

Announcements

Assignments:

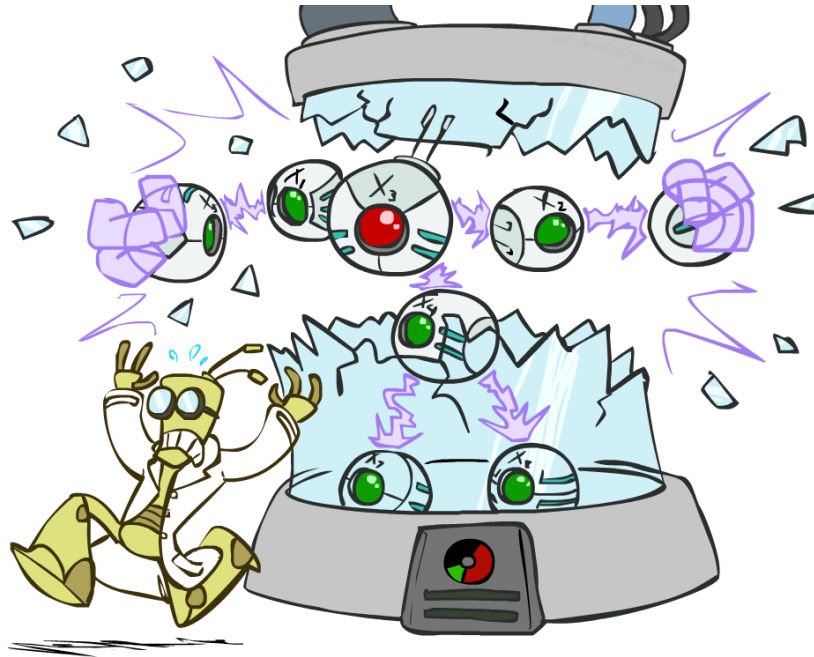
- P4
 - Due tonight, 10 pm
 - Check your slip days!
- HW9 (online)
 - Due Tue 11/5, 10 pm

Midterm:

- Tue 11/12, in-class
 - See Piazza for details

AI: Representation and Problem Solving

Bayes Nets: Independence



Instructors: Pat Virtue & Fei Fang

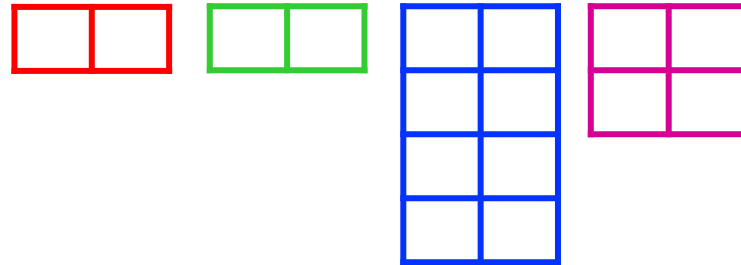
Slide credits: CMU AI and <http://ai.berkeley.edu>

Bayesian Networks

One node per random variable

DAG

One CPT per node: $P(\text{node} \mid \text{Parents}(\text{node}))$

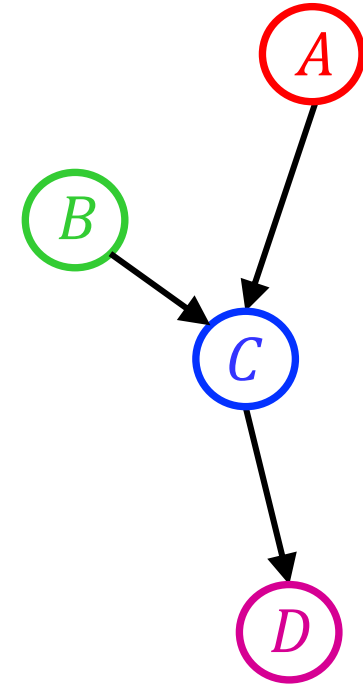


$$P(A, B, C, D) = P(A) P(B) P(C|A, B) P(D|C)$$

Encode joint distributions as product of conditional distributions on each variable

$$P(X_1, \dots, X_N) = \prod_i P(X_i \mid \text{Parents}(X_i))$$

Bayes net



Danielle Belgrave, Microsoft Research



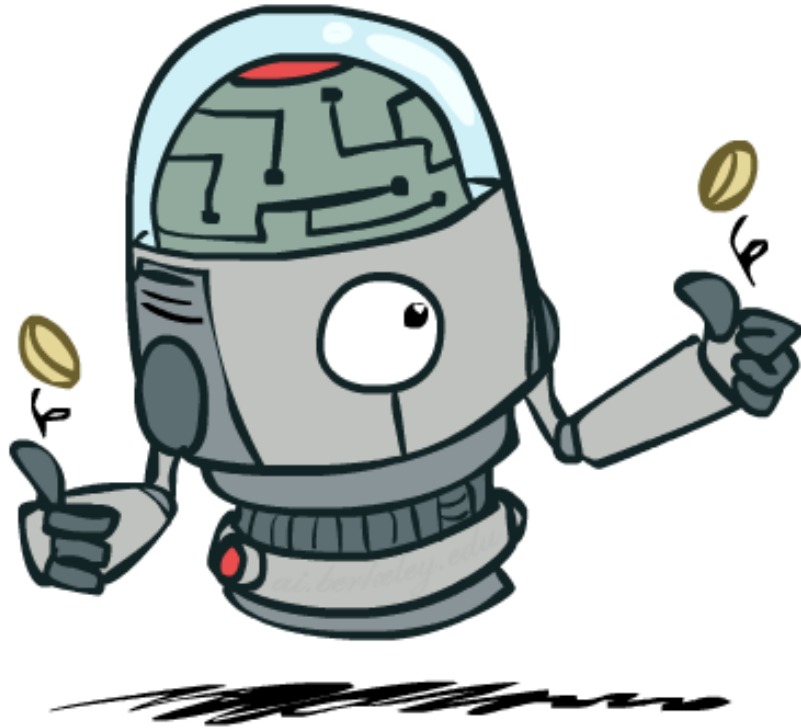
Danielle Belgrave
Principal Researcher

<https://www.microsoft.com/en-us/research/people/dabelgra/>

Developmental Profiles of Eczema, Wheeze, and Rhinitis:
Two Population-Based Birth Cohort Studies
Danielle Belgrave, et al. *PLOS Medicine*, 2014

<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001748>

Independence



Independence

Two variables X and Y are (absolutely) **independent** if

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

- 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:

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

Example: two dice rolls $Roll_1$ and $Roll_2$

- $P(Roll_1=5, Roll_2=5) = P(Roll_1=5) P(Roll_2=5) = 1/6 \times 1/6 = 1/36$
- $P(Roll_2=5 | Roll_1=5) = P(Roll_2=5)$



Example: Independence

n fair, independent coin flips:

$P(X_1)$

H	0.5
T	0.5

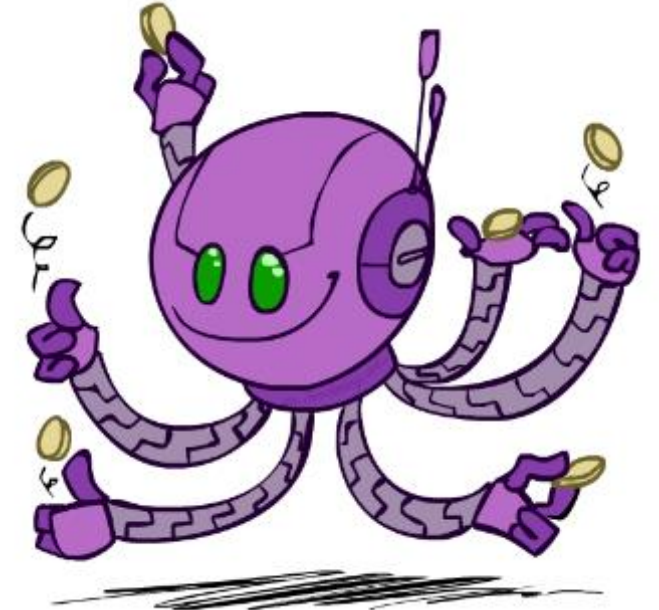
$P(X_2)$

H	0.5
T	0.5

...

$P(X_n)$

H	0.5
T	0.5



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

2^n



Piazza Poll 1

Are T and W independent?

$$P_1(T, W)$$

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

$$P(T)$$

T	P
hot	0.5
cold	0.5

$$P(W)$$

W	P
sun	0.6
rain	0.4

Piazza Poll 1

Are T and W independent?

No

$$P_1(T, W)$$

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

$$P(T)$$

T	P
hot	0.5
cold	0.5

$$P(W)$$

W	P
sun	0.6
rain	0.4

$$P_2(T, W) = P(T)P(W)$$

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

Conditional Independence

$P(\text{Toothache}, \text{Cavity}, \text{Catch})$

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

- $P(+\text{catch} \mid +\text{toothache}, +\text{cavity}) = P(+\text{catch} \mid +\text{cavity})$

The same independence holds if I don't have a cavity:

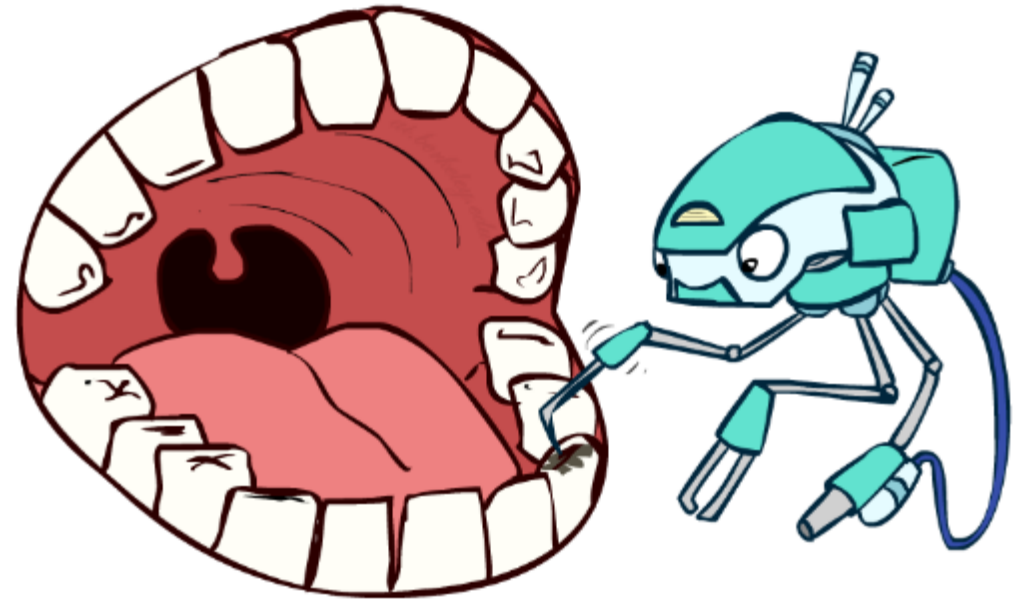
- $P(+\text{catch} \mid +\text{toothache}, -\text{cavity}) = P(+\text{catch} \mid -\text{cavity})$

Catch is *conditionally independent* of Toothache given Cavity:

- $P(\text{Catch} \mid \text{Toothache}, \text{Cavity}) = P(\text{Catch} \mid \text{Cavity})$

Equivalent statements:

- $P(\text{Toothache} \mid \text{Catch}, \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity})$
- $P(\text{Toothache}, \text{Catch} \mid \text{Cavity}) = P(\text{Toothache} \mid \text{Cavity}) P(\text{Catch} \mid \text{Cavity})$
- One can be derived from the other easily



Conditional Independence

Unconditional (absolute) independence very rare (why?)

