AI: Representation and Problem Solving

Bayes Nets: Independence

Instructors: Tuomas Sandholm and Vincent Conitzer

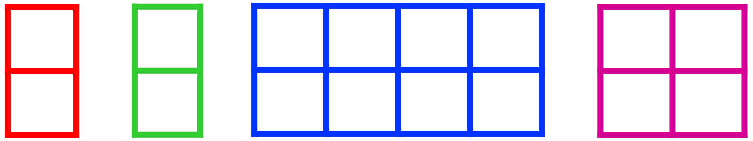
Slide credits: CMU AI and http://ai.berkeley.edu

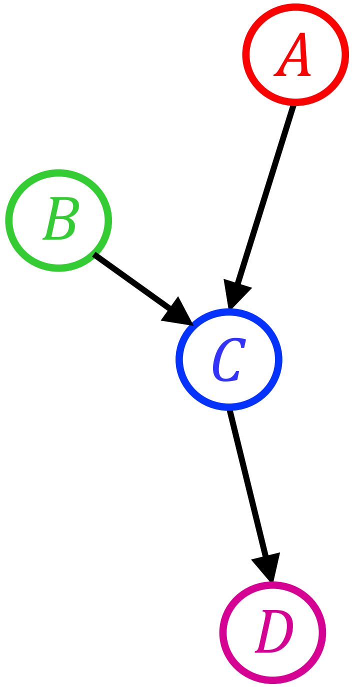
Announcements

- P4 due tonight!

Bayesian Networks

- One node per random variable
- Directed Acyclic Graph
- One CPT per node: P(node | *Parents*(node))





Bayes net

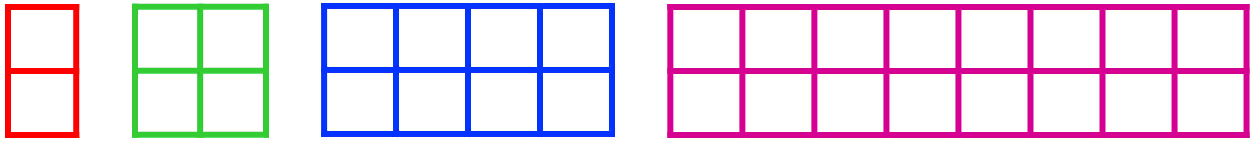
$$P(A, B, C, D) = P(A)\, P(B)\, P(C|A, B)\, P(D|C)$$

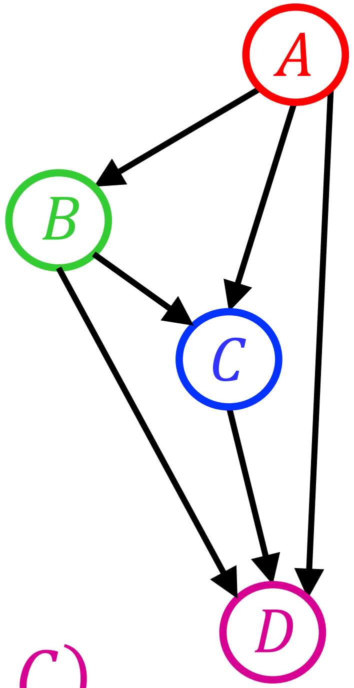
Encode joint distributions as product of conditional distributions on each variable

$$P(X_1, \dots, X_N) = \prod_i P(X_i |\, Parents(X_i))$$

Any distribution **can** be a Bayes net

- One node per random variable
- Directed-Acyclic-Graph
- One CPT per node: P(node | *Parents*(node))





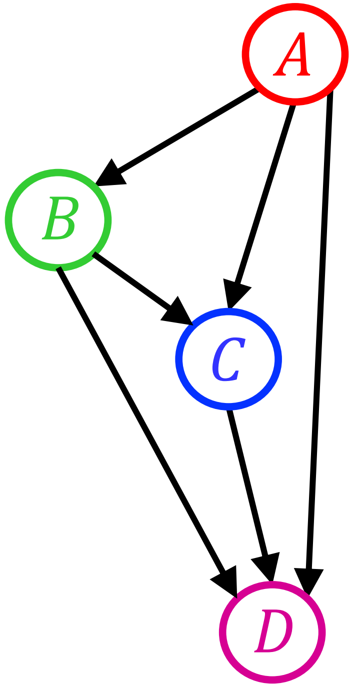
Bayes net

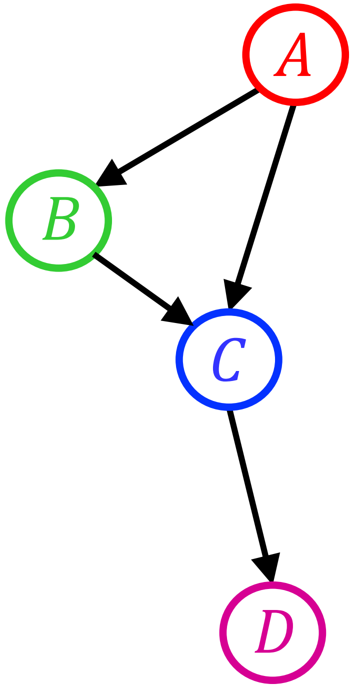
$$P(A,B,C,D) = P(A)\,P(B|A)\,P(C|A,B)\,P(D|A,B,C)$$

Encode joint distributions as product of conditional distributions on each variable

$$P(X_1,\ldots,X_N) = \prod_i P(X_i|\,Parents(X_i))$$

Joint distribution from Bayes nets





$P(A)\ P(B|A)\ P(C|A,B)\ P(D|A,B,C)$

$P(A)\ P(B|A)\ P(C|A,B)\ P(D|C)$

If these encode the same distribution then

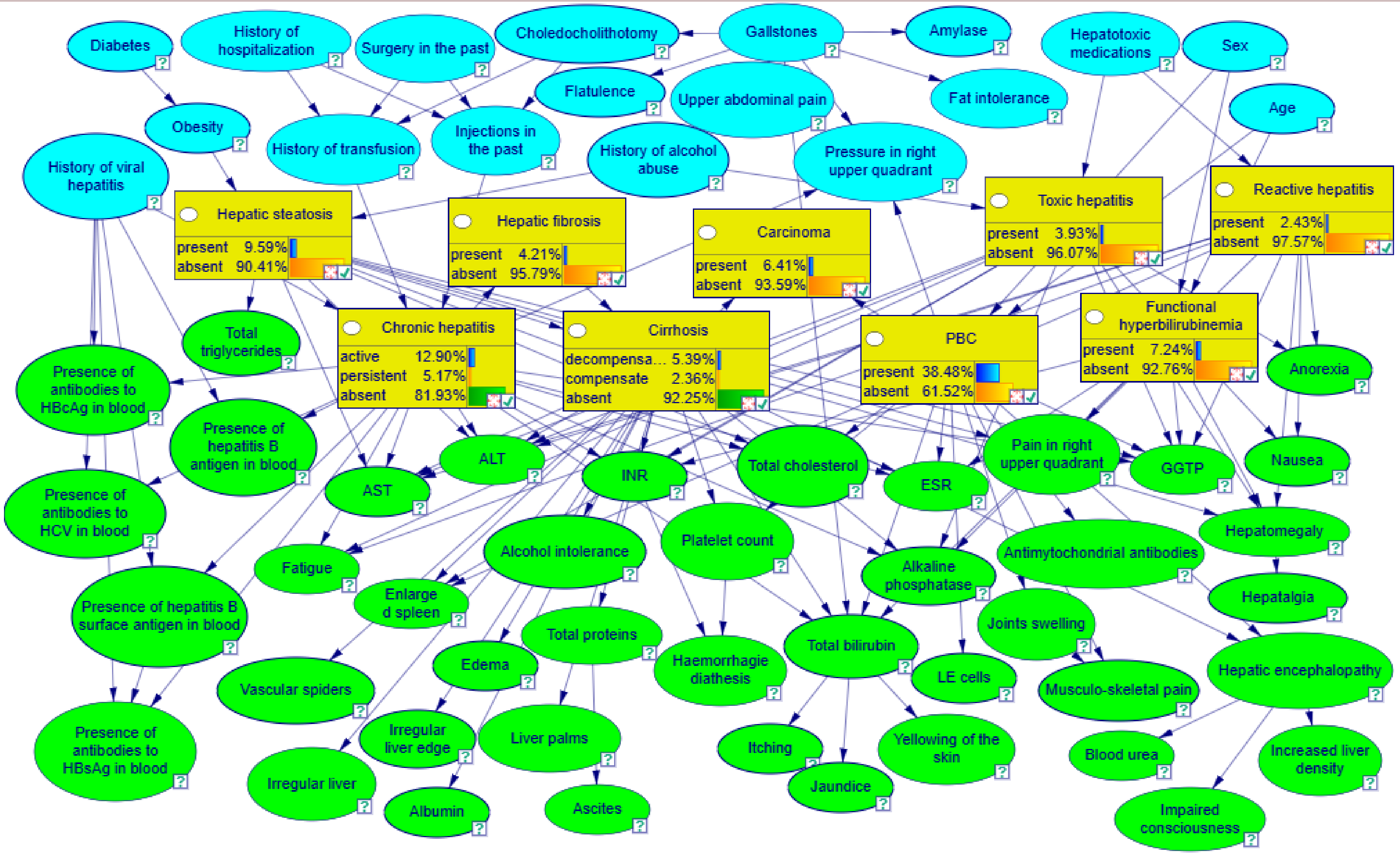
$$P(D \mid A, B, C) = P(D \mid C)$$

The power of **conditional independencies**

- For a Bayes net $P(X_1, \dots, X_N) = \prod_i P(X_i | Parents(X_i))$

- When $|Parents(X_i)|$ is small, the conditional probability tables are much smaller
  - Makes them easier to estimate from data

Example: Liver Disorders



https://demo.bayesfusion.com/bayesbox.html

Independence

- Two variables X and Y are *independent* if

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

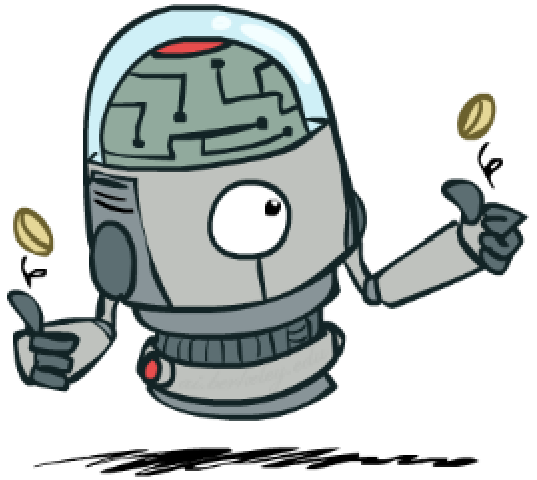
  - This says that their joint distribution ***factors*** into a product of two simpler distributions
  - Combine with product rule $P(x,y) = P(x|y)P(y)$ we obtain another form:

$$\forall x,y\ P(x \mid y) = P(x) \quad \text{or} \quad \forall x,y\ P(y \mid x) = P(y)$$

