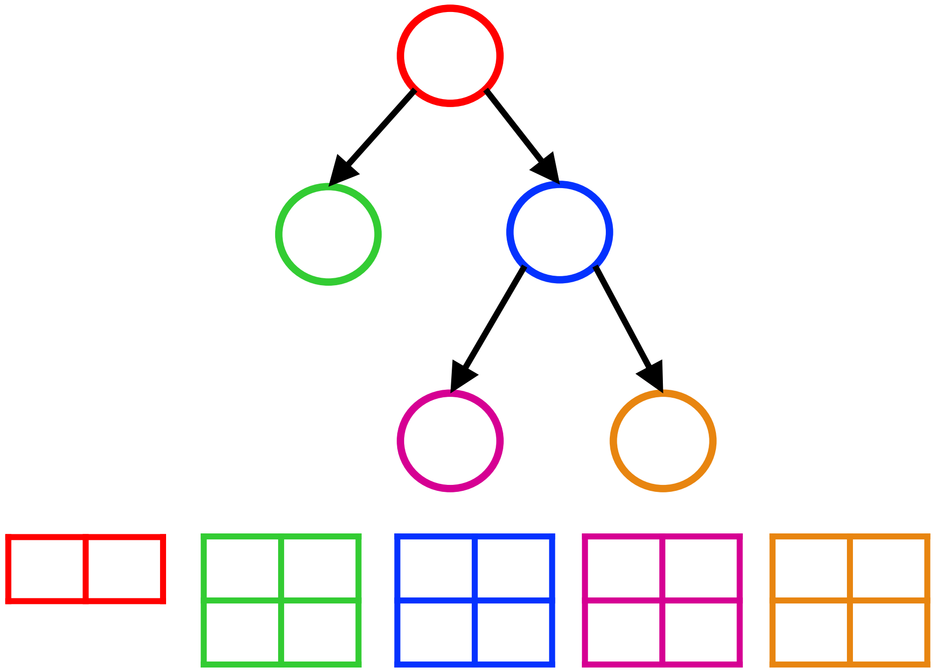
Warm-up as you walk in

Each node in a Bayes net represents a conditional probability distribution.

What distribution do you get when you multiply all of these conditional probabilities together?



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

Announcements

Midterm:

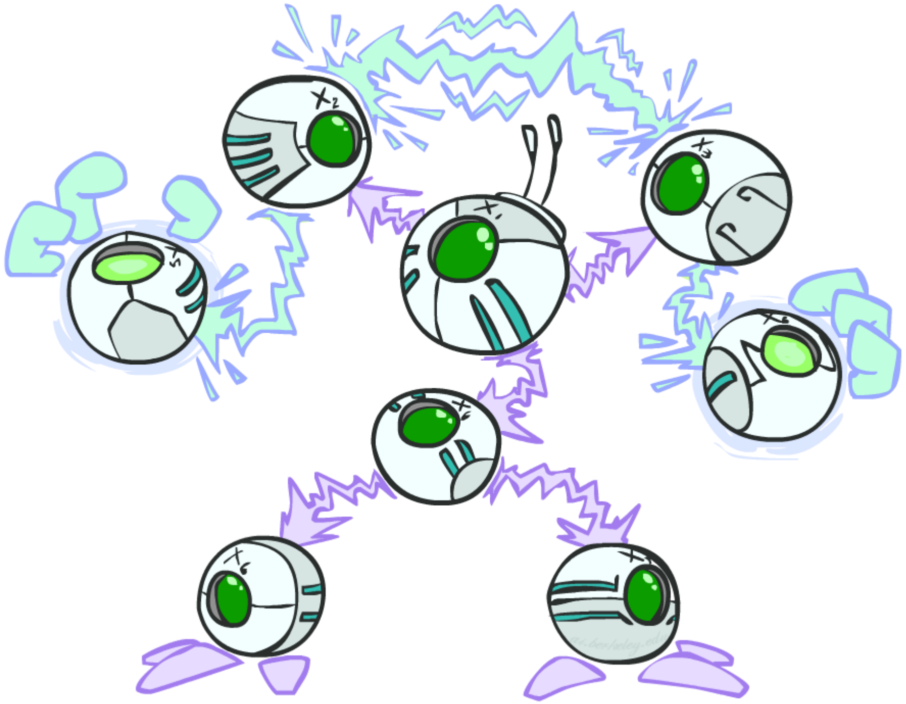
- Mon 4/8, in-class
- See Piazza for details