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

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

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

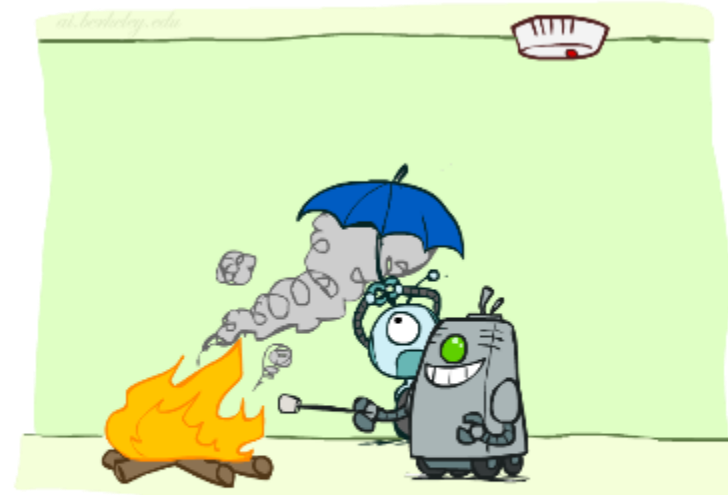
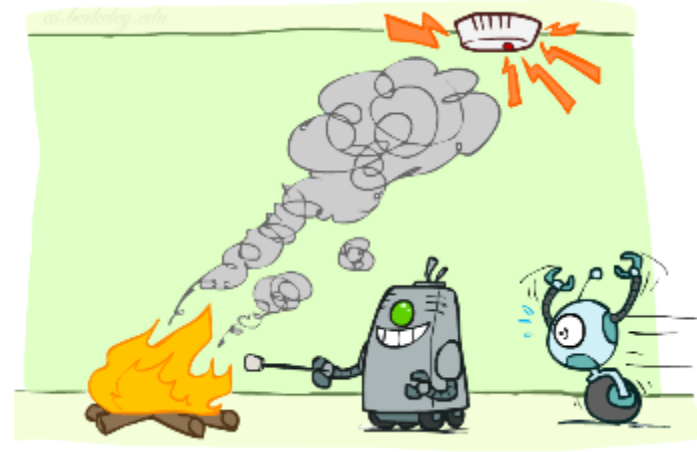
or, equivalently, if and only if

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

Conditional Independence

What about this domain:

- Fire
- Smoke
- Alarm



Conditional Independence

What about this domain:

- Traffic
- Umbrella
- Raining



Conditional Independence and the Chain Rule

Chain rule:

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

Trivial decomposition:

$$P(\text{Rain}, \text{Traffic}, \text{Umbrella}) =$$

With assumption of conditional independence:

$$P(\text{Rain}, \text{Traffic}, \text{Umbrella}) =$$



Conditional Independence and the Chain Rule

Chain rule:

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

Trivial decomposition:

$$P(\text{Rain}, \text{Traffic}, \text{Umbrella}) = P(\text{Rain}) P(\text{Traffic} \mid \text{Rain}) P(\text{Umbrella} \mid \text{Rain}, \text{Traffic})$$

With assumption of conditional independence:

$$P(\text{Rain}, \text{Traffic}, \text{Umbrella}) = P(\text{Rain}) P(\text{Traffic} \mid \text{Rain}) P(\text{Umbrella} \mid \text{Rain})$$

Bayes nets / graphical models help us express conditional independence assumptions



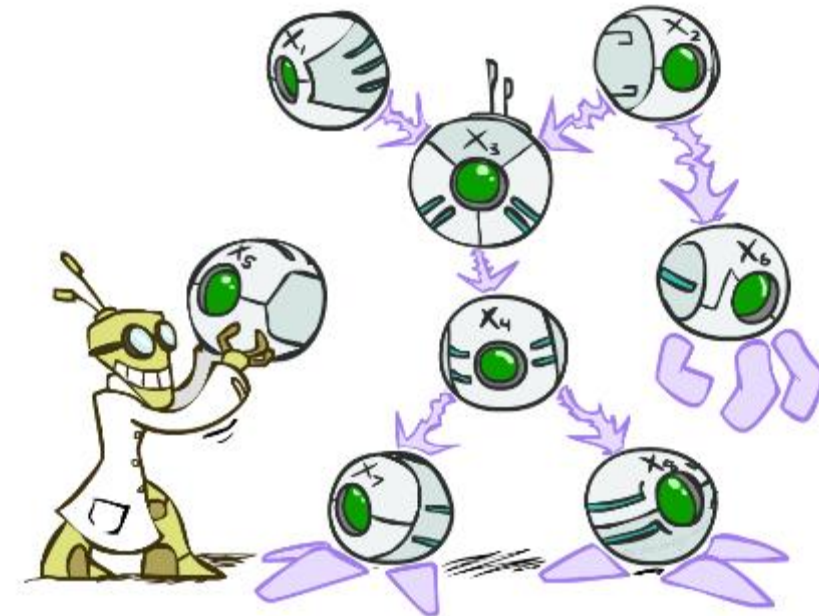
Bayes' Nets: Big Picture

Two problems with using full joint distribution tables as our probabilistic models:

- Unless there are only a few variables, the joint is WAY too big to represent explicitly
- Hard to learn (estimate) anything empirically about more than a few variables at a time

Bayes nets: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)

- More properly called **graphical models**
- We describe how variables locally interact
- Local interactions chain together to give global, indirect interactions



Graphical Model Notation

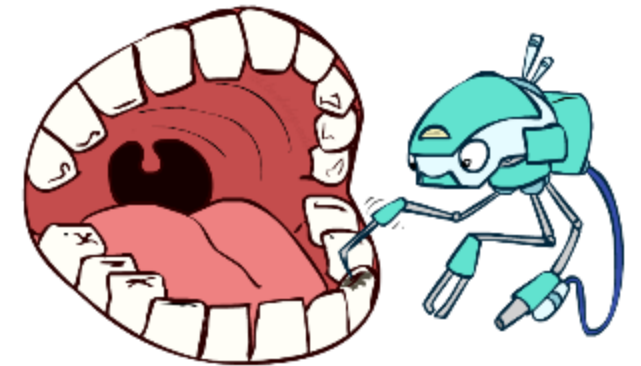
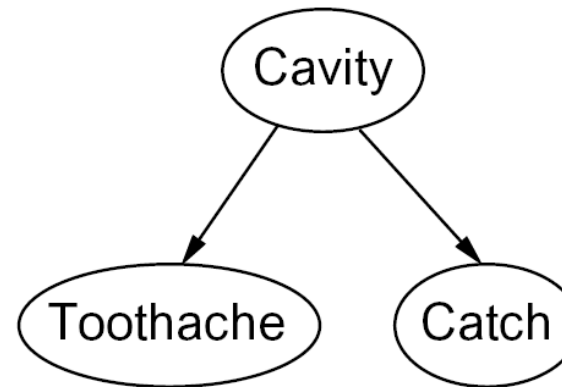
Nodes: variables (with domains)

- Can be assigned (observed) or unassigned (unobserved)

Arcs: interactions

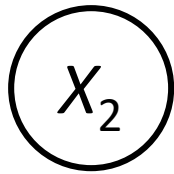
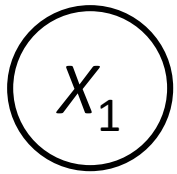
- Similar to CSP constraints
- Indicate “direct influence” between variables
- Formally: encode conditional independence

For now: imagine that arrows mean direct causation (in general, they don't!)

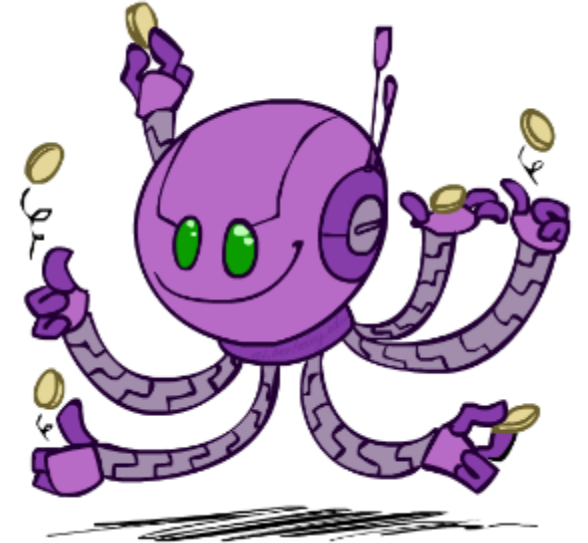
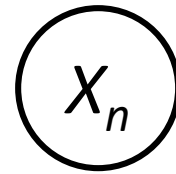


Example: Coin Flips

N independent coin flips



...



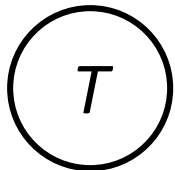
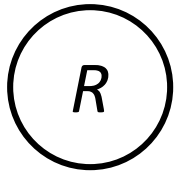
No interactions between variables: **absolute independence**

Example: Traffic

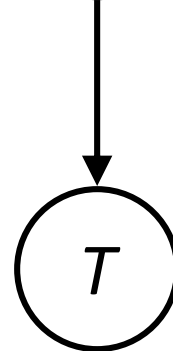
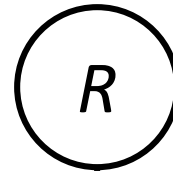
Variables:

- R : It rains
- T : There is traffic

Model 1: independence



Model 2: rain causes traffic



Why is an agent using model 2 better?

Example: Traffic II

Let's build a causal graphical model!