- Example: two dice rolls R_1 and R_2

$P(R_1=5, R_2=5) = P(R_1=5)\, P(R_2=5) = 1/6 \times 1/6 = 1/36$

$P(R_2=5 \mid R_1=5) = P(R_2=5)$



Example: Independence

- n fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

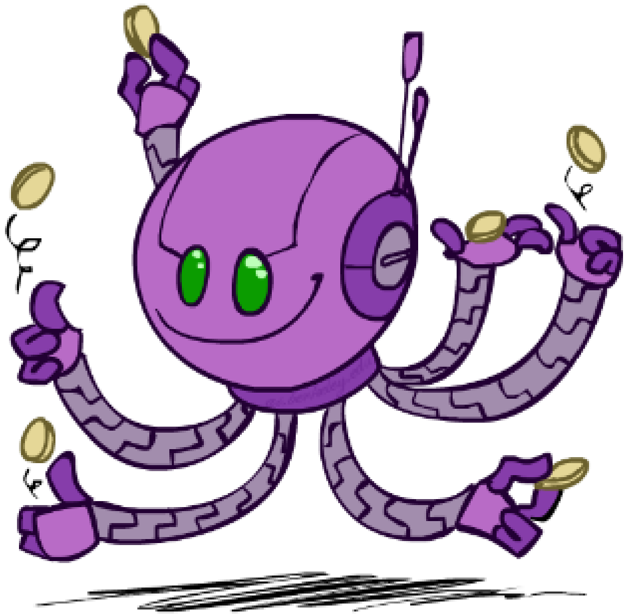
$P(X_2)$

H	0.5
T	0.5

. . .

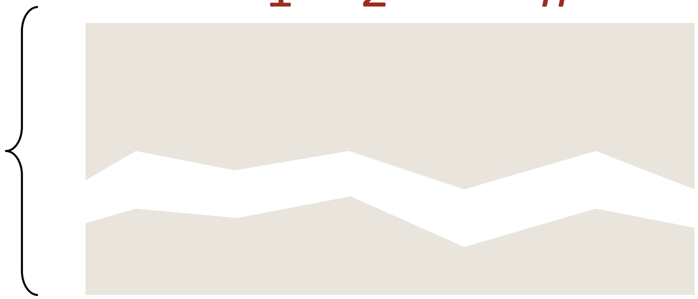
$P(X_n)$

H	0.5
T	0.5



$P(X_1, X_2, \ldots, X_n)$

2^n



Poll 1

- Are T and W independent?

$P(T)$

T	P
hot	0.5
cold	0.5

$P(T,W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P(W)$

W	P
sun	0.6
rain	0.4

Conditional Independence

- Absolute (unconditional) independence very rare

- *Conditional independence* is our most basic and robust form of structural knowledge about uncertain environments.

- X is conditionally independent of Y given Z
  if and only if:

$$\forall x,y,z \qquad P(x \mid y, z) = P(x \mid z)$$

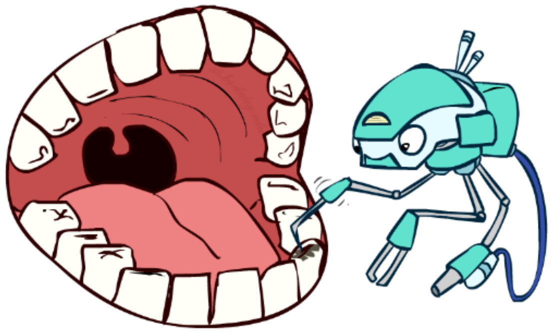
  or, equivalently, if and only if

$$\forall x,y,z \qquad P(x, y \mid z) = P(x \mid z)\, P(y \mid z)$$

Conditional independence

- X and Y are independent if
  - $P(X, Y) = P(X)\,P(Y)$, or
  - $P(X \mid Y) = P(X)$
- X and Y are **conditionally independent given Z** if
  - $P(X, Y \mid Z) = P(X \mid Z)\,P(Y \mid Z)$, or
  - $P(X \mid Y, Z) = P(X \mid Z)$

Conditional independence



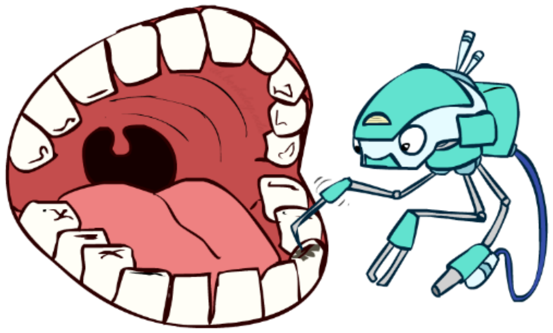
- P(Toothache, Cavity, (p)Robe)

- If I have a cavity, the probability that the probe catches in it **doesn't** depend on whether I have a toothache:
  - $P(+r \mid +toothache, +cavity) = P(+r \mid +cavity)$

- The same independence holds if I don't have a cavity:
  - $P(+r \mid +toothache, -cavity) = P(+r \mid -cavity)$

- Probe is *conditionally independent* of Toothache given Cavity:
  - $P(R \mid T, C) = P(R \mid C)$

Conditional independence



Equivalent statements:

- P(Toothache | Probe , Cavity) = P(Toothache | Cavity)

- P(Toothache, Probe | Cavity) = P(Toothache | Cavity) P(Probe | Cavity)

- One can be derived from the other easily

Independence Rules

- Independence
  If A and B are independent, then:

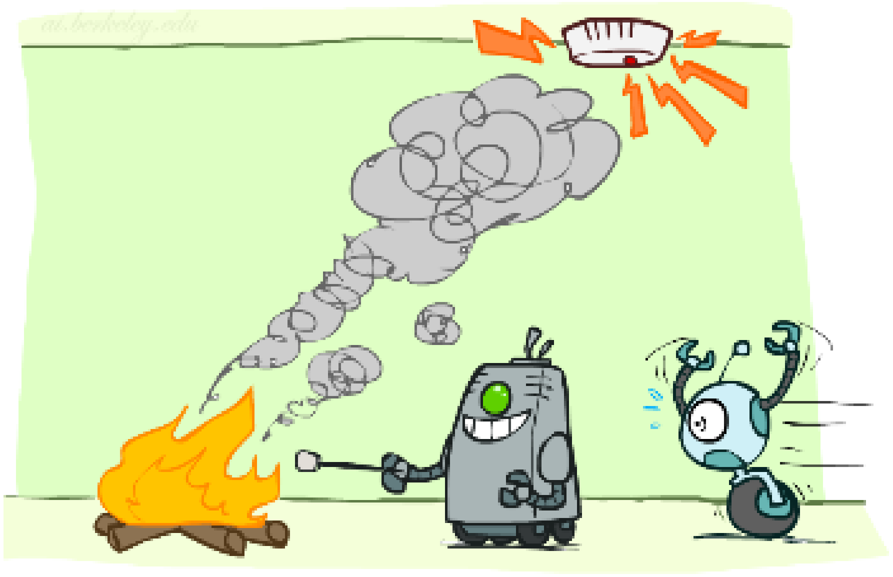
$$P(A,B) = P(A)P(B)$$
$$P(A \mid B) = P(A)$$
$$P(B \mid A) = P(B)$$

- Conditional independence
  If A and B are conditionally independent given C, then:

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

Conditional Independence and Bayes Nets

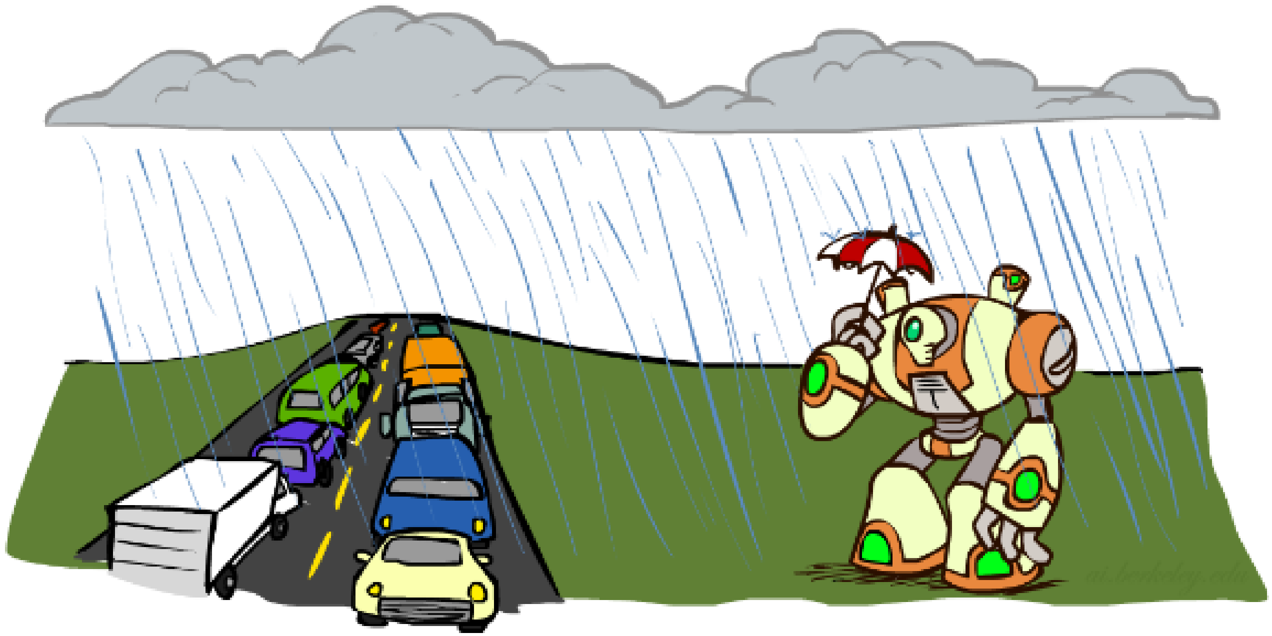
- Fire, Smoke, Alarm
  - Causal story to create Bayes net