Course survey:

- Thanks!
- A few notes
  - In-class polls
  - Lectures
  - Recitation
  - TAs
  - Beyond Pac-man/Gridworld

AI: Representation and Problem Solving

Bayes Nets Independence



Instructors: Pat Virtue & Stephanie Rosenthal

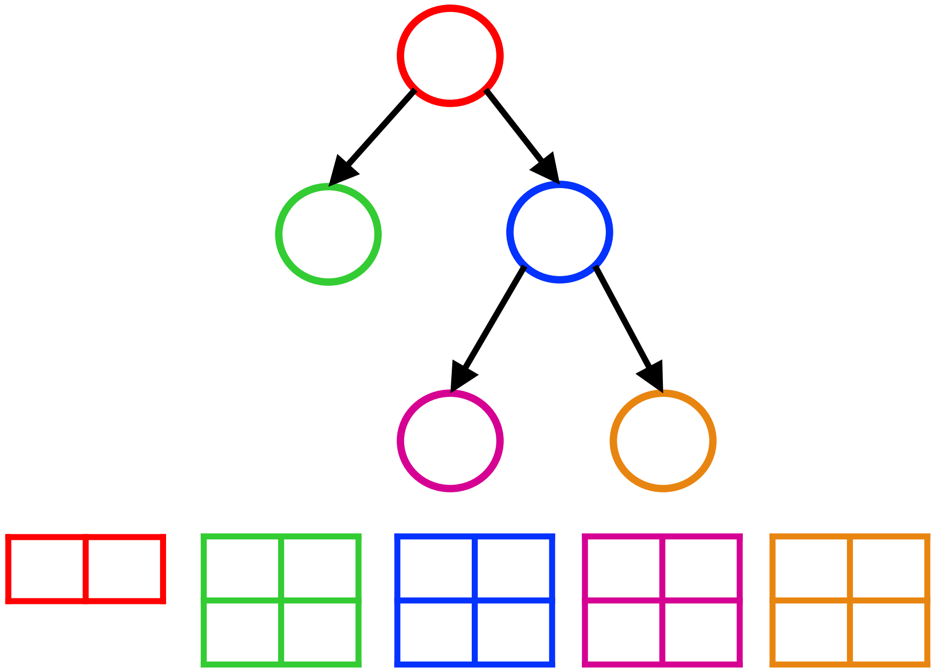
Slide credits: CMU AI and http://ai.berkeley.edu

Piazza Poll 1

Each node in a Bayes net represents a conditional probability distribution.

What distribution do you get when you multiply all of these conditional probabilities together?

A) Marginal
B) Conditional
C) Joint
D) Poisson



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

Piazza Poll 1

Each node in a Bayes net represents a conditional probability distribution.

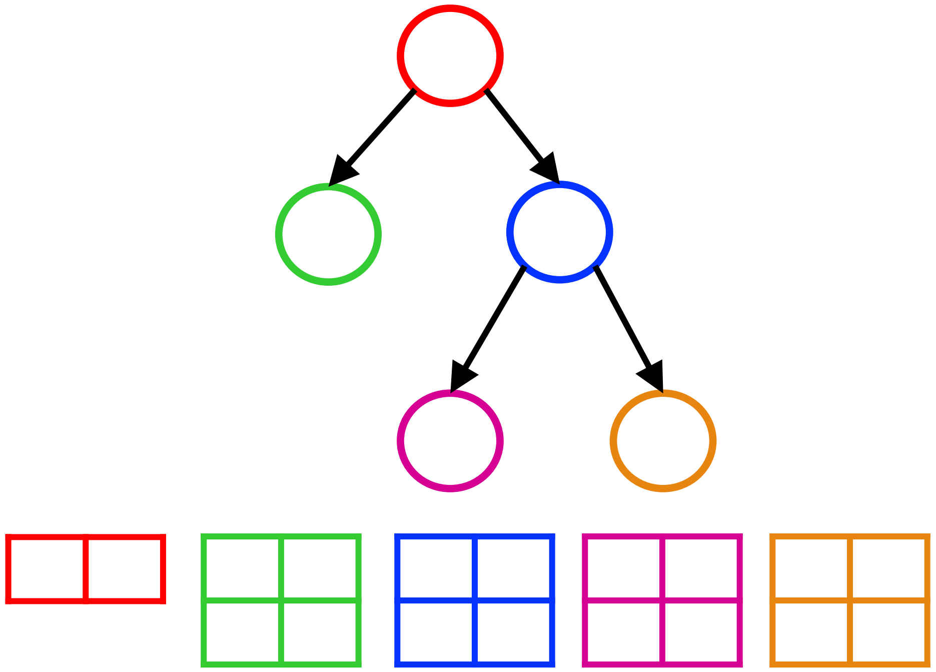
What distribution do you get when you multiply all of these conditional probabilities together?