Variables

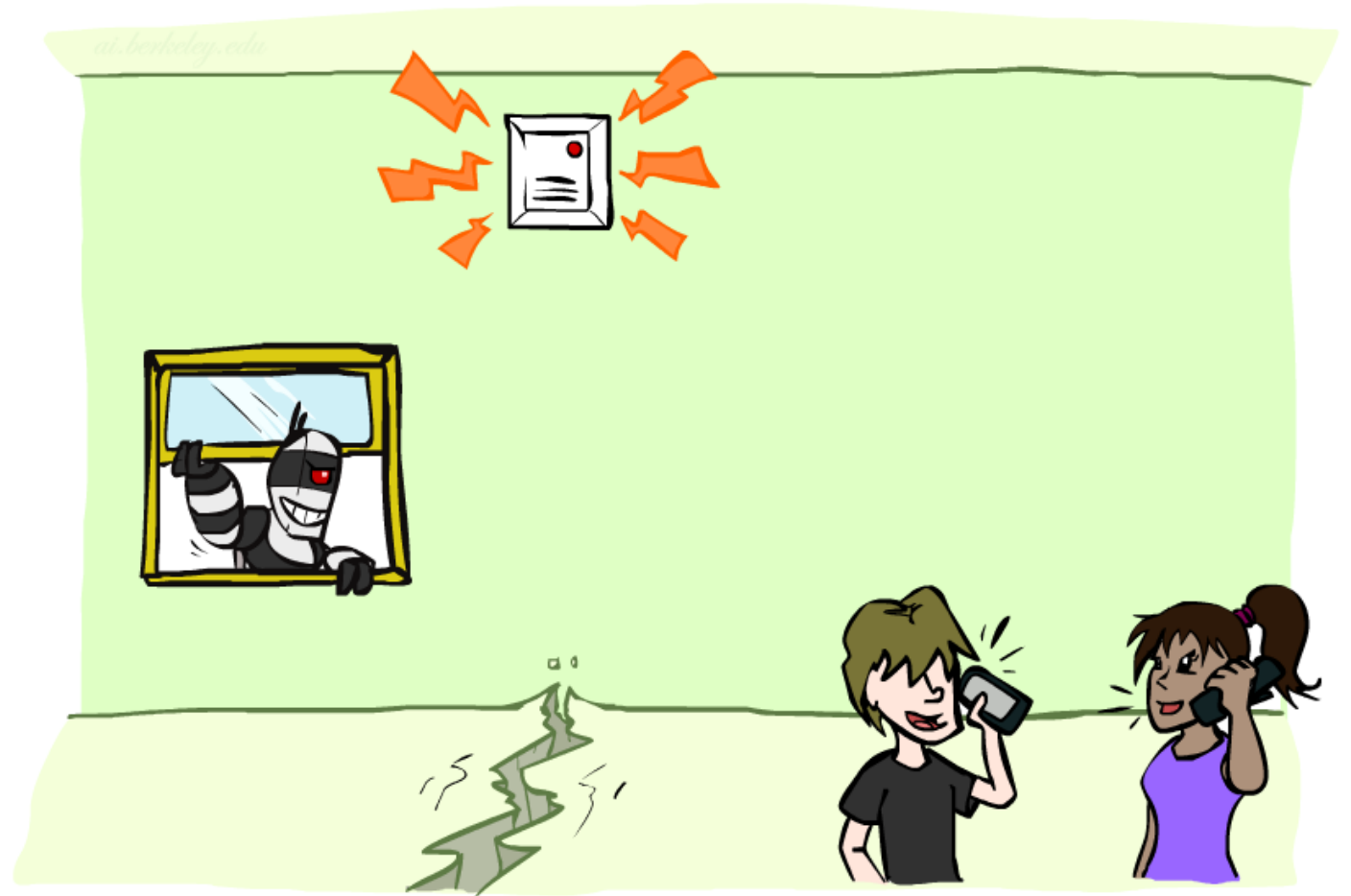
- T: Traffic
- R: It rains
- L: Low pressure
- D: Roof drips
- B: Ballgame
- C: Cavity



Example: Alarm Network

Variables

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



Semantics Example

Joint distribution factorization example

Generic chain rule

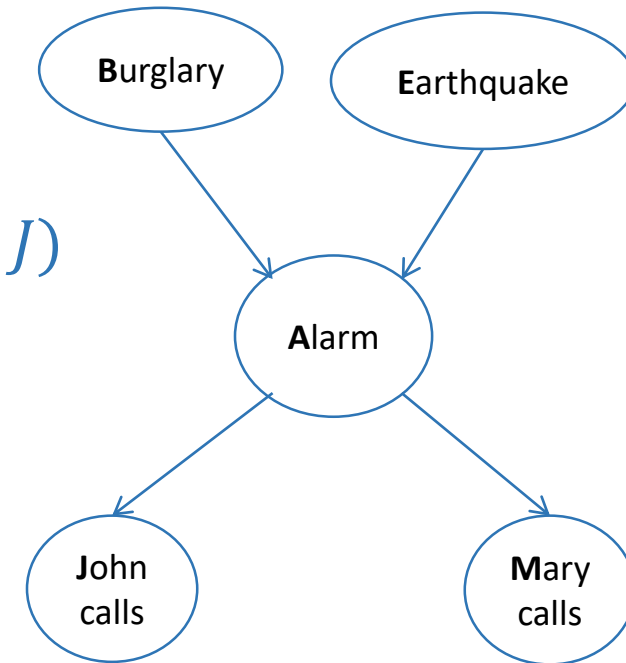
- $P(X_1 \dots X_n) = \prod_i P(X_i | X_1 \dots X_{i-1})$