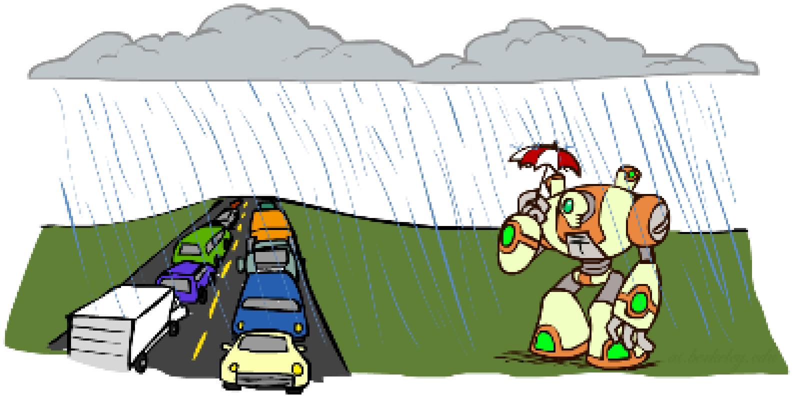
- From Bayes Net: $P(S, F, A) = P(F)\, P(S \mid F)\, \mathbf{P(A \mid S)}$
- Joint distribution: $P(S, F, A) = P(F)\, P(S \mid F)\, \mathbf{P(A \mid S, F)}$

Conditional Independence and Bayes Nets

- What about this domain:
  - Traffic
  - Umbrella
  - Raining



Conditional Independence and the Chain Rule



- Chain rule:

  $P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i \mid x_1, \ldots, x_{i-1})$

- Trivial decomposition:
- $P(Rain, Traffic, Umbrella) =$

- With assumption of conditional independence:
- $P(Rain, Traffic, Umbrella) =$

- Bayes nets / graphical models help us express conditional independence assumptions

Poll 2

Choose the true statement(s):

(A) If X and Y are conditionally independent given Z, then X and Y are independent

(B) If X and Y are independent, then X and Y are also conditionally independent given Z

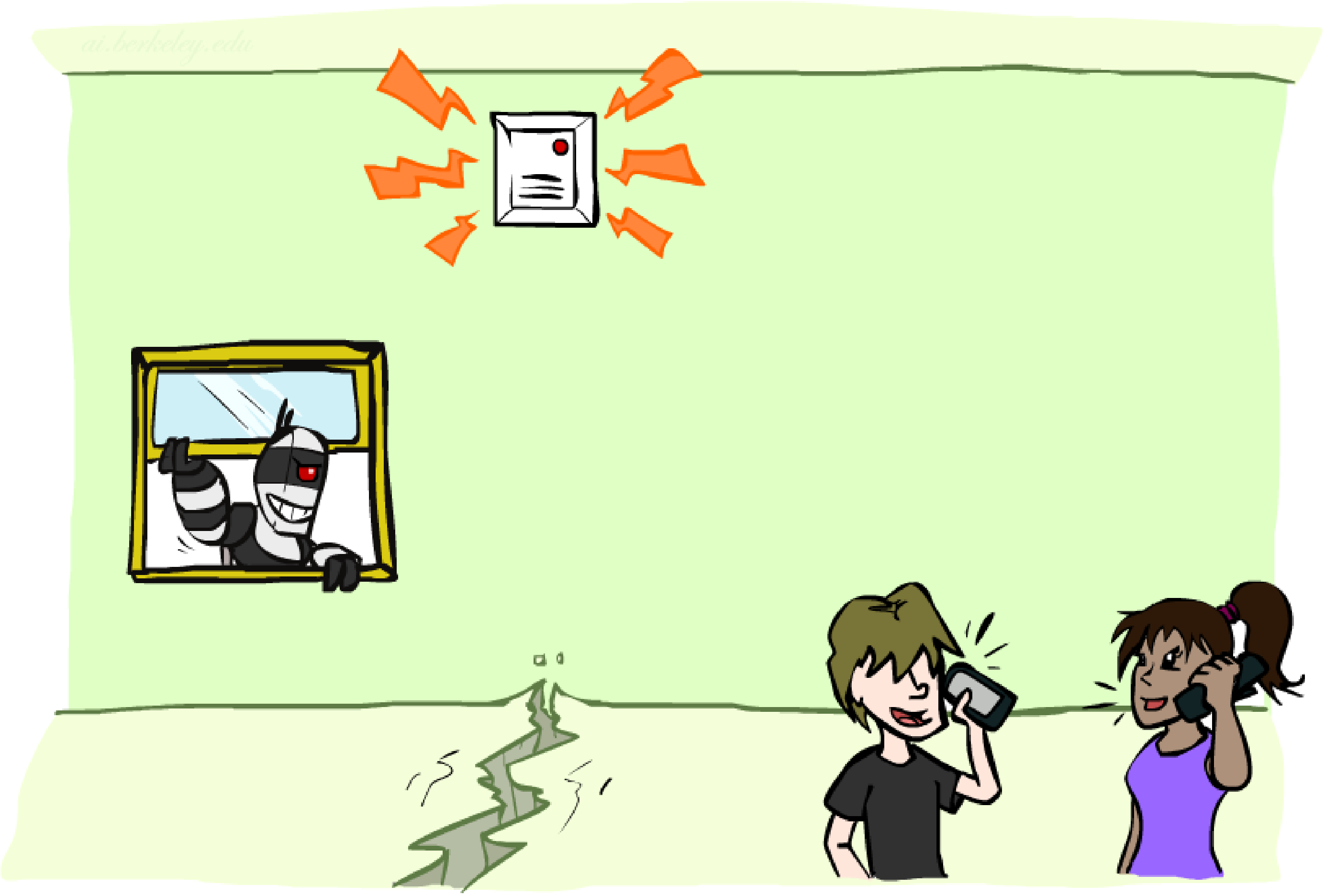
(C) Neither is true

Example: Traffic II

- Let's build a causal Bayes net!
- Variables
  - T: Traffic
  - R: It rains
  - L: Low pressure
  - D: Roof drips
  - H: Hockey game
  - C: Cavity

Example: alarm network

- Variables
  - **B:** Burglary
  - **A:** Alarm goes off
  - **M**: Mary calls
  - **J:** John calls
  - **E:** Earthquake!



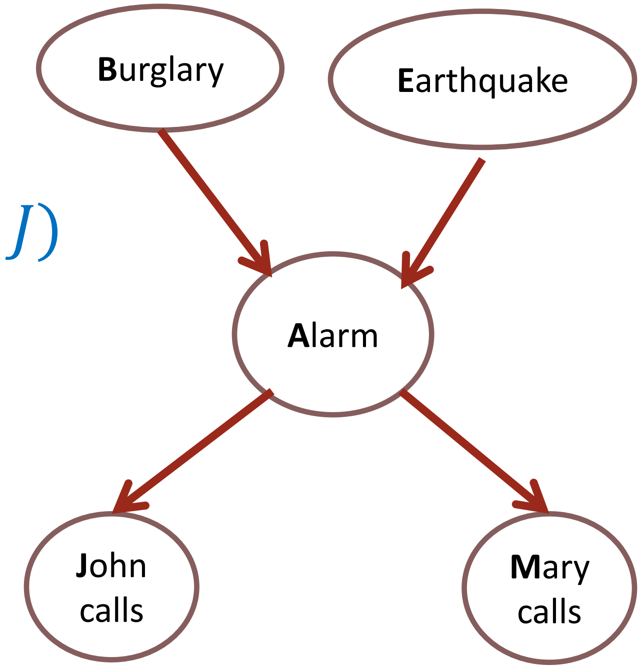
Analyzing the structure of the alarm network

- Joint distribution factorization example

- Generic chain rule
  - $P(X_1 \dots X_n) = \prod_i P(X_i | X_1 \dots X_{i-1})$

$$P(B,E,A,J,M) = P(B)\,P(E|B)\,P(A|B,E)\,P(J|B,E,A)\,P(M|B,E,A,J)$$

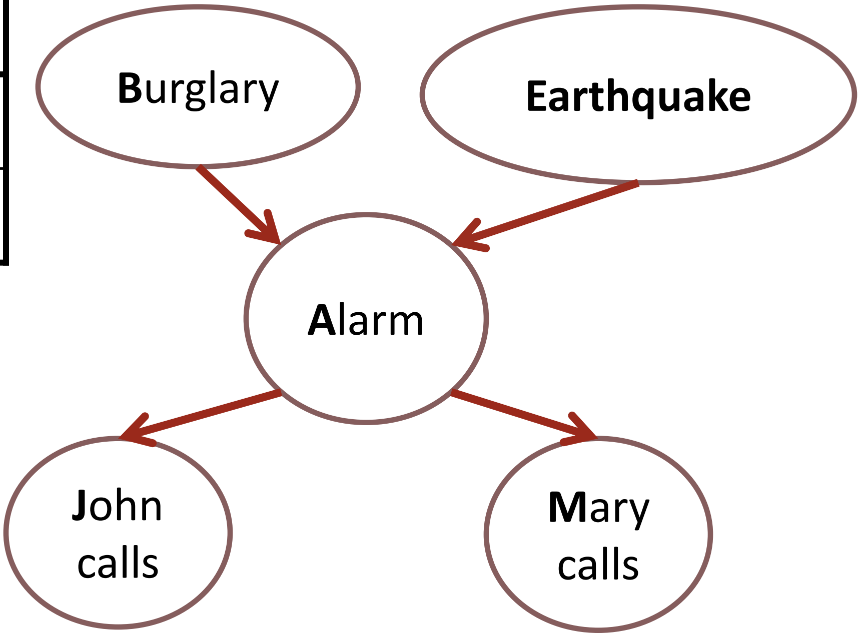
$$P(B,E,A,J,M) = P(B)\,P(E)\quad P(A|B,E)\,P(J|A)\qquad P(M|A)$$



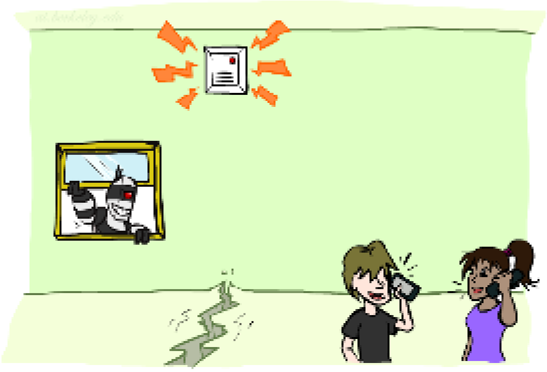
- Bayes nets
  - $P(X_1 \dots X_n) = \prod_i P(X_i | Parents(X_i))$