A) Marginal

B) Conditional

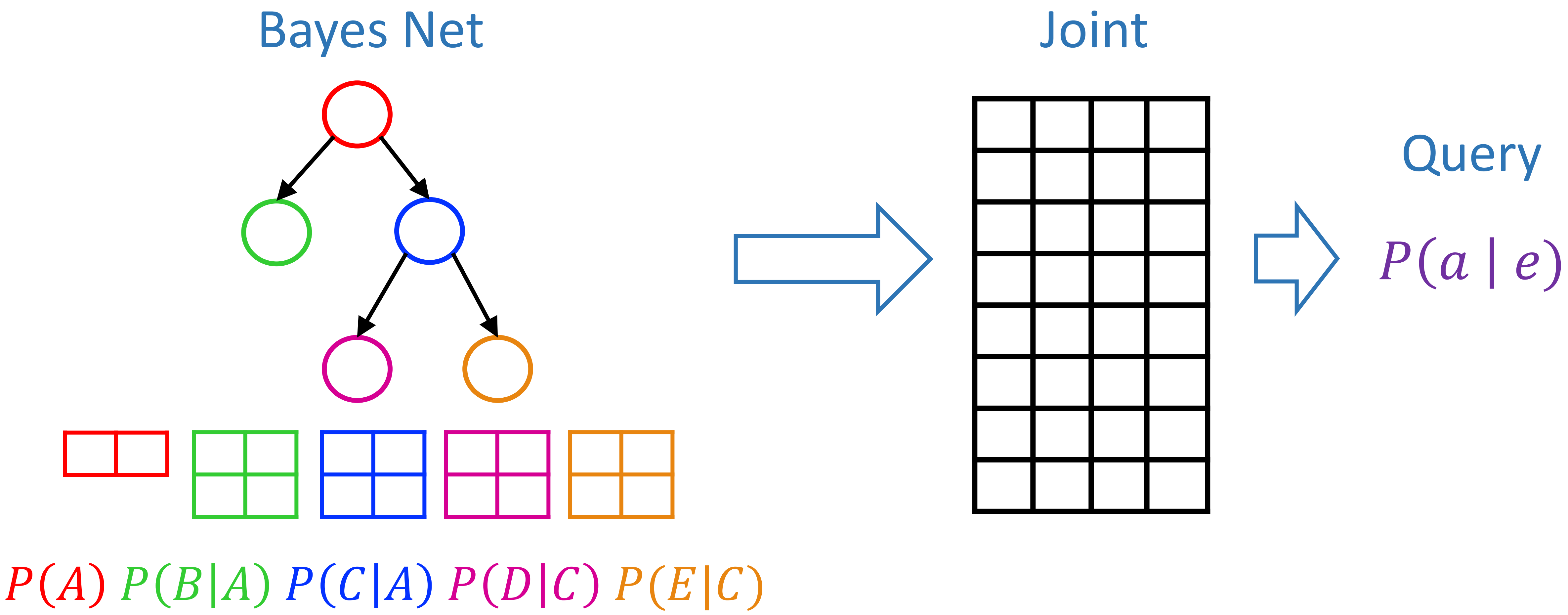
C) Joint

D) Poisson



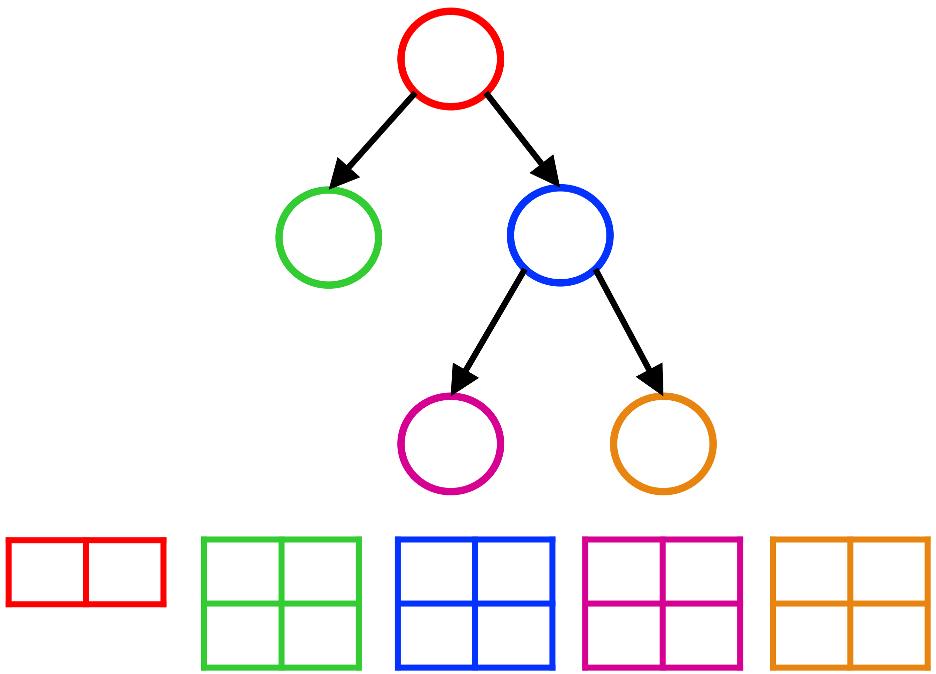
$$P(A,B,C,D,E) = P(A)\,P(B|A)\,P(C|A)\,P(D|C)\,P(E|C)$$

Answer Any Query from Condition Probability Tables



Bayes Net Independence Assumptions