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

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

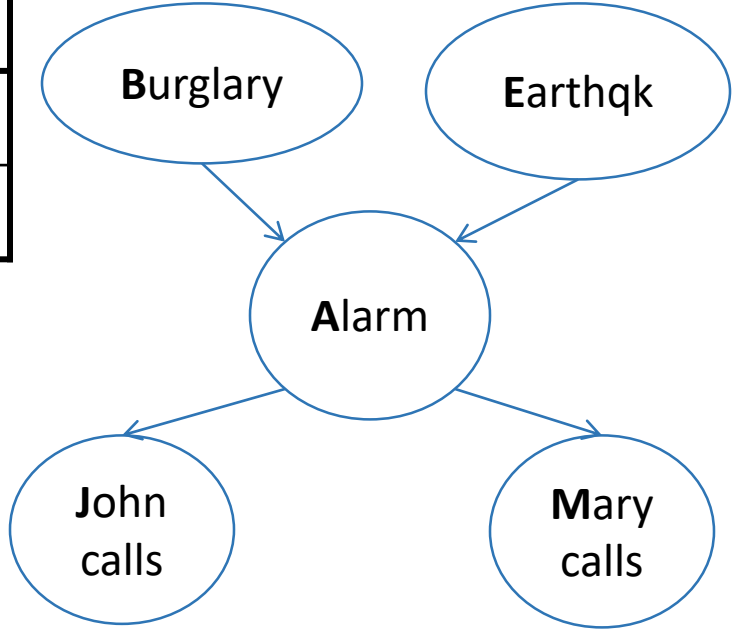
Bayes nets

- $P(X_1 \dots X_n) = \prod_i P(X_i | \text{Parents}(X_i))$



Example: Alarm Network

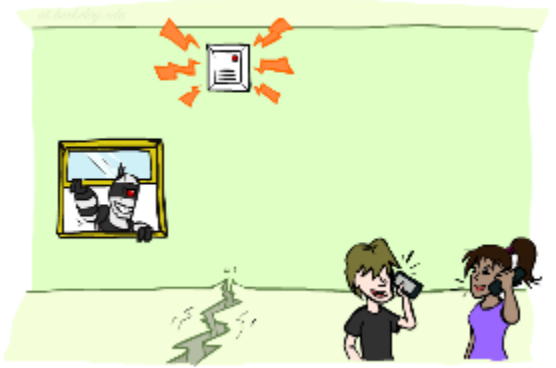
B	P(B)
+b	0.001
-b	0.999



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

E	P(E)
+e	0.002
-e	0.998

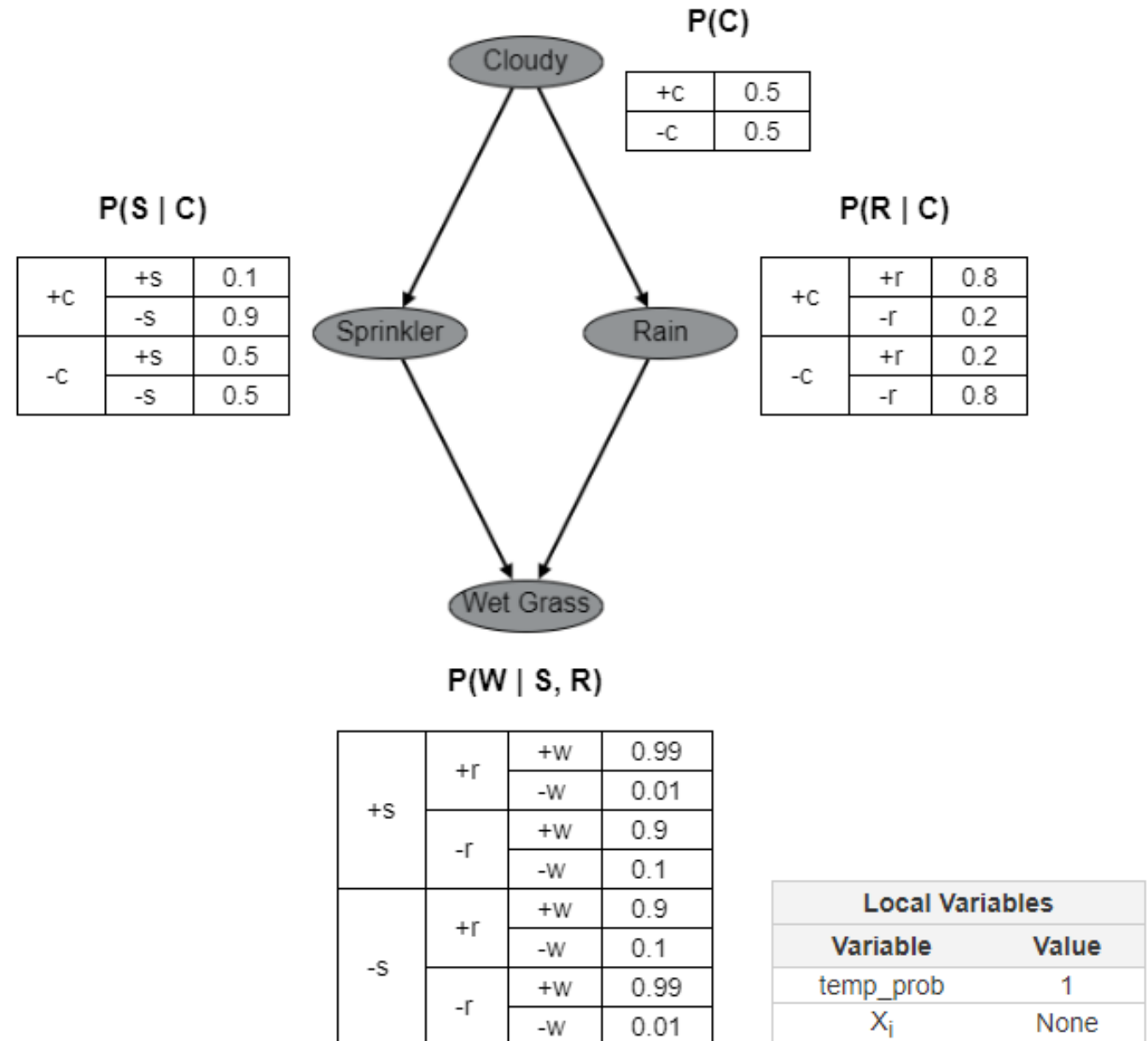


B	E	A	P(A B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

Piazza Poll 2

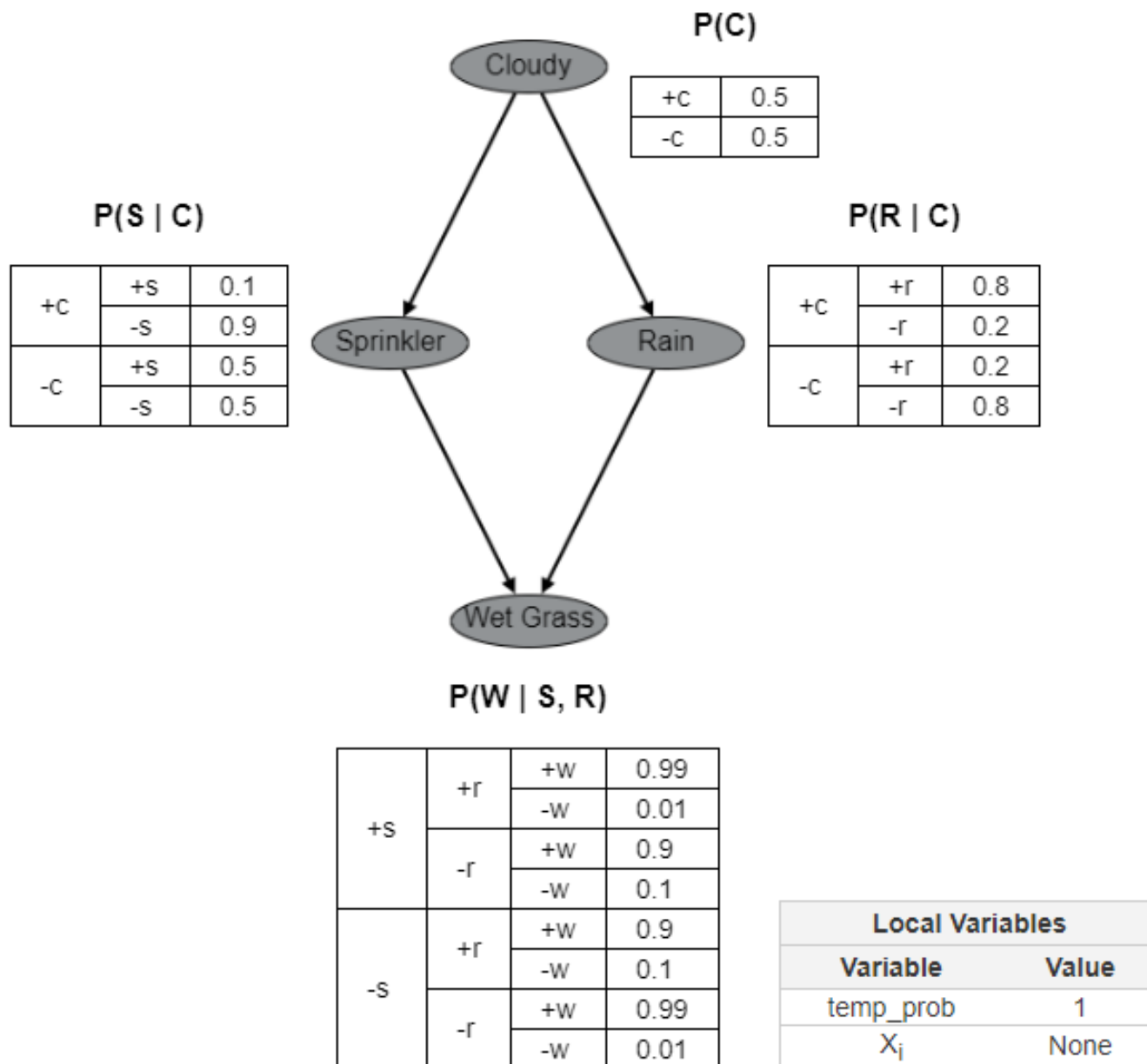
Use the demo from the course website to compute $P(-c, +s, -r, +w)$.

- A. 0.0
- B. 0.0004
- C. 0.001
- D. 0.036
- E. 0.18
- F. 0.198
- G. 0.324



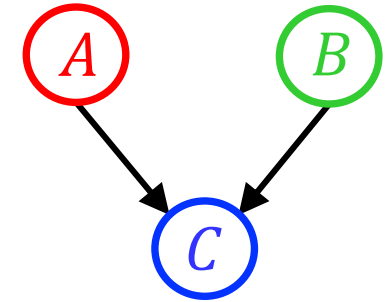
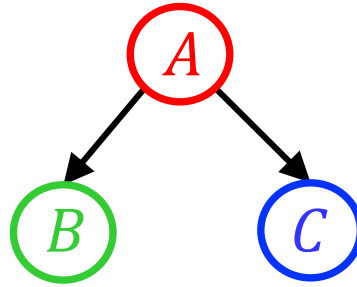
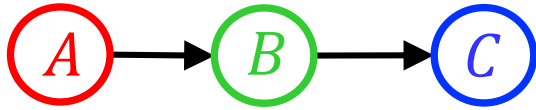
Piazza Poll 2

Use the demo from the course website to compute $P(-c, +s, -r, +w)$.



Piazza Poll 3

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

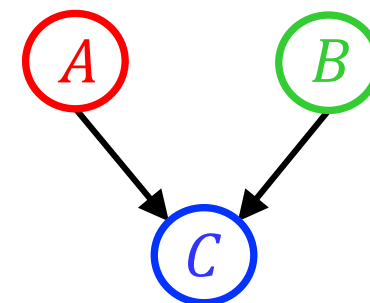
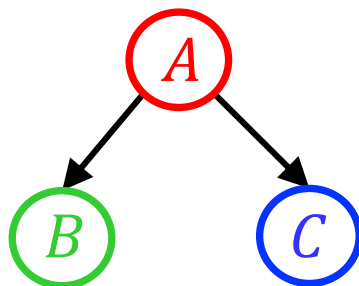
$P(A) P(B) P(C|A, B)$

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

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, 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)



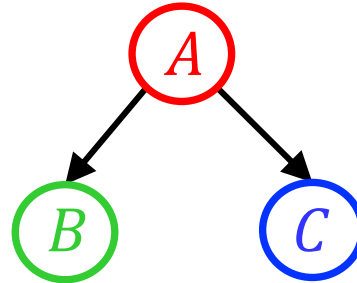
$$P(A) P(B|A) P(C|A, B)$$

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

Assumption:

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

C is independent from A given B



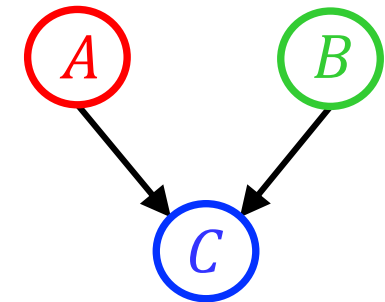
$$P(A) P(B|A) P(C|A, B)$$

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

Assumption:

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

C is independent from B given A



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

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

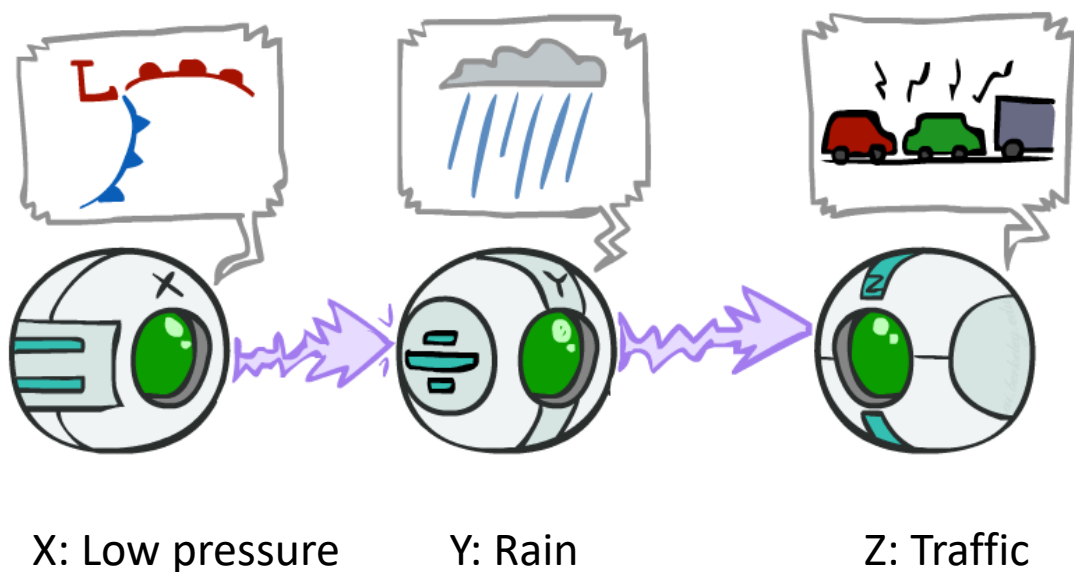
Assumption:

$$P(B|A) = P(B)$$

A is independent from B given { }

Causal Chains

This configuration is a “causal chain”



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

■ 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.
- Example:
 - Low pressure causes rain causes traffic, high pressure causes no rain causes no traffic
 - In numbers:

$$P(+y \mid +x) = 1, P(-y \mid -x) = 1, \\ P(+z \mid +y) = 1, P(-z \mid -y) = 1$$

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?