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998



A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

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

Common Effect

- Chain rule:

  $P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i \mid x_1, \ldots, x_{i-1})$

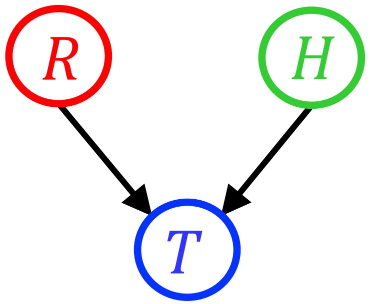
- Trivial decomposition:

  $P(Rain, Hockey, Traffic) = P(Rain)\, P(Hockey \mid Rain)\, P(Traffic \mid Rain, Hockey)$

- With assumption of conditional independence:
- $P(Rain, Hockey, Traffic) =$

- Bayes nets / graphical models help us express conditional independence assumptions

Common Effect



- Chain rule:

  $P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i \mid x_1, \ldots, x_{i-1})$

- Trivial decomposition:

  $P(Rain, Hockey, Traffic) = P(Rain)\, P(Hockey \mid Rain)\, P(Traffic \mid Rain, Hockey)$

- With assumption of conditional independence:

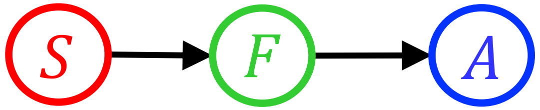
  $P(Rain, Hockey, Traffic) = P(Rain)\, P(Hockey)\, P(Traffic \mid Rain, Hockey)$

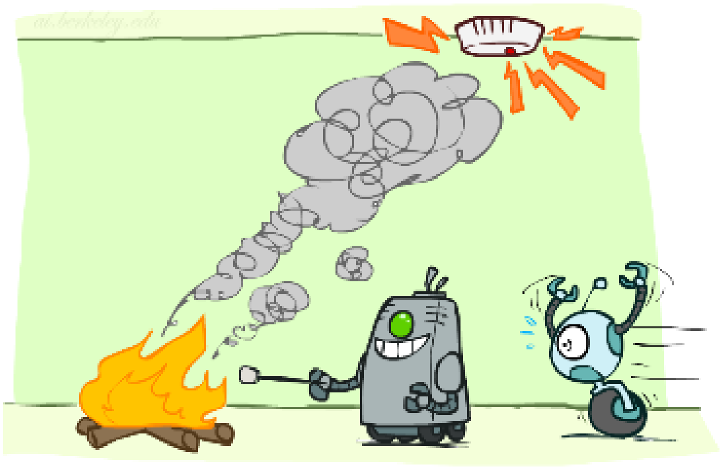
- Bayes nets / graphical models help us express conditional independence assumptions

Conditional Independence Semantics

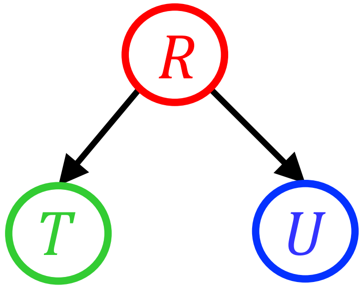
- Important local relationships within a Bayes net

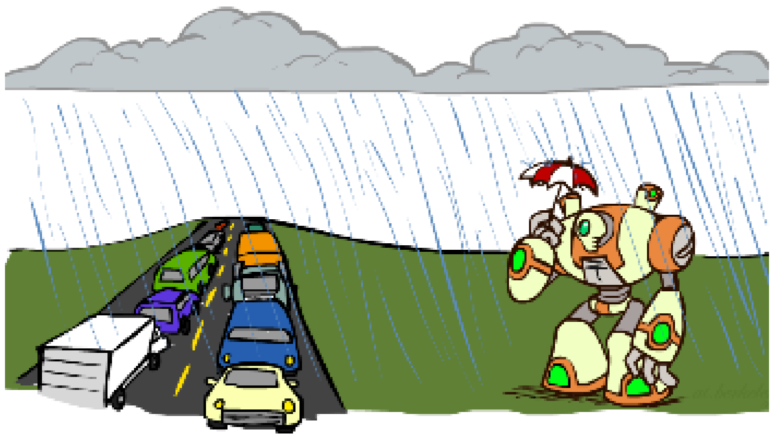
Causal Chain



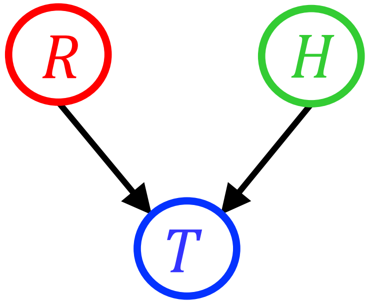


Common Cause



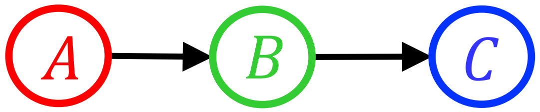


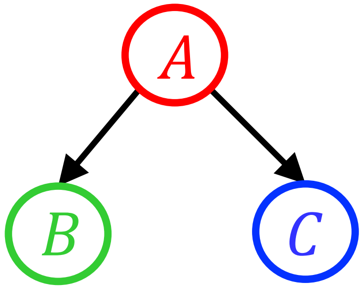
Common Effect

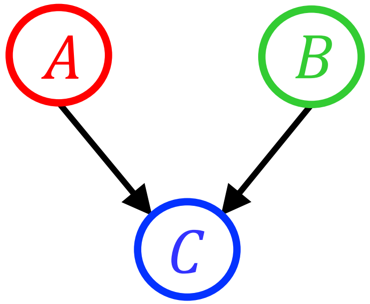


Poll 3: Conditional independence from Bayes Nets

- Match the product of CPTs to the Bayes net.



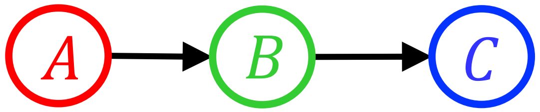


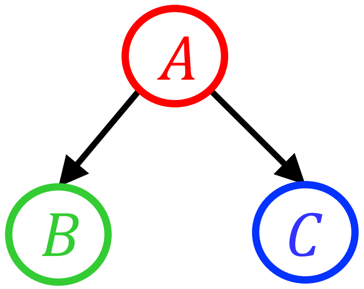


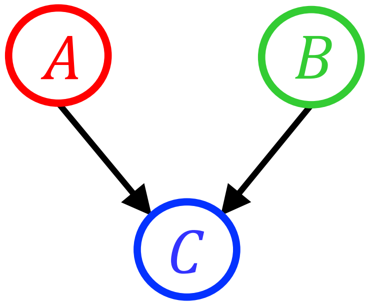
- I. $P(A)\,P(B|A)\,P(C|B)$ $P(A)\,P(B|A)\,P(C|A)$ $P(A)\,P(B)\,P(C|A,B)$
- II. $P(A)\,P(B)\,P(C|A,B)$ $P(A)\,P(B|A)\,P(C|B)$ $P(A)\,P(B|A)\,P(C|A)$
- III. $P(A)\,P(B|A)\,P(C|B)$ $P(A|B,C)\,P(B)\,P(C)$ $P(A)\,P(B|A)\,P(C|B)$

Conditional Independence Semantics

- For the following Bayes nets, write the joint $P(A, B, C)$
  1. Using the chain rule (with top-down order A,B,C)
  2. Using Bayes net semantics (product of CPTs)

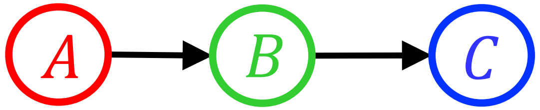






Conditional Independence Semantics

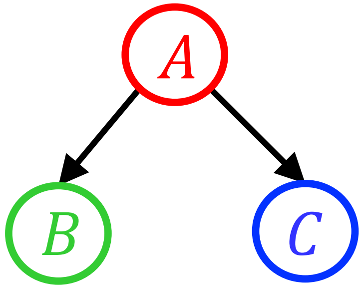
- For the following Bayes nets, we write the joint $P(A, B, C)$
  1. Using the chain rule (with top-down order A,B,C)
  2. Using Bayes net semantics (product of CPTs)



$P(A)\ P(B|A)\ P(C|A,B)$

$P(A)\ P(B|A)\ P(C|B)$

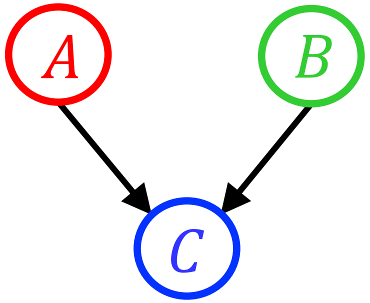
Assumption:
$P(C|A,B) = P(C|B)$
C is independent from A given B



$P(A)\ P(B|A)\ P(C|A,B)$

$P(A)\ P(B|A)\ P(C|A)$

Assumption:
$P(C|A,B) = P(C|A)$
C is independent from B given A



$P(A)\ P(B|A)\ P(C|A,B)$

$P(A)\ P(B)\ P(C|A,B)$

Assumption:
$P(B|A) = P(B)$
A is independent from B given { }

Causal Chains

- This configuration is a "causal chain"

X: Low pressure Y: Rain Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

Guaranteed X independent of Z ?
No!

- One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed
  - Low pressure *always* causes rain
  - Rain *always* causes traffic
  - High pressure *always* causes no rain
  - No rain *always* causes no traffic

  - Then:
    P(+z | +x) = 1
    P(+z | -x) = 0

Causal Chains

- This configuration is a "causal chain"

X: Low pressure Y: Rain Z: Traffic

$$P(x, y, z) = P(x)P(y|x)P(z|y)$$

- Guaranteed X independent of Z given Y?