Bayes Net



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

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

CPTs to Joint

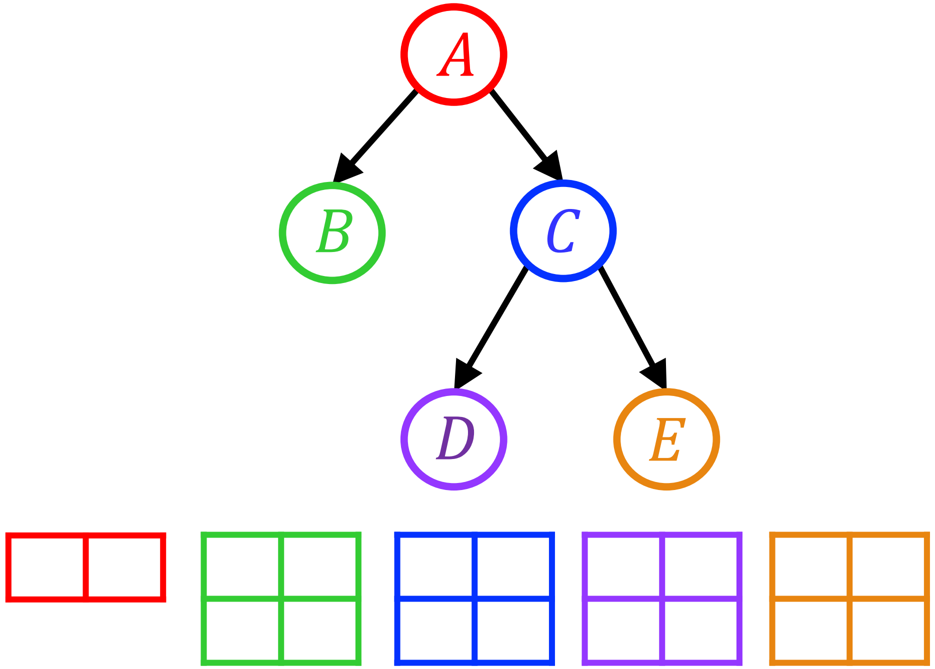
- Chain rule
- Bayes net model
  - Causality (or not)
  - Markov assumptions