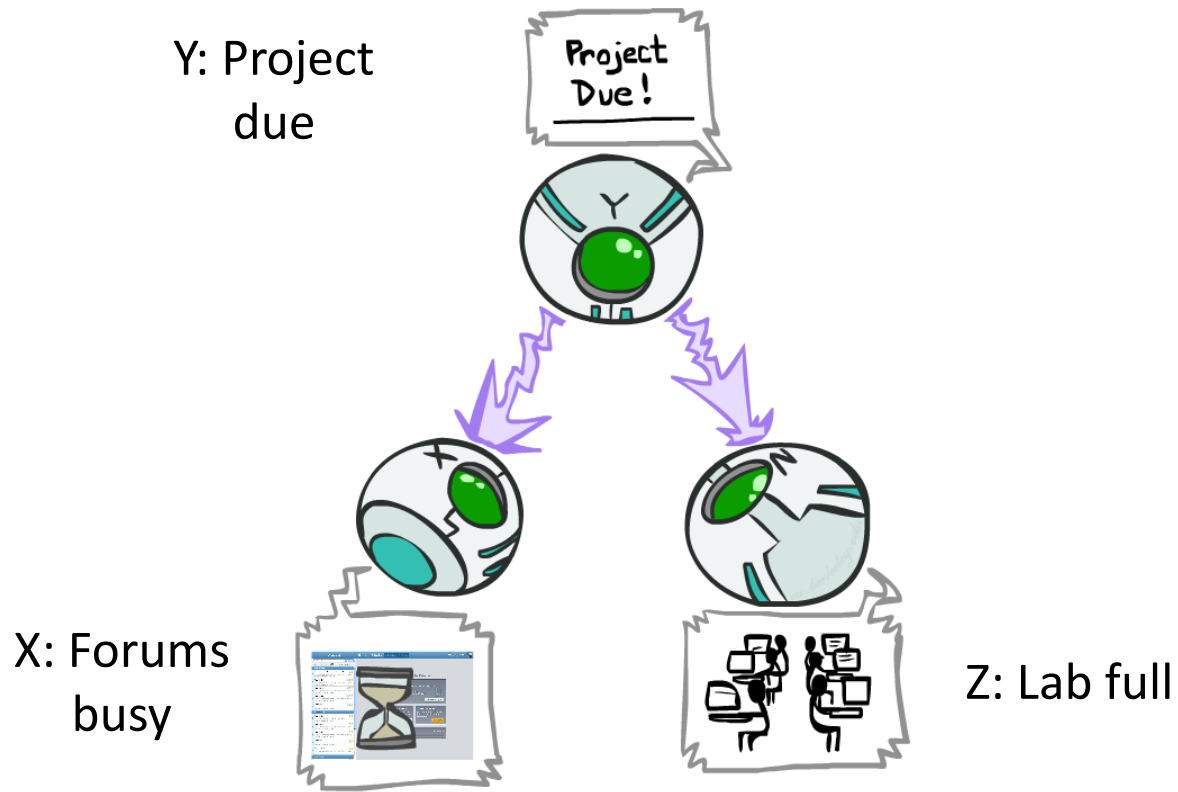
$$\begin{aligned} P(z|x, y) &= \frac{P(x, y, z)}{P(x, y)} \\ &= \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)} \\ &= P(z|y) \end{aligned}$$

Yes!

- Evidence along the chain “blocks” the influence

Common Cause

This configuration is a “common cause”



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

- 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:

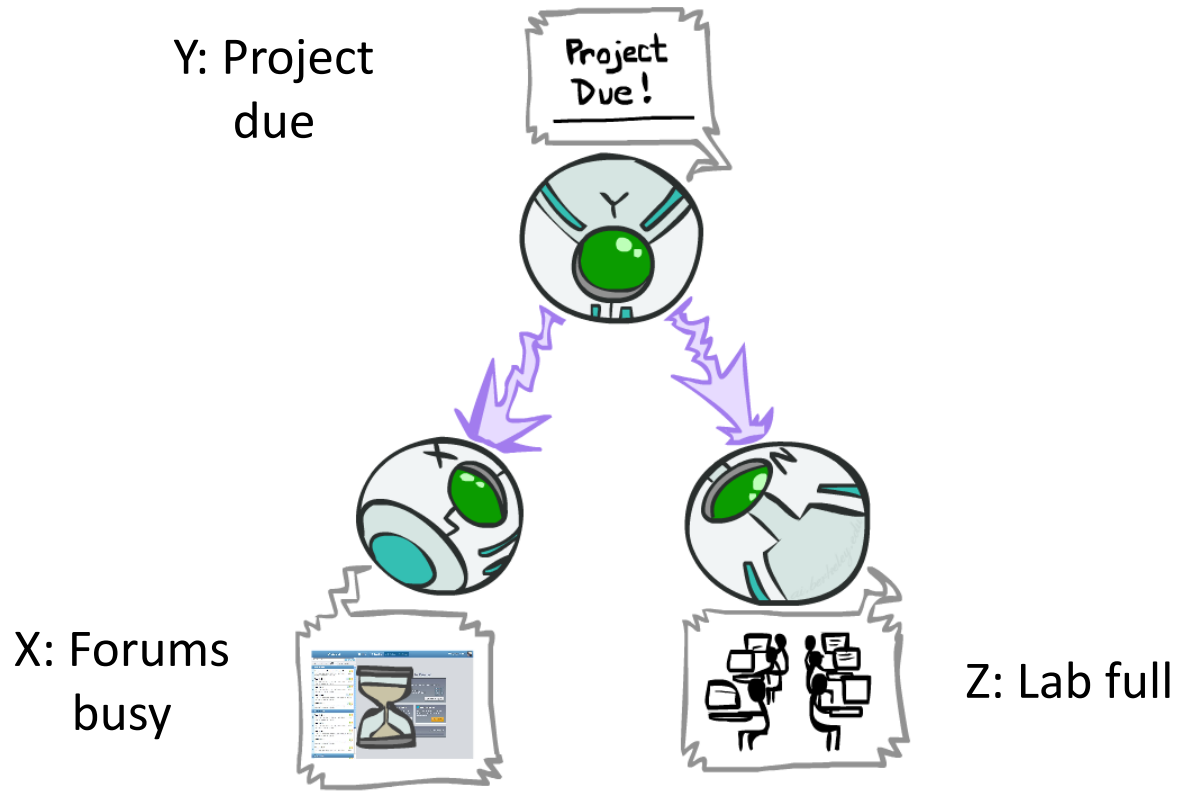
- Project due causes both forums busy and lab full

- In numbers:

$$P(+x \mid +y) = 1, P(-x \mid -y) = 1, \\ P(+z \mid +y) = 1, P(-z \mid -y) = 1$$

Common Cause

This configuration is a “common cause”



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

- Guaranteed X and Z independent given Y?

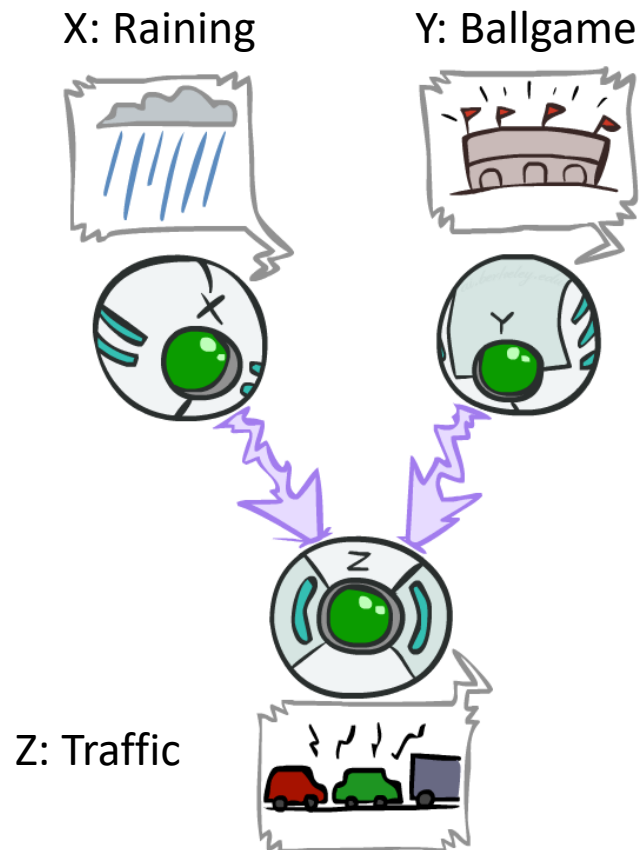
$$\begin{aligned} P(z|x, y) &= \frac{P(x, y, z)}{P(x, y)} \\ &= \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)} \\ &= P(z|y) \end{aligned}$$

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

Bayes Net Independence



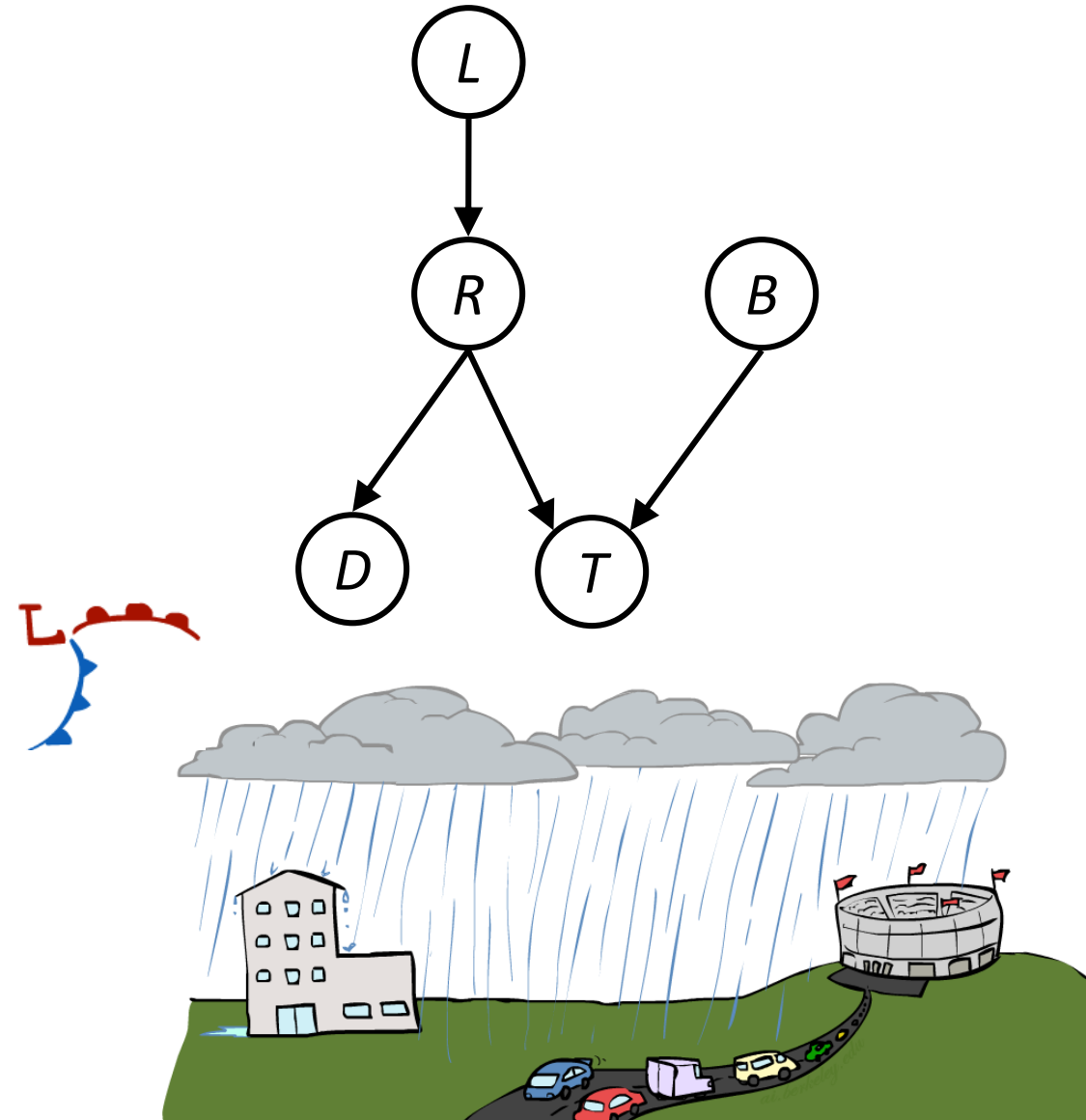
Reachability

Recipe: shade evidence nodes, look for paths in the resulting graph

Attempt 1: if two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent

Almost works, but not quite

- Where does it break?
- Answer: the v-structure at T doesn't count as a link in a path unless "active"