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

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

$$= P(z|y)$$

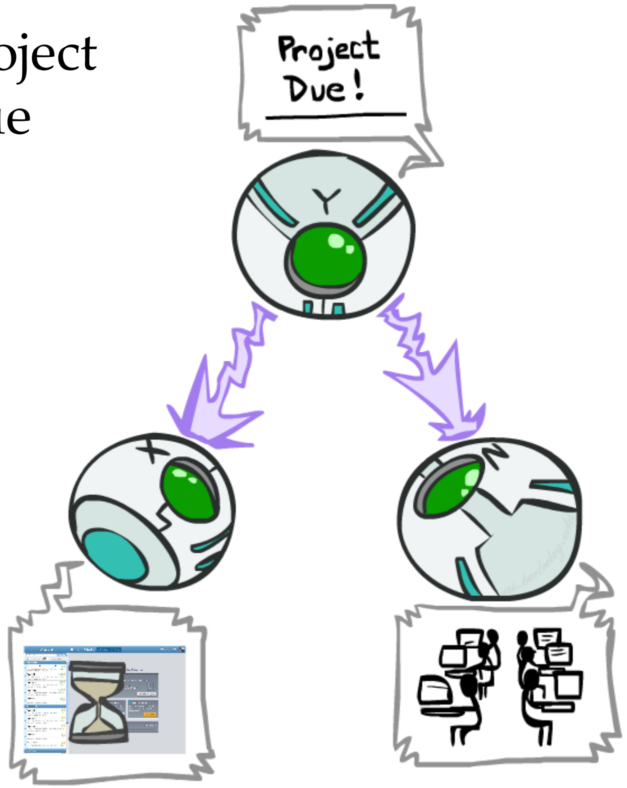
Yes!

- Evidence along the chain "blocks" the influence from the beginning to the end

Common Cause

- This configuration is a "common cause"

Y: Project due



X: Forums busy

Z: Lab full

$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

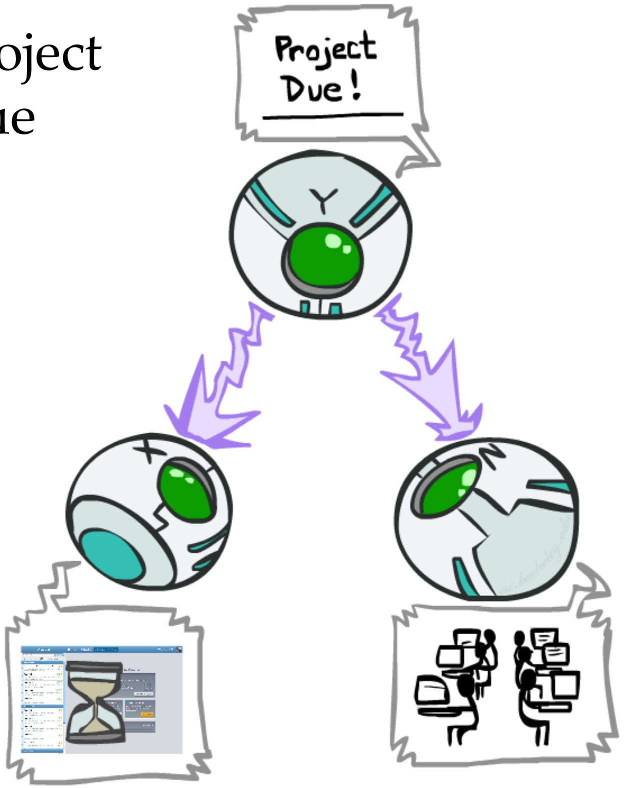
- Guaranteed X independent of Z ? *No!*
  - *One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.*
  - Example:
    - Project due *always* causes both forums busy and lab full;
      not project due *always* causes both forums not busy and lab not full
    - Then:

      P(+x | +z) = 1
      P(+x | -z) = 0

Common Cause

- This configuration is a "common cause"

Y: Project due



X: Forums busy

Z: Lab full

$$P(x, y, z) = P(y)P(x|y)P(z|y)$$

- Guaranteed X and Z independent given Y?

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

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

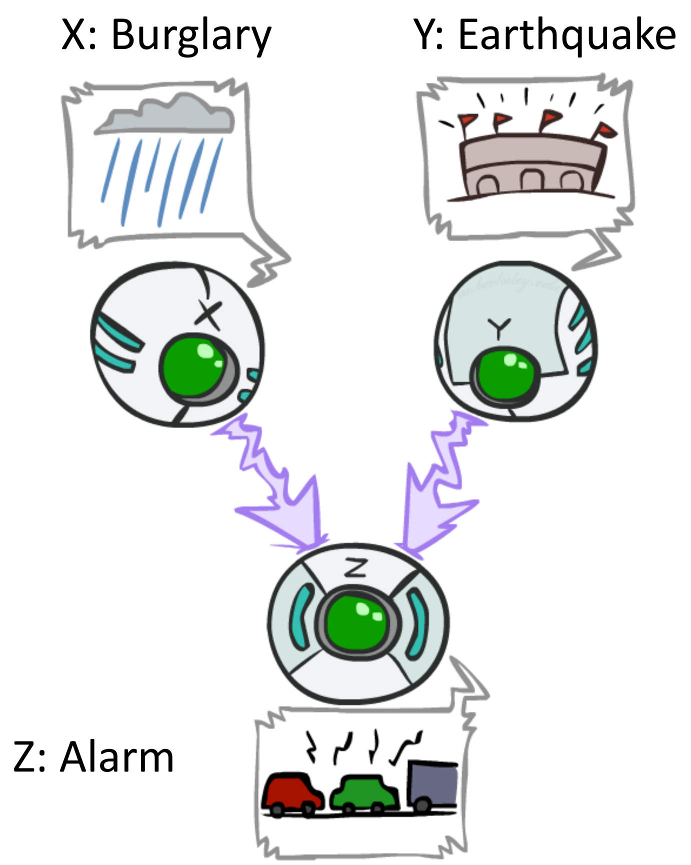
$$= P(z|y)$$

Yes!

- Observing the cause blocks influence between effects.

Common Effect

- Last configuration: two causes of one effect (v-structures)

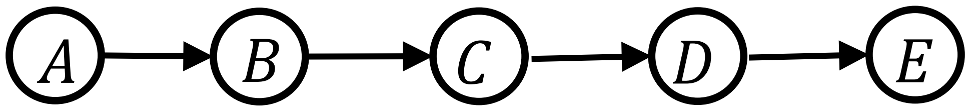


- Are X and Y independent?
  - *Yes*: burglary and earthquake cause alarm, but they are not correlated
  - Still need to prove they must be (try it!)
- Are X and Y independent given Z?
  - *No*: observing alarm puts the burglary and the earthquake in competition as explanation.
- This is backwards from the other cases
  - Observing an effect activates influence between possible causes. **Explaining away** effect**.**

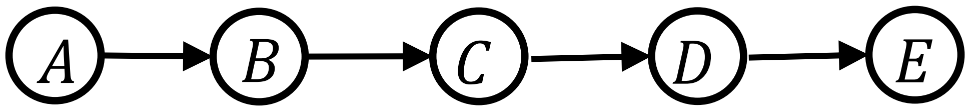
Bayes Net Independence

Answering Independence Questions

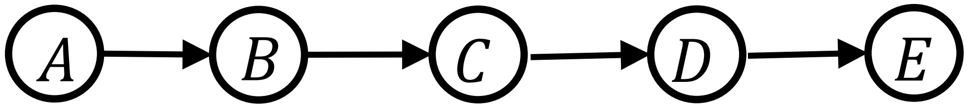
- Is A independent from E?



- Is A independent from E given C?

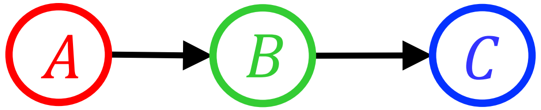


- Is A independent from C given E?

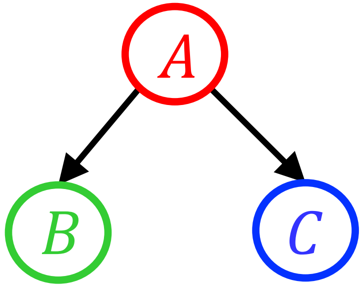


Summary from before

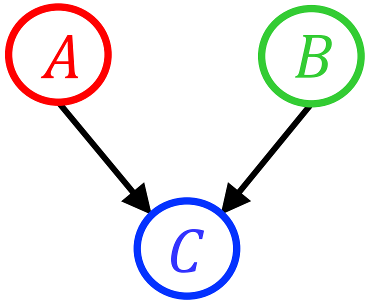
Causal chain



Common cause



Common effect (v-structure)

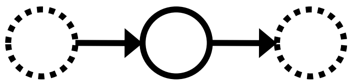


Given a Bayes net, which conditional independences hold?

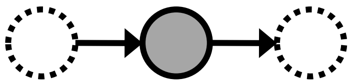
- Question: Given a Bayes net, are X and Y conditionally independent given some "evidence" variables (E, F, G…)?
- Influence is generally exerted via edges in Bayes net
- Idea: trace **undirected paths** from X to Y and see if X and Y exert influence on each other via any path
  - No active paths => conditional independence

Active and inactive paths

- A path is active if **each consecutive triple of nodes on the path** is active
  - If we get inactivity (independence) at any point, the influence of X on Y is blocked
- Shaded nodes are evidence (or "given" nodes)
- Causal chain



Active

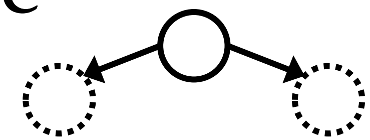


Inactive

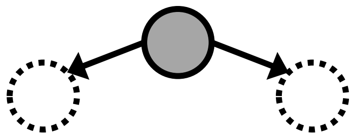
- *Knowing about intermediate node gives independence*

Active and inactive paths

- A path is active if **each consecutive triple of nodes on the path** is active
  - If we get inactivity (independence) at any point, the influence of X on Y is blocked
- Shaded nodes are evidence (or "given" nodes)
- Common cause



Active

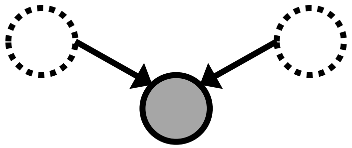


Inactive

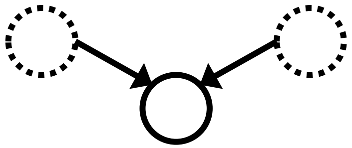
- *Children are dependent, but become independent when conditioned on parent*

Active and inactive paths

- A path is active if **each consecutive triple of nodes on the path** is active
  - If we get inactivity (independence) at any point, the influence of X on Y is blocked
- Shaded nodes are evidence (or "given" nodes)
- Common effect



Active

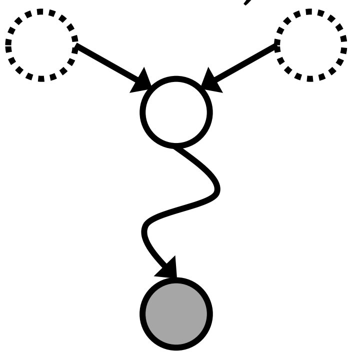


Inactive

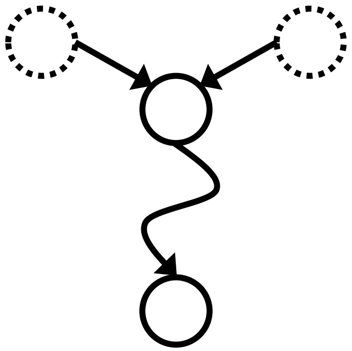
- *Parents were independent, but become dependent if the child is known (**explaining away** effect)*

Active and inactive paths

- A path is active if **each consecutive triple of nodes on the path** is active
  - If we get inactivity (independence) at any point, the influence of X on Y is blocked
- Shaded nodes are evidence (or "given" nodes)
- Common effect



Active



Inactive

- Explaining away effect also holds if: some **descendant** of child is observed!