Bayes Nets Representation

- One node per random variable
- DAG
- One CPT per node: P(node | *Parents*(node))
- Encodes joint distribution as product of conditional distributions on each variable

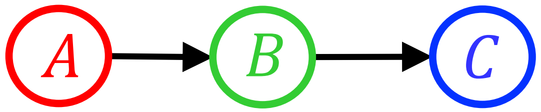
$$P(X_1, \ldots, X_N) = \prod_i P(X_i | \, Parents(X_i))$$

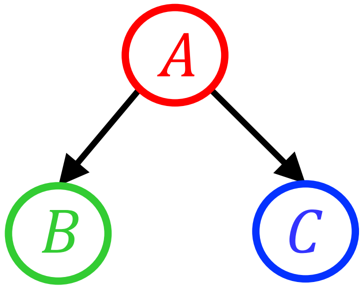
Bayes net

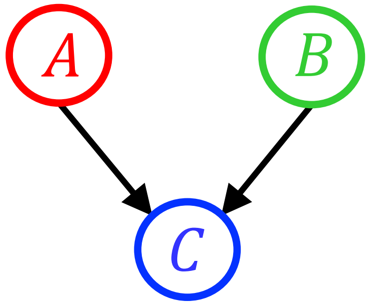


Piazza Poll 2

Match the product of CPTs to the Bayes net.



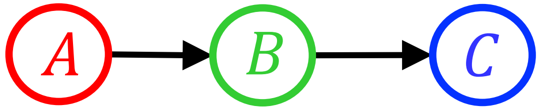


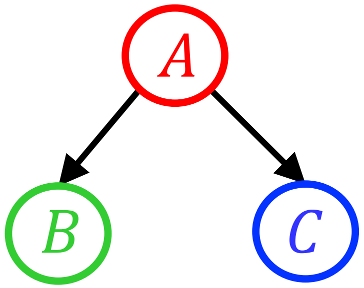


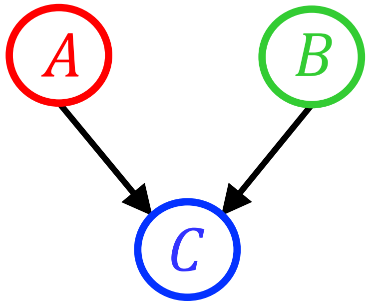
I.	$P(A)\,P(B\|A)\,P(C\|B)$	$P(A)\,P(B\|A)\,P(C\|A)$	$P(A)\,P(B)\,P(C\|A,B)$
II.	$P(A)\,P(B)\,P(C\|A,B)$	$P(A)\,P(B\|A)\,P(C\|B)$	$P(A)\,P(B\|A)\,P(C\|A)$
III.	$P(A)\,P(B\|A)\,P(C\|A)$	$P(A)\,P(B)\,P(C\|A,B)$	$P(A)\,P(B\|A)\,P(C\|B)$

Piazza Poll 2

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

Bayes Nets in the Wild

P(baton) P(bacon)

P(Third | pass, the)

P(Third | pass)

Wheel of Fortune

Bayes Nets in the Wild

Example: Speech Recognition

"artificial ………."

Find most probable next word given "artificial" and the audio for second word.

$P(F, S, A) = P(F)\,P(S \mid F)\,\underline{P(A \mid F, S)}$

Bayes Nets in the Wild

$F \to S \to A$

Example: Speech Recognition

"artificial ...."

$P(F)\,P(S \mid F)\,\underline{P(A \mid S)}$

Find most probable next word given "artificial" and the audio for second word.

Which second word gives the highest probability?