Active / Inactive Paths

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

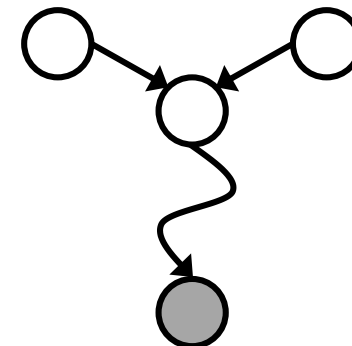
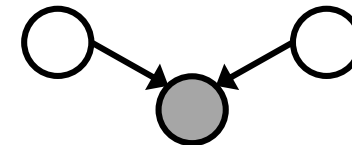
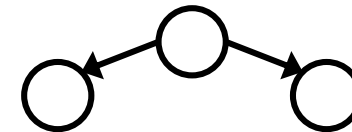
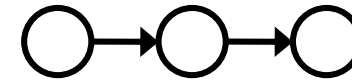
- Yes, if X and Y “d-separated” by Z
- Consider all (undirected) paths from X to Y
- No active paths = independence!

A path is active if each triple is active:

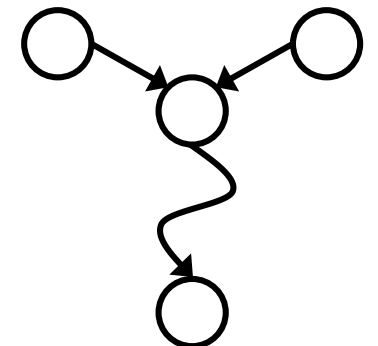
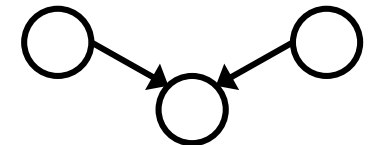
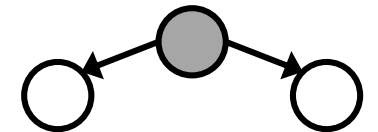
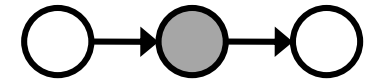
- Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
- Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
- Common effect (aka v-structure)
 $A \rightarrow B \leftarrow C$ where B or one of its descendents is observed

All it takes to block a path is a single inactive segment

Active Triples

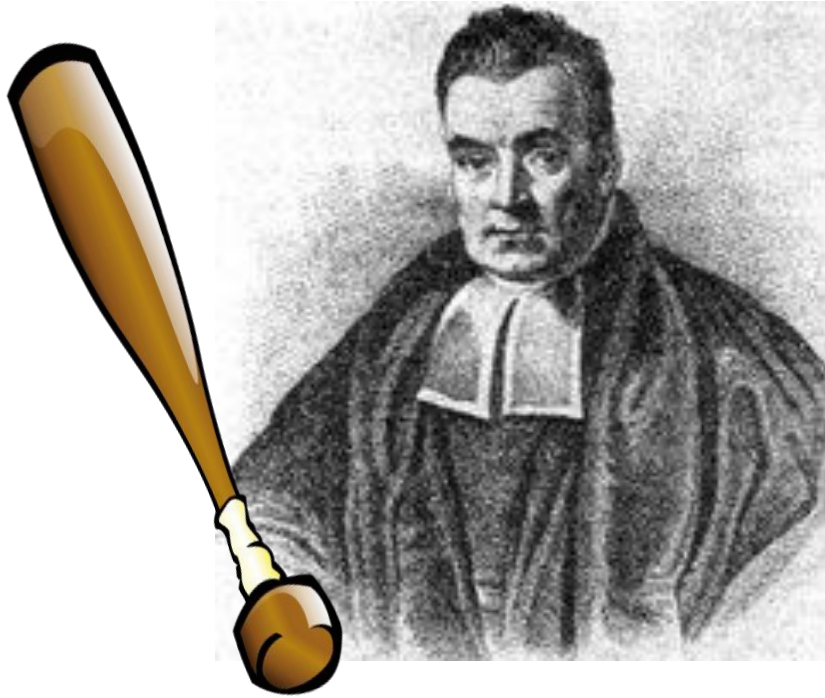


Inactive Triples



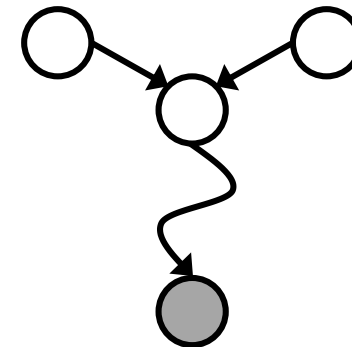
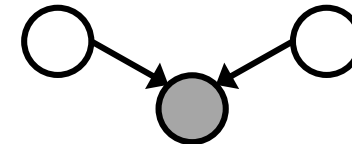
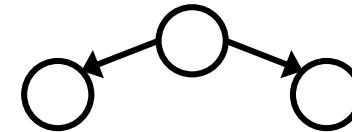
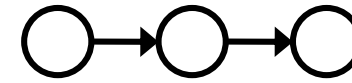
Bayes Ball

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

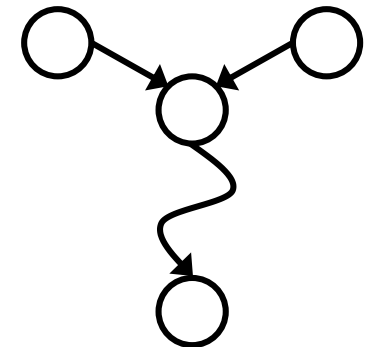
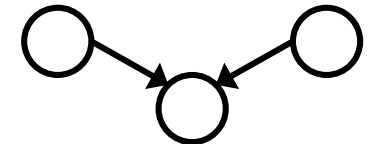
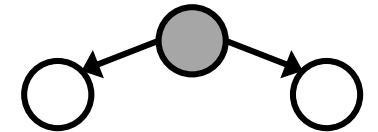
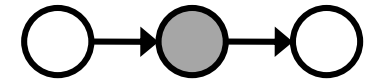


Shachter, Ross D. "Bayes-Ball: Rational Pastime (for Determining Irrelevance and Requisite Information in Belief Networks and Influence Diagrams)." *Proceedings of the Fourteenth conference on Uncertainty in Artificial Intelligence*. 1998.

Active Triples



Inactive Triples

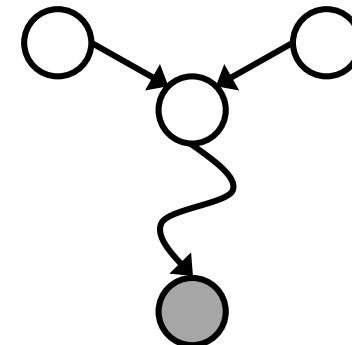
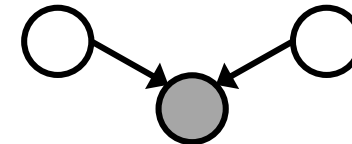
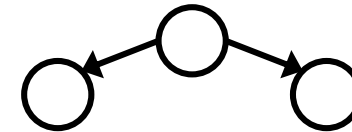
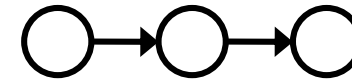


Bayes Ball

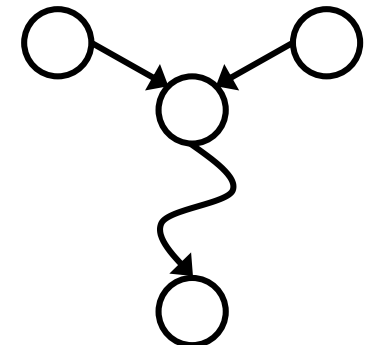
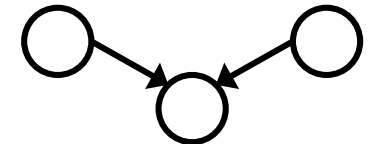
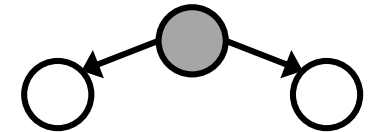
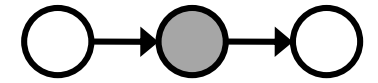
Question: Are X and Y conditionally independent given evidence variables {Z}?

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 reaches Y, then X and Y are NOT conditionally independent given Z

Active Triples

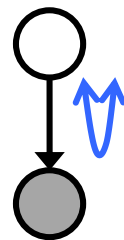
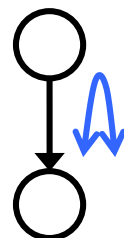
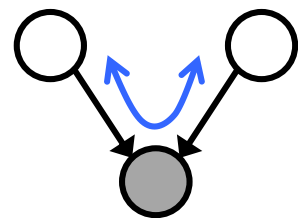
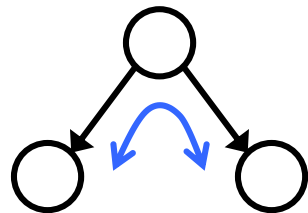
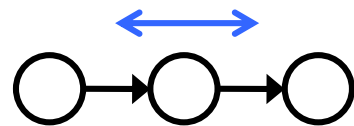


Inactive Triples

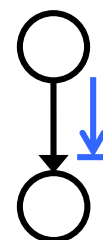
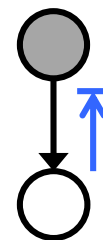
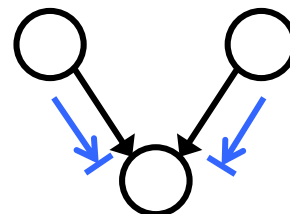
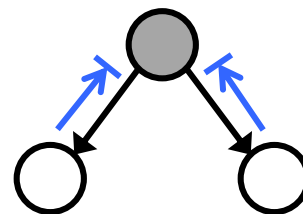
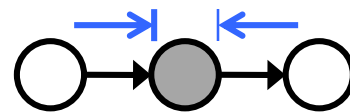