Summary

- Question: Are X and Y conditionally independent given evidence variables {Z}?
- Consider all (undirected) paths from X to Y
- A path is active if **each consecutive triple** is active:
  - Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
  - Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
  - Common effect (aka v-structure) $A \rightarrow B \leftarrow C$ where B *or one of its descendants* is observed
- No active paths => independence

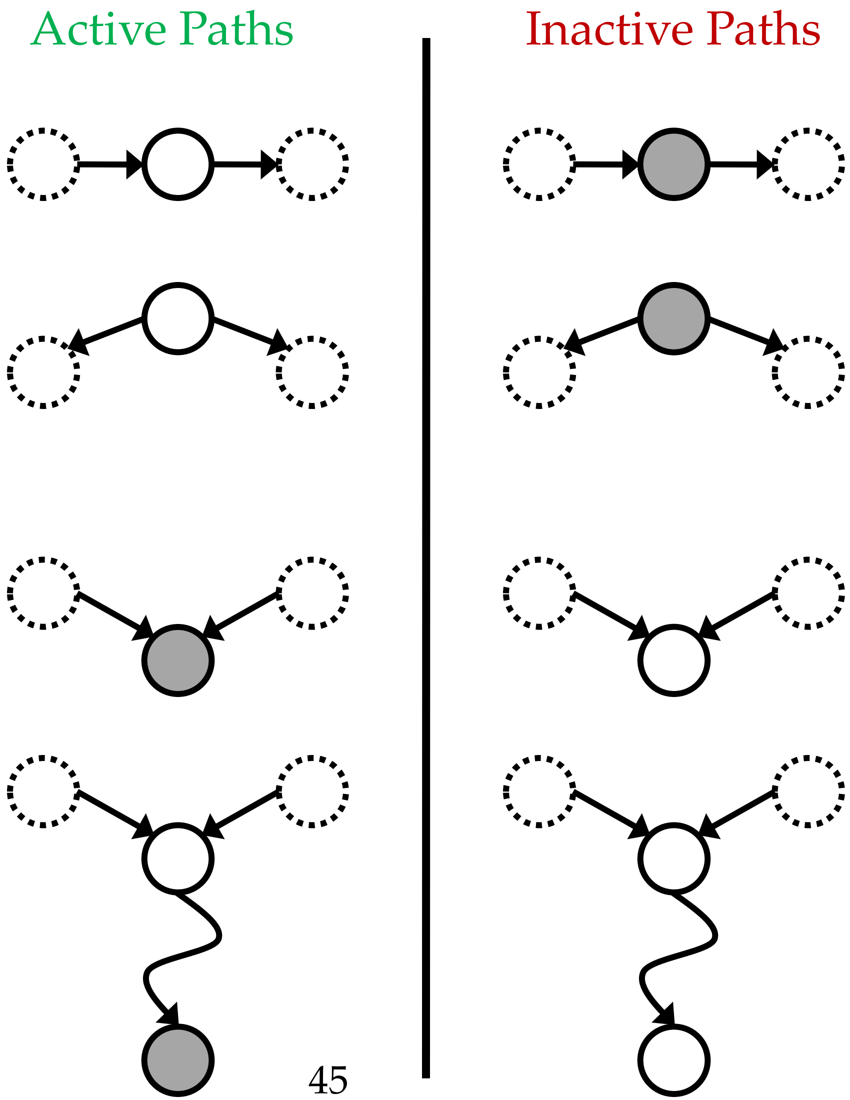
Important note

- We look at all paths along **undirected** edges
- But when going down a path and looking at triplets, we need to look at the **direction** of the edges
- Common cause and common effect induce opposite phenomena: observing parent causes independence, observing child causes dependence

Bayes Ball

- Question: Are X and Y conditionally independent given evidence variables {Z}?

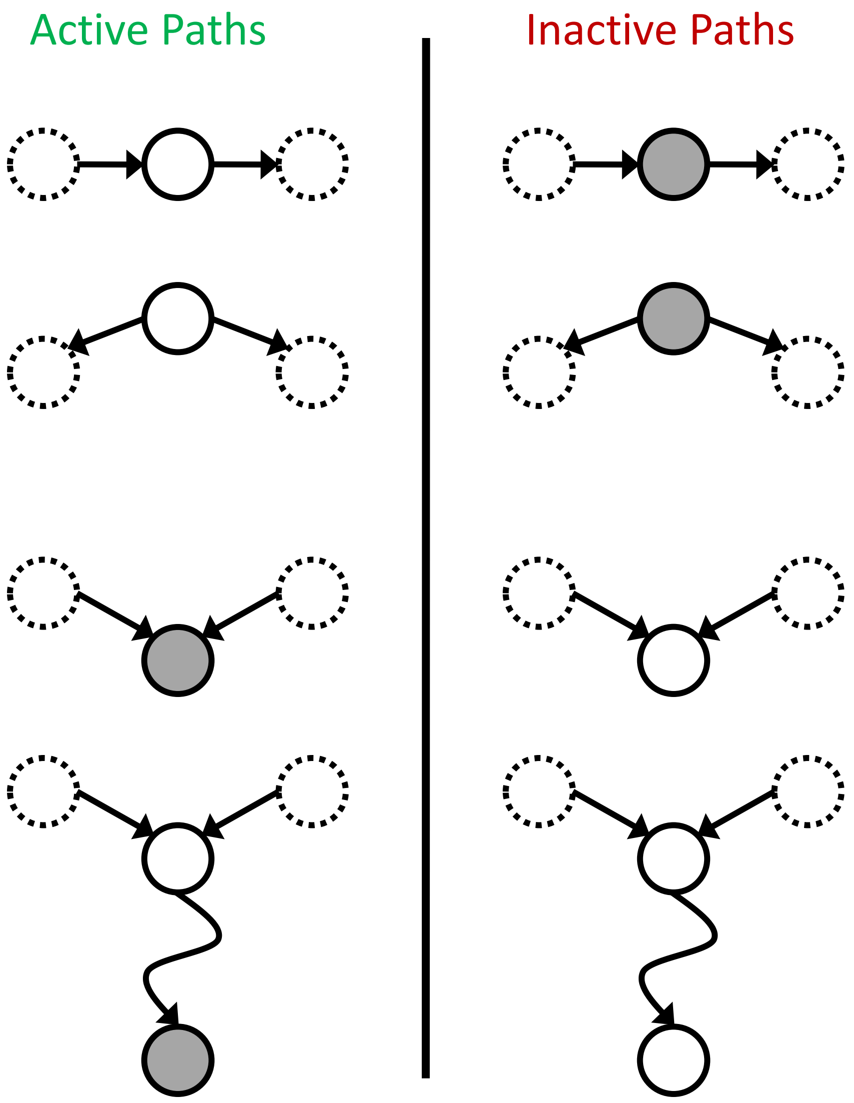
Shachter, Ross D. "Bayes-Ball: Rational Pastime (for Determining Irrelevance and Requisite Information in Belief Networks and Influence Diagrams)." *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence.* 1998.



Bayes Ball

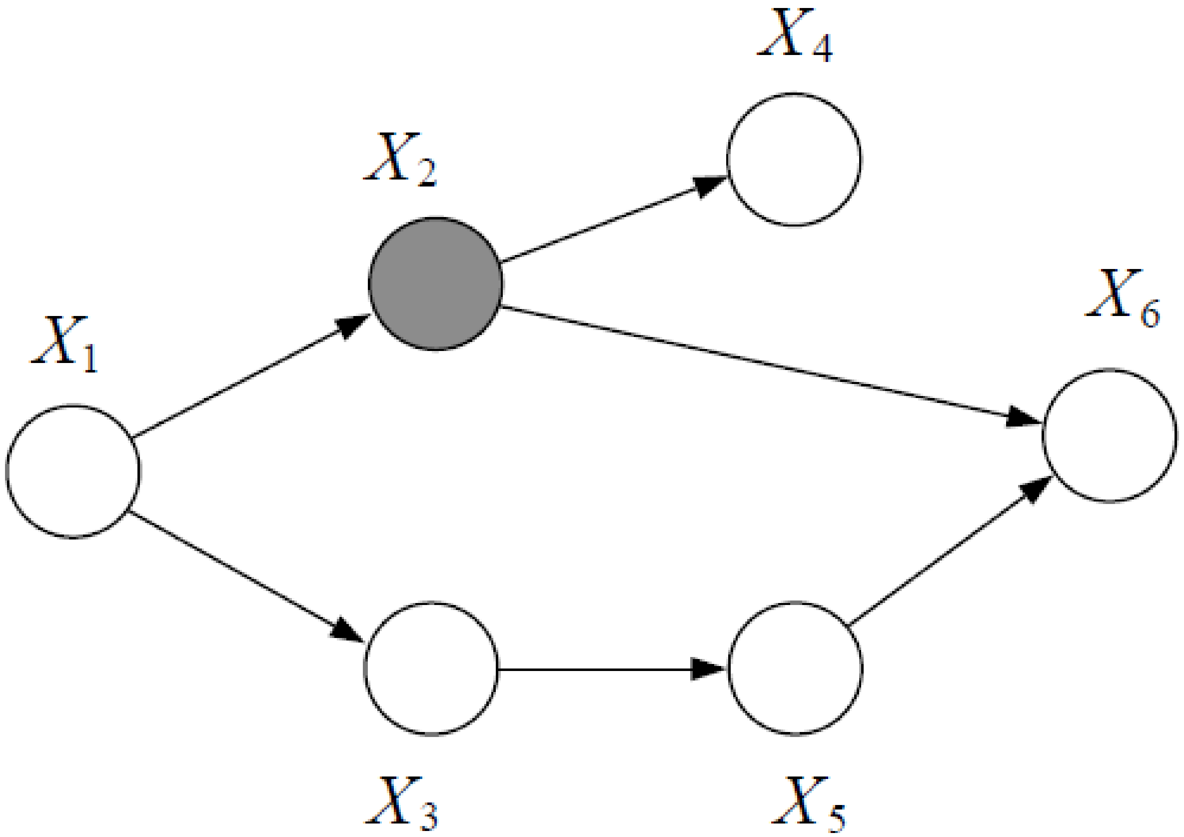
- Question: Are X and Y conditionally independent given evidence variables {Z}?