$P(\textbf{limb} \mid \text{artificial}, \text{audio})$

$P(\textbf{intelligence} \mid \text{artificial}, \text{audio})$

$P(\textbf{flavoring} \mid \text{artificial}, \text{audio})$

<u>Break down problem</u>

n-gram probability * audio probability

$P(\textbf{limb} \mid \text{artificial}) * P(\text{audio} \mid \textbf{limb})$

$P(\textbf{intelligence} \mid \text{artificial}) * P(\text{audio} \mid \textbf{intelligence})$

$P(\textbf{flavoring} \mid \text{artificial}) * P(\text{audio} \mid \textbf{flavoring})$

Bayes Nets in the Wild

First → Second → Audio

$$
\begin{aligned}
second^* &= \text{argmax}_{\boldsymbol{second}}\, P(\boldsymbol{second} \mid artificial, audio) \\
&= \text{argmax}_{\boldsymbol{second}}\, \frac{P(\boldsymbol{second}, artificial, audio)}{P(artificial, audio)} \\
&= \text{argmax}_{\boldsymbol{second}}\, P(\boldsymbol{second}, artificial, audio) \\
&= \text{argmax}_{\boldsymbol{second}}\, P(artificial)\, P(\boldsymbol{second} \mid artificial)\, P(audio \mid artificial, \boldsymbol{second}) \\
&= \text{argmax}_{\boldsymbol{second}}\, P(artificial)\, P(\boldsymbol{second} \mid artificial)\, P(audio \mid \boldsymbol{second}) \\
&= \text{argmax}_{\boldsymbol{second}}\, P(\boldsymbol{second} \mid artificial)\, P(audio \mid \boldsymbol{second})
\end{aligned}
$$

Definition of cond. prob.

Denominator doesn't affect argmax

Chain rule

Bayes Net

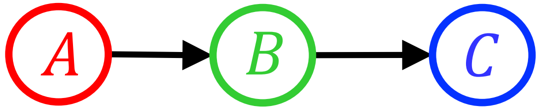
First factor doesn't affect argmax

n-gram probability * audio probability

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)



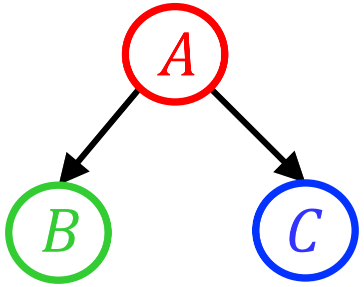
$P(A)\ P(B|A)\ P(C|A, B)$

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

Assumption:

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

C is independent from A given B



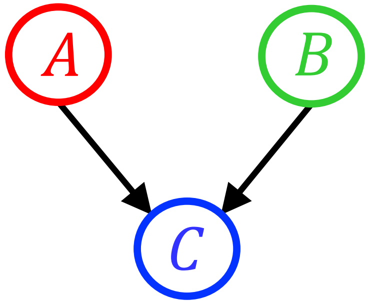
$P(A)\ P(B|A)\ P(C|A, B)$

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

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"

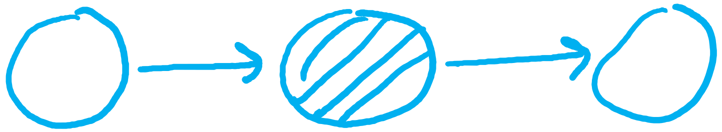
X: Low pressure   Y: Rain   Z: Traffic

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

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

      P(+y | +x) = 1, P(-y | - x) = 1,
      P(+z | +y) = 1, P(-z | -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?

