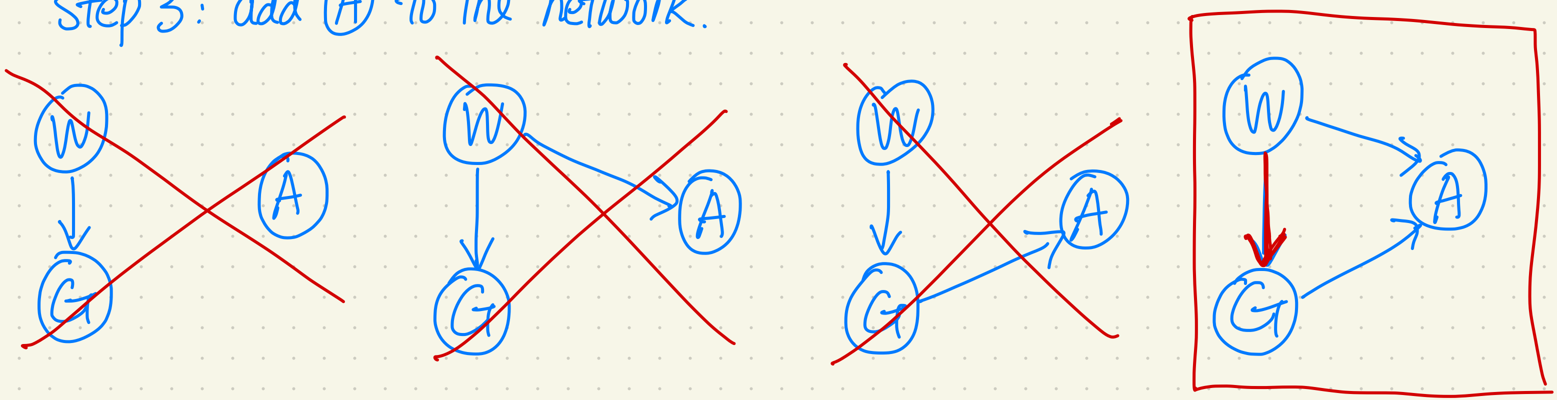
Construct a correct Bayesian network based on the variable ordering: W, G, and A.

step 1 : W

step 2 : add G to the network.

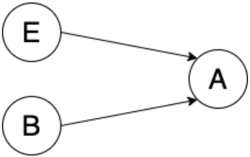


step 3: add A to the network.

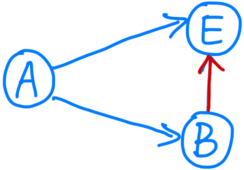


Example 3: Construct a Bayes Net

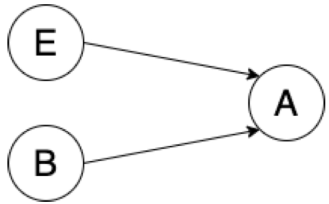
Consider the Bayesian network.



Construct a correct Bayesian network
based on the variable ordering: A, B, and E.



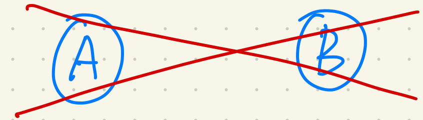
Consider the Bayesian network.

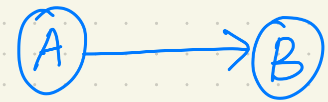


Construct a correct Bayesian network based on the variable ordering: A, B, and E.

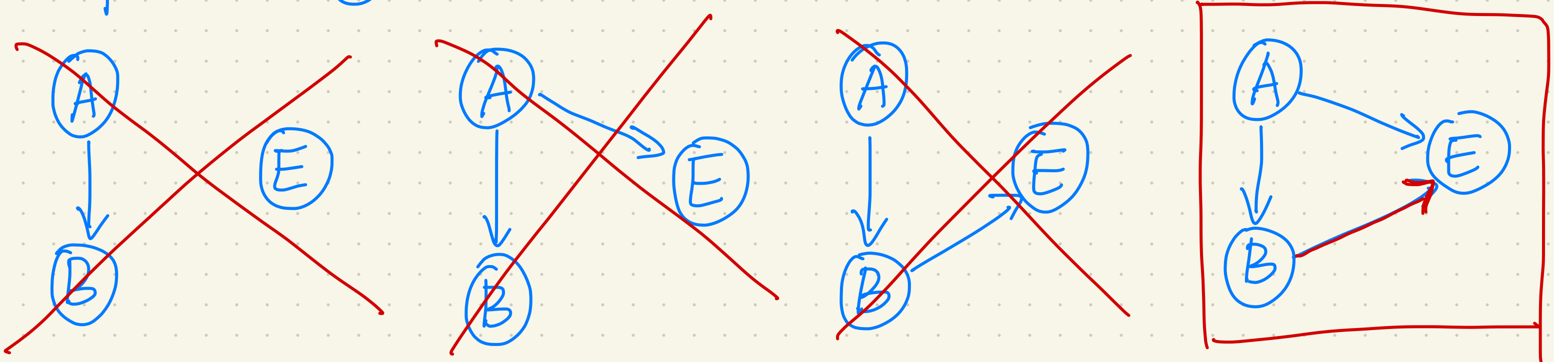
Step 1: A

Step 2: add B to the network.





step 3: add E to the network.



Constructing a Compact Bayesian Network

How can we choose the variable ordering such that
the resulting Bayes Net requires as few probabilities as possible?

If causes precede effects, we get a more compact Bayesian Network.

Revisiting the Learning Goals

By the end of the lecture, you should be able to

- Compute a joint probability given a Bayesian network.
- Identify the conditional independence assumptions required by a Bayesian network.
- Given a Bayesian network, determine if two variables are independent or conditionally independent given a third variable.
- Given a scenario with independent assumptions and a given order of the variables, construct a Bayesian network by adding the variables to the network based on the given order.