Bayes Ball

Active Paths

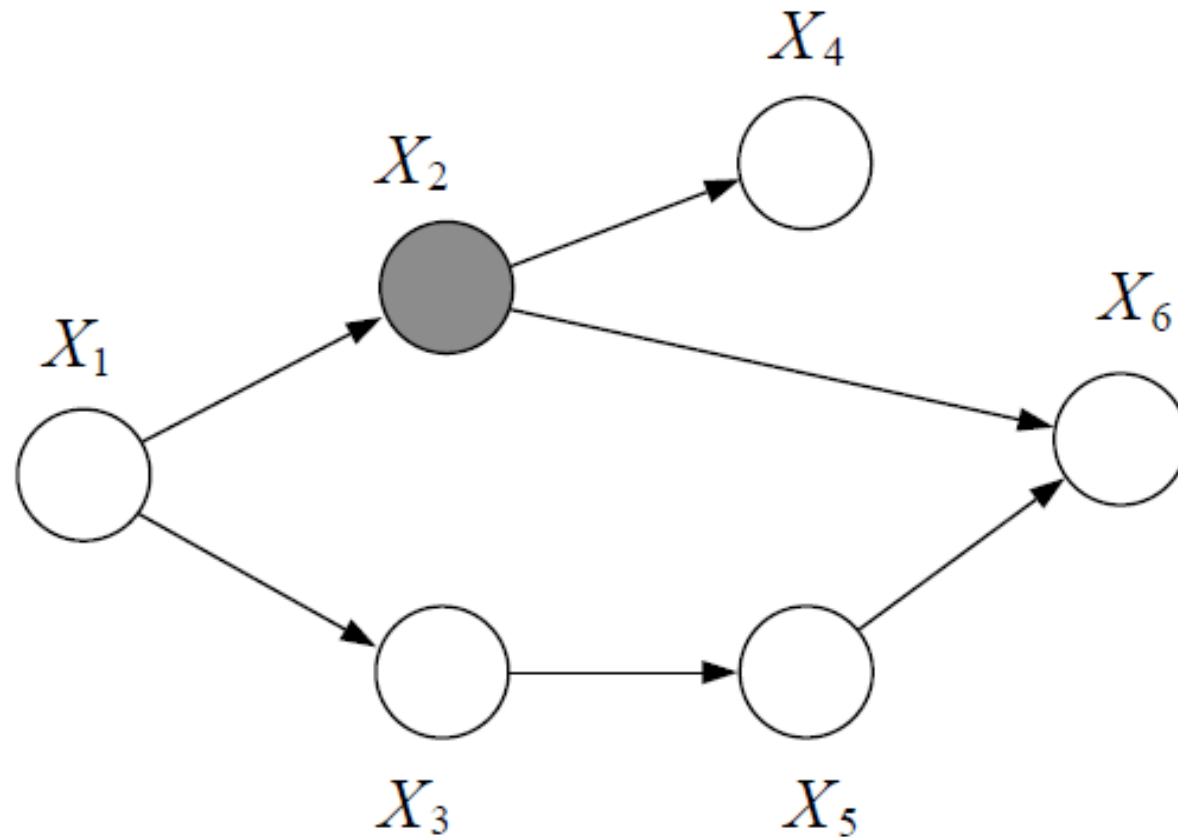


Inactive Paths



Piazza Poll 4

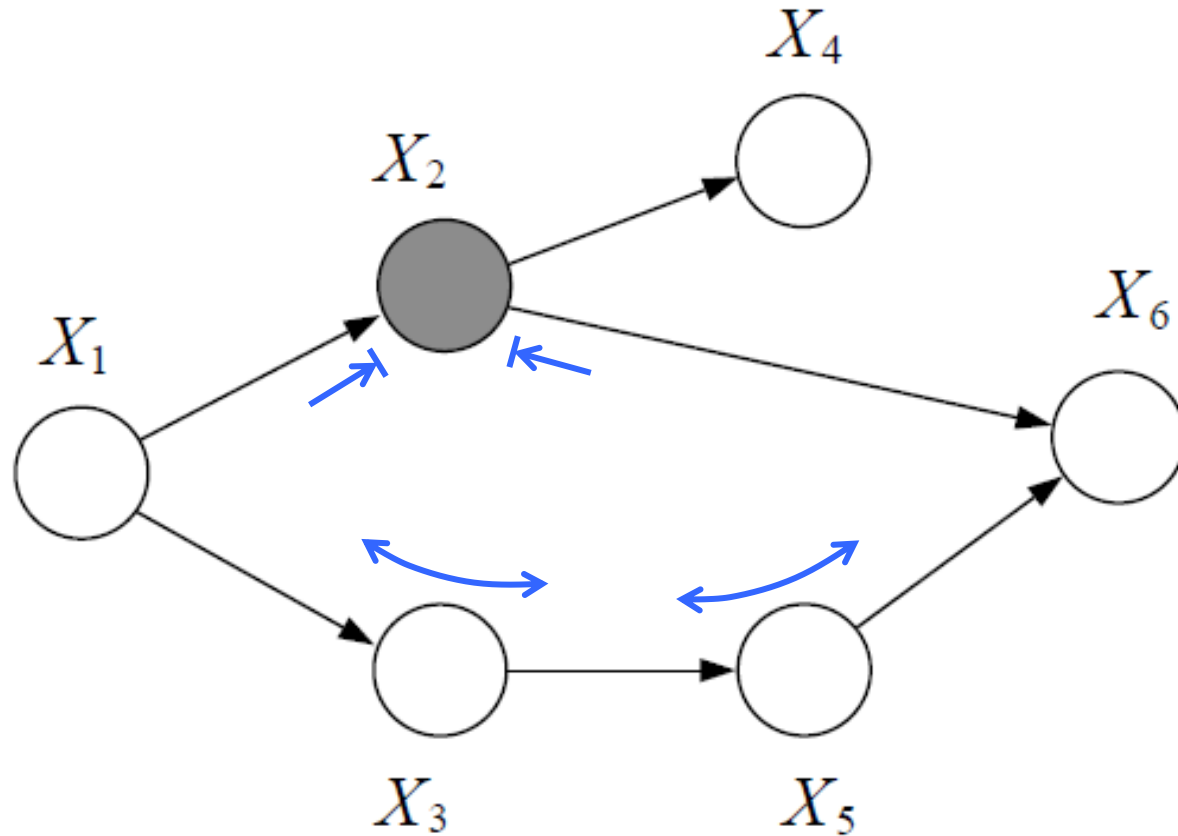
Is X_1 independent from X_6 given X_2 ?



Piazza Poll 4

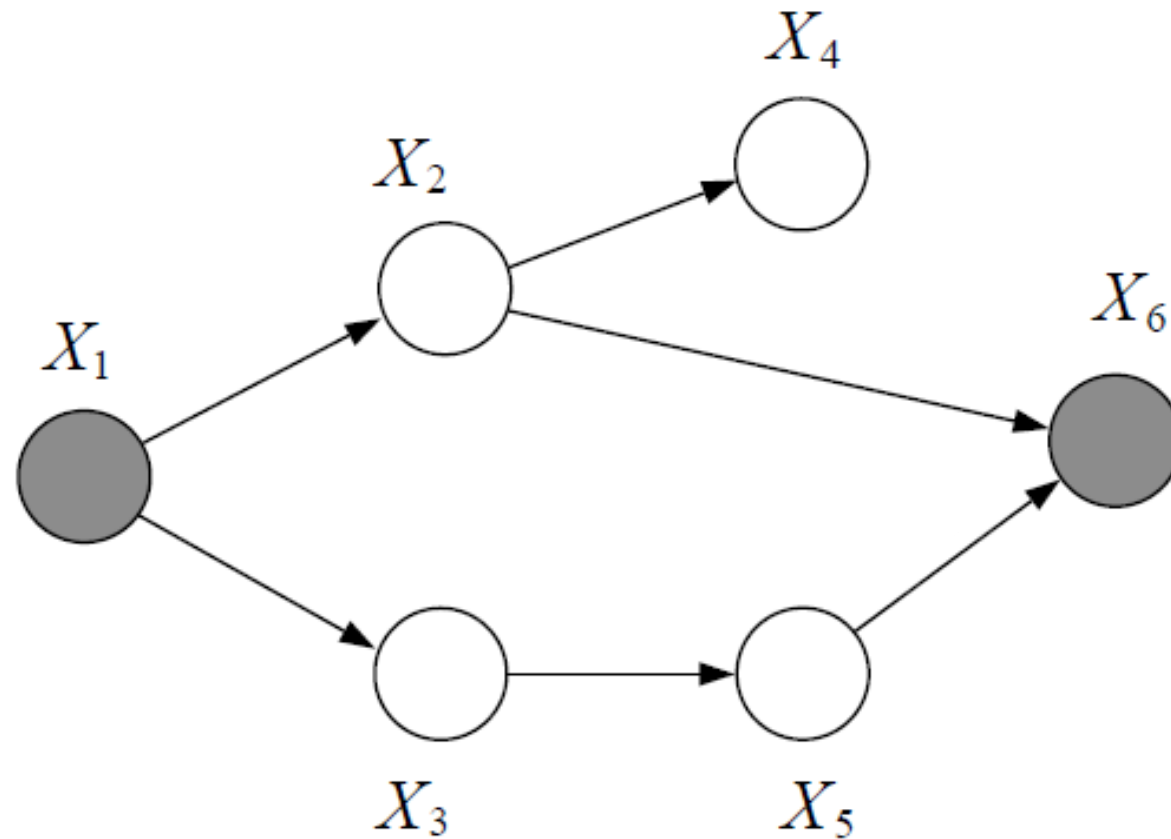
Is X_1 independent from X_6 given X_2 ?

No, the Bayes ball can travel through X_3 and X_5 .