1. Shade in Z
2. Drop a ball at X
3. The ball can pass through any *active* path and is blocked by any *inactive* path (ball can move either direction on an edge)
4. If the ball can reach Y, then X and Y are NOT conditionally independent given Z



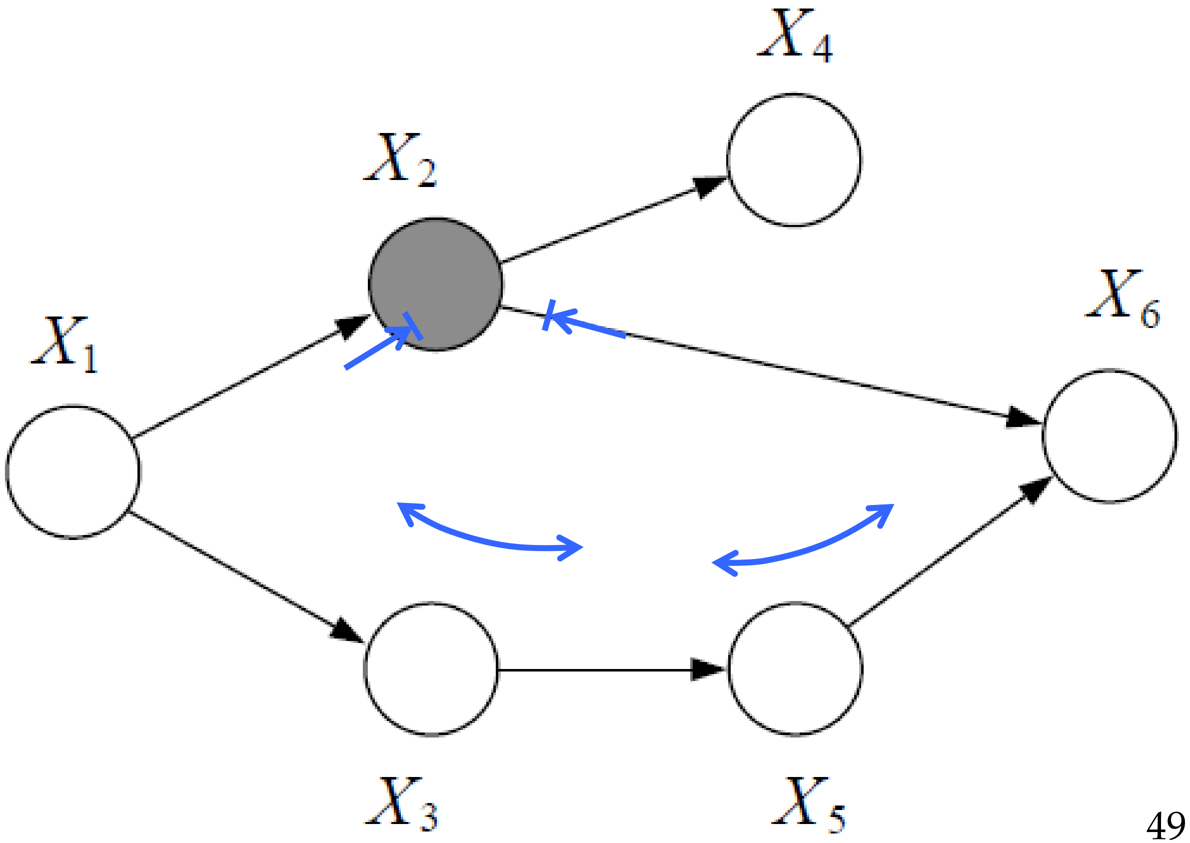
Question

- Is X_1 independent from X_6 given X_2?



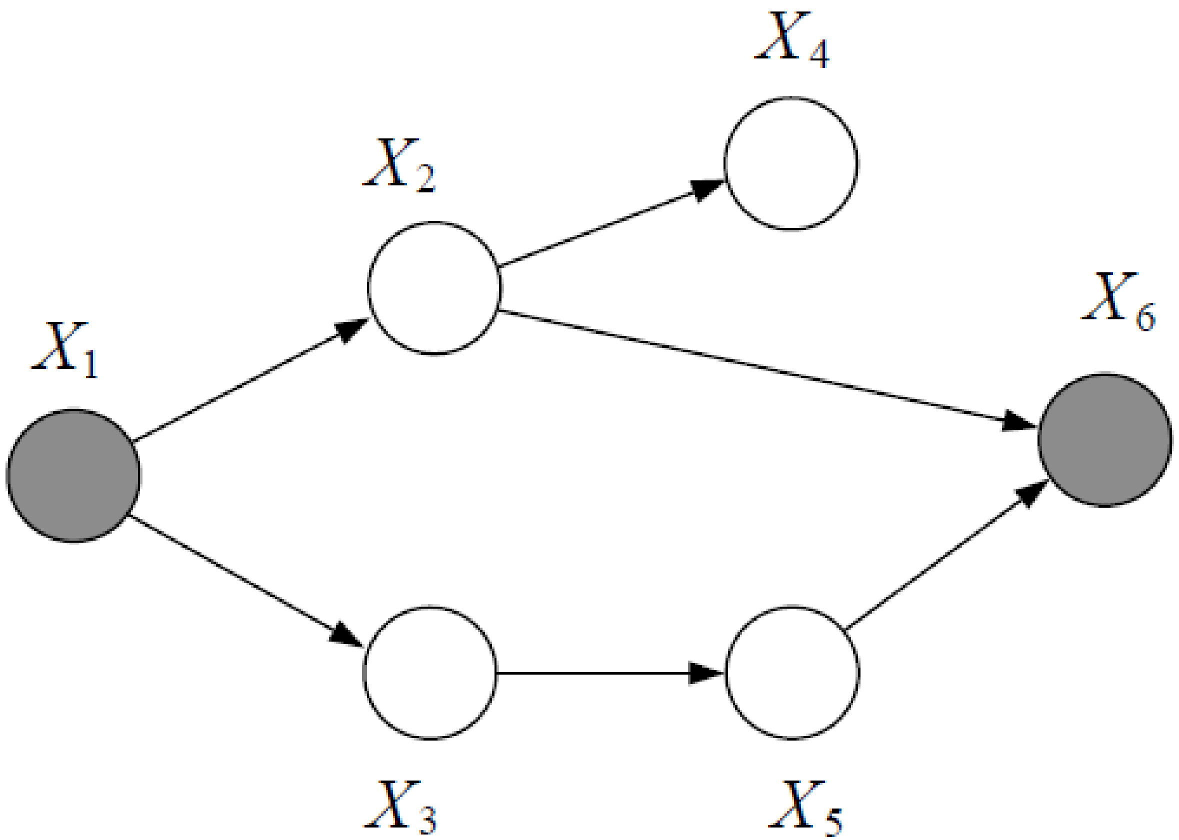
Question

- Is X_1 independent from X_6 given X_2?
- No, the Bayes ball can travel through X_3 and X_5.



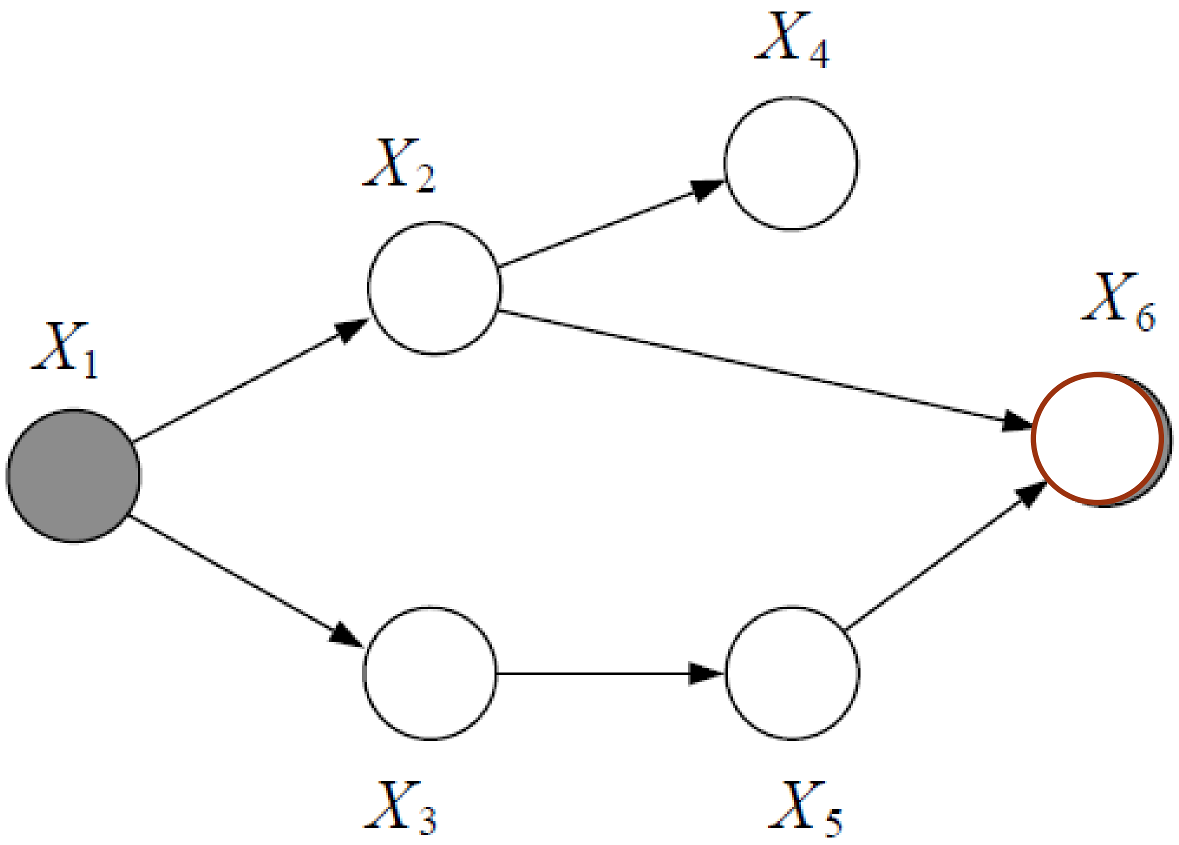
Poll 4

- Is X_2 independent from X_3 given X_1 and X_6?



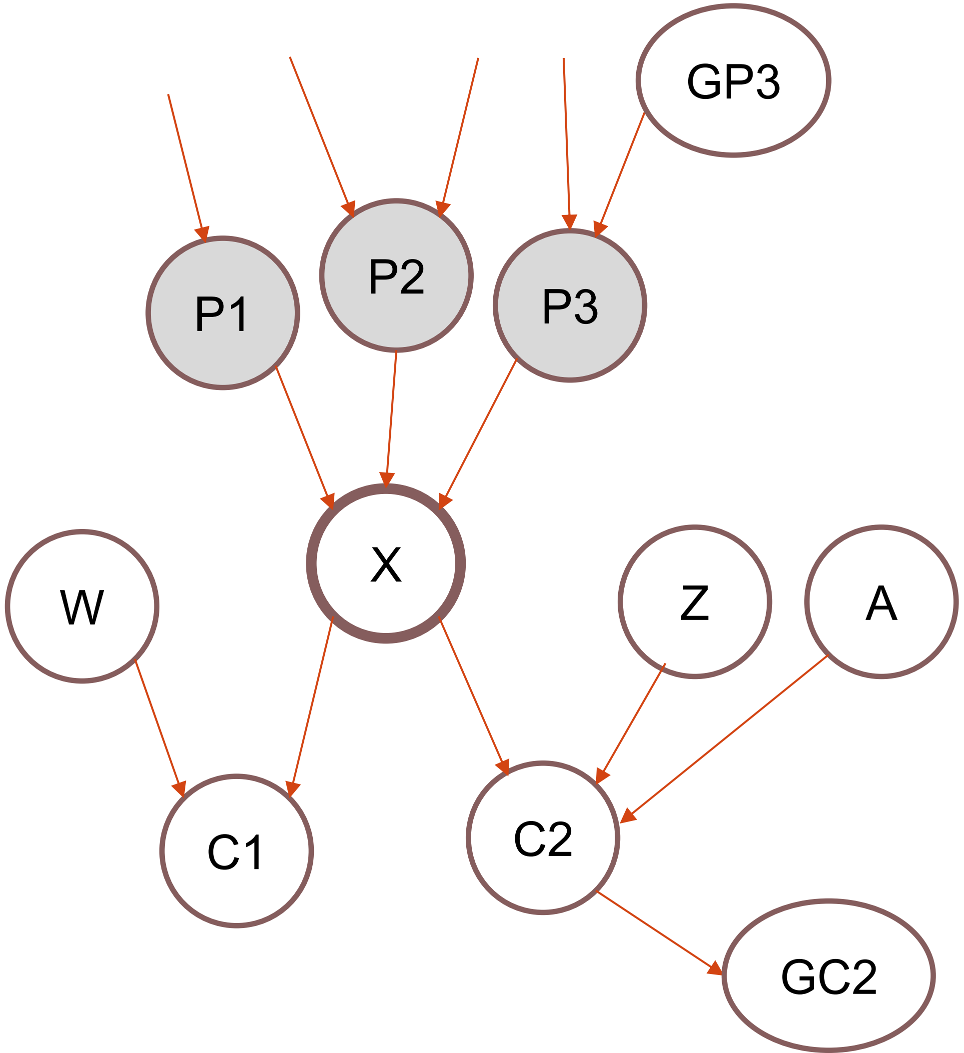
What if?

- Is X_2 independent from X_3 given X_1 and X_6?



Conditional independence semantics

- ***Every variable is conditionally independent of its non-descendants given its parents***

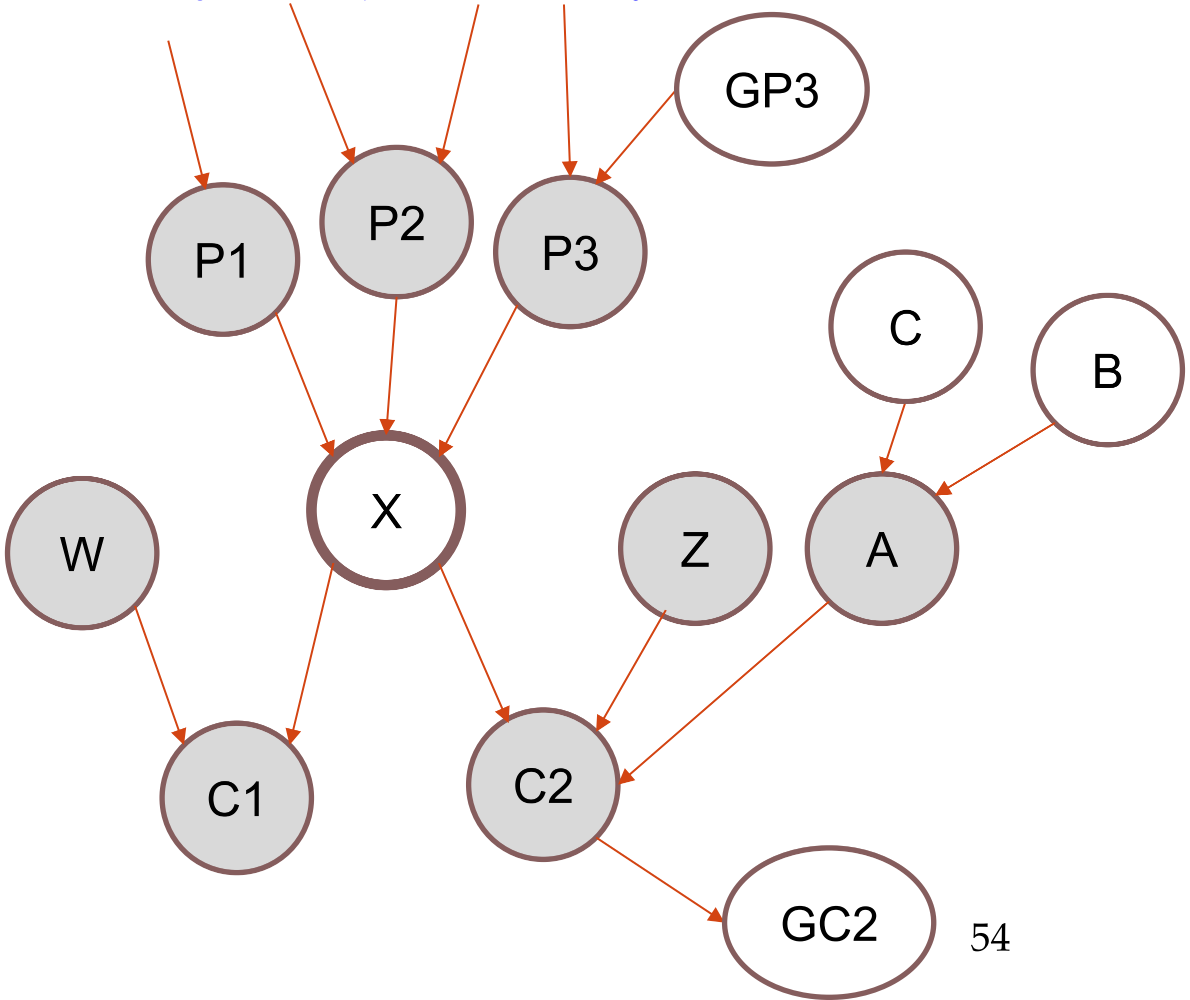


Markov blanket

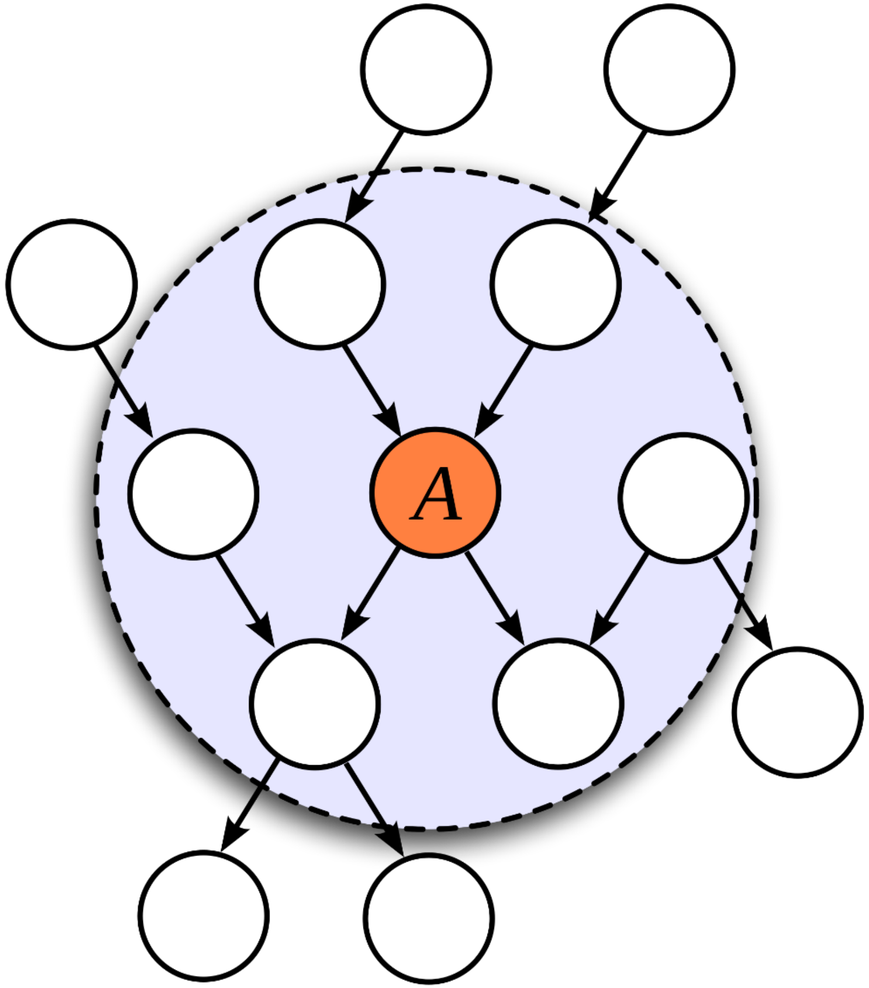
- Markov blanket of X - subset of variables such that all other variables are independent of X conditioned on the blanket

Markov blanket

- A variable's Markov blanket consists of parents, children, children's other parents
- ***Every variable is conditionally independent of all other variables given its Markov blanket***



Markov blanket



Summary – things to know

- Answer queries from joint distribution (marginalization, conditioning)
- Independence, conditional independence implied by a Bayes net
  - "Bayes ball" method
  - Reasoning about causal chain, common cause, common effect