Piazza Poll 5

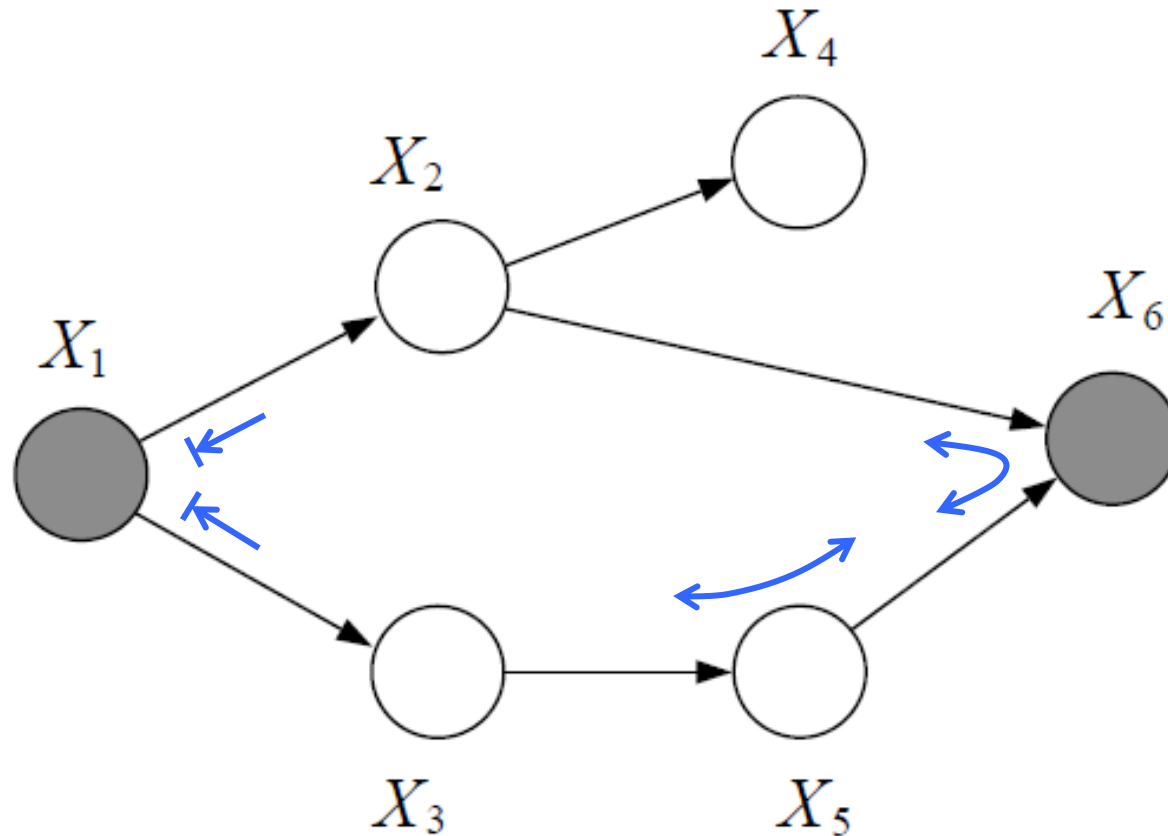
Is X_2 independent from X_3 given X_1 and X_6 ?



Piazza Poll 5

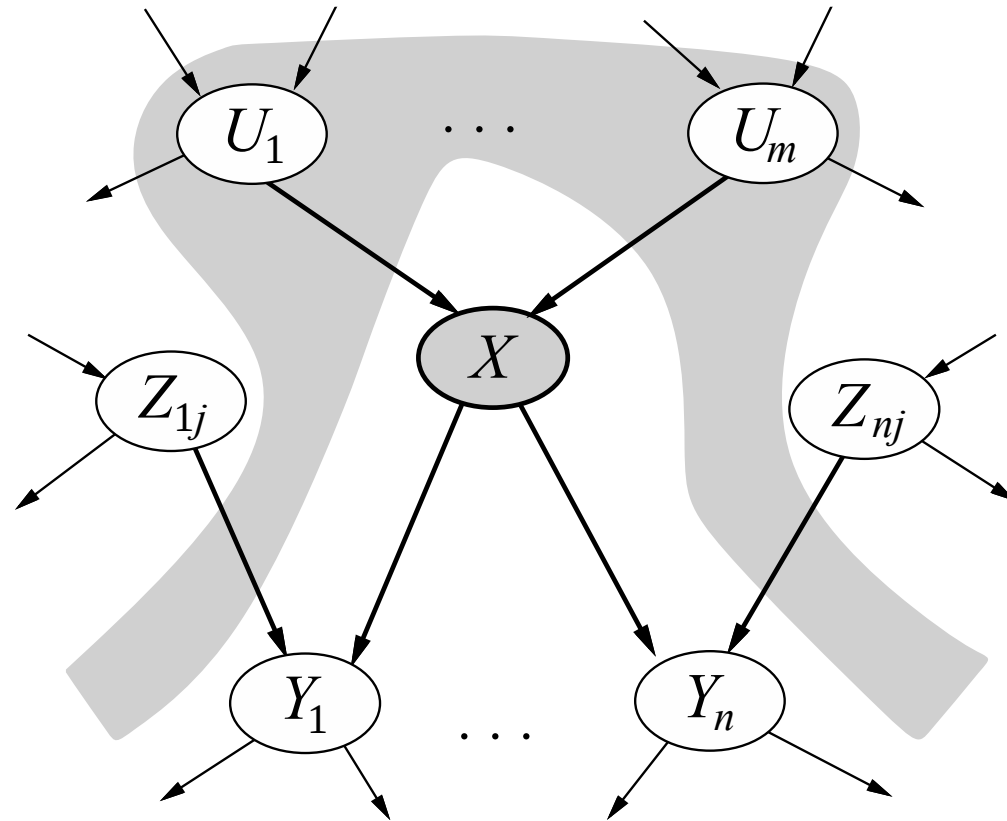
Is X_2 independent from X_3 given X_1 and X_6 ?

No, the Bayes ball can travel through X_5 and X_6 .



Conditional Independence Semantics

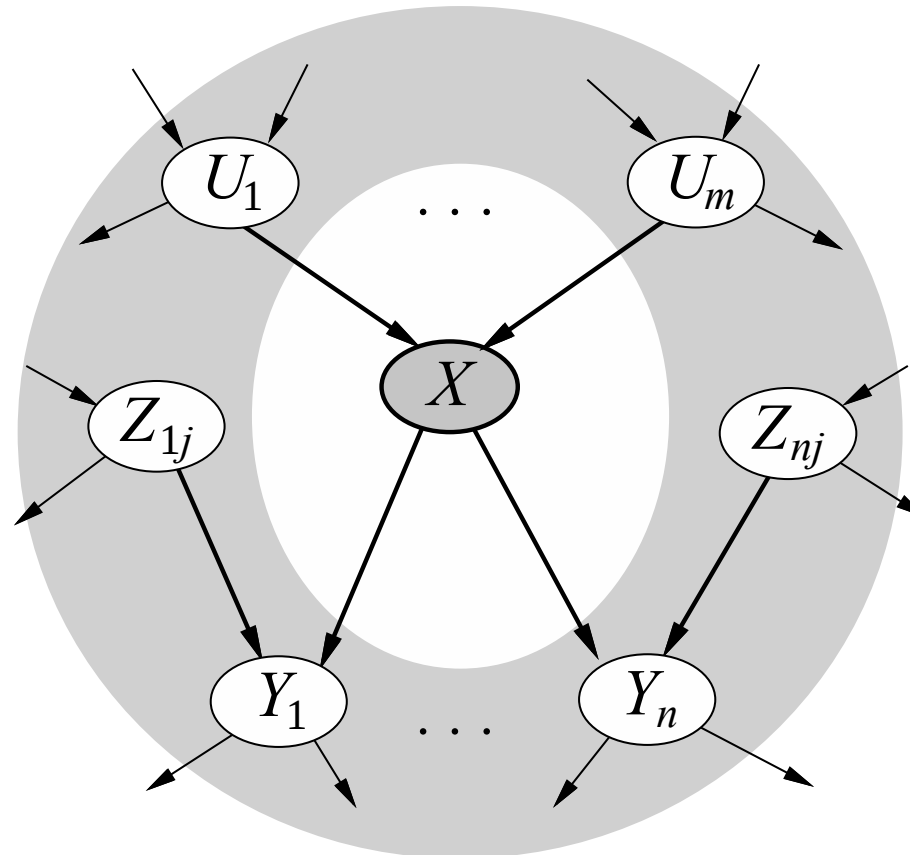
Every variable is conditionally independent of its non-descendants given its parents



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



Answer Any Query from Joint Distribution

Joint distributions are the best!

Problems with joints

- We aren't given the joint table
 - Usually some set of conditional probability tables
- Huge
 - n variables with d values
 - d^n entries

Joint

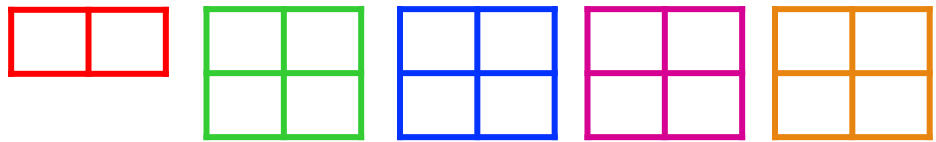
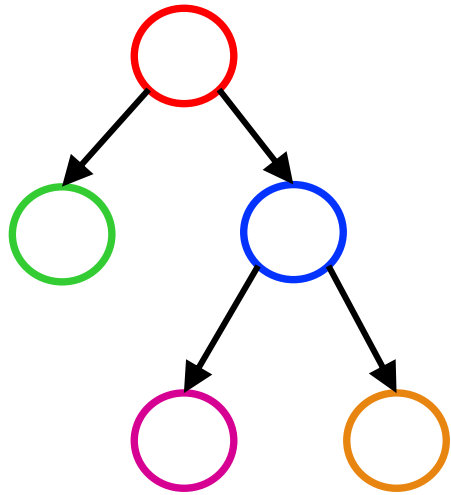


Query

$$P(a \mid e)$$

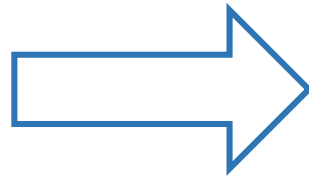
Answer Any Query from Bayes Net

Bayes Net



$P(A)$ $P(B|A)$ $P(C|A)$ $P(D|C)$ $P(E|C)$

Joint



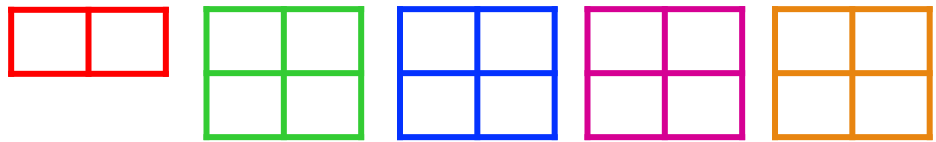
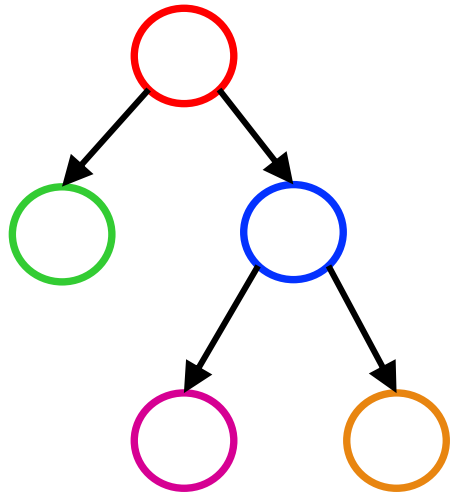


Query

$P(a | e)$

Next: Answer Any Query from Bayes Net

Bayes Net



$P(A)$ $P(B|A)$ $P(C|A)$ $P(D|C)$ $P(E|C)$

Query



$P(a | e)$