shortcut

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

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

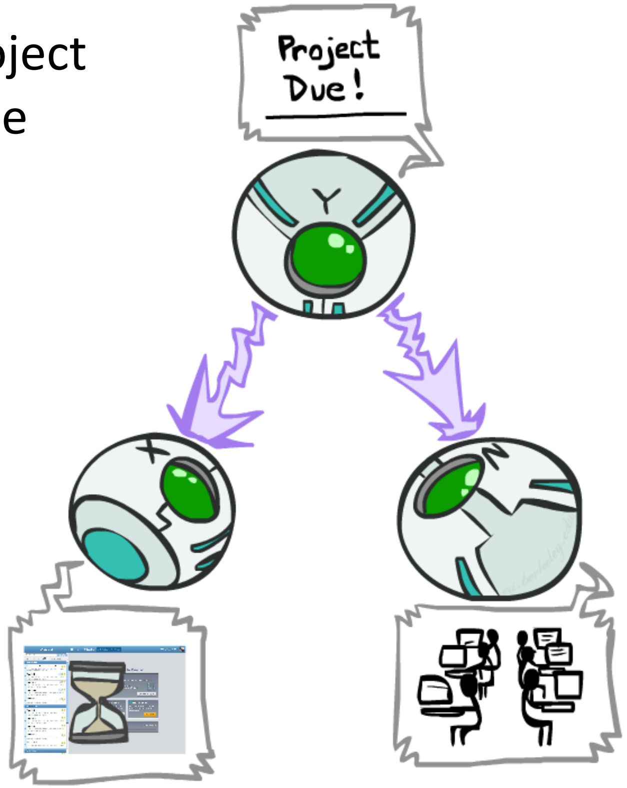
$$= P(z|y)$$

Yes!

- Evidence along the chain "blocks" the influence

Common Cause

This configuration is a "common cause"



Y: Project due

X: Forums busy

Z: Lab full

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

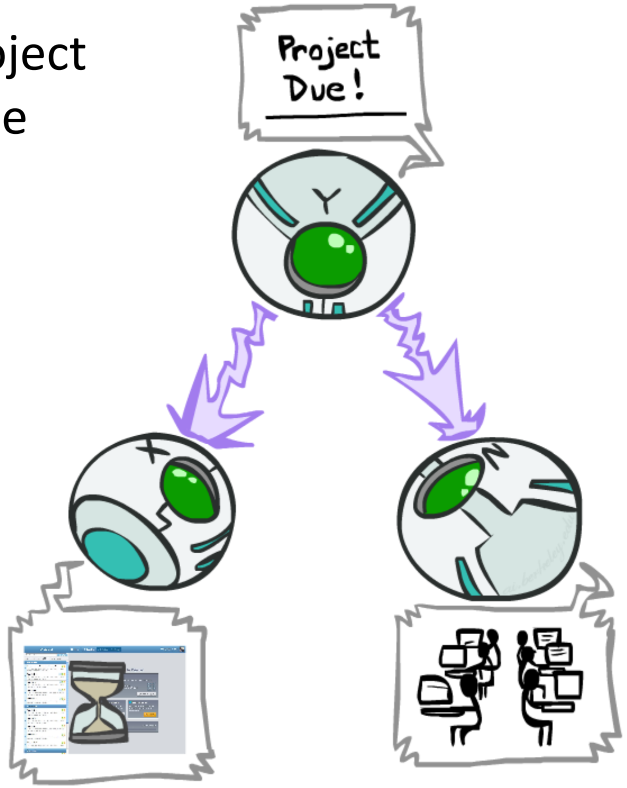
- Guaranteed X independent of Z ? *No!*
  - One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.
  - Example:
    - Project due causes both forums busy and lab full
    - In numbers:

P(+x | +y) = 1, P(-x | -y) = 1,
P(+z | +y) = 1, P(-z | -y) = 1

Common Cause

This configuration is a “common cause”

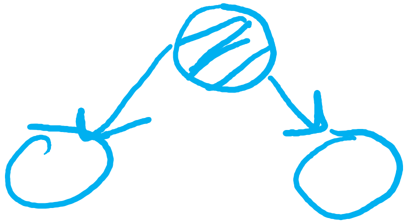
Y: Project due



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?

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

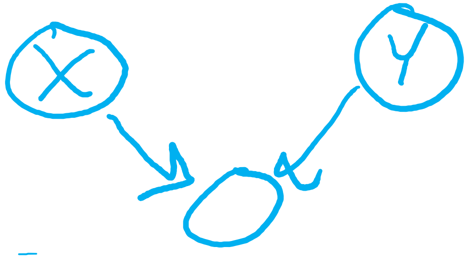
$$= \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}$$

$$= P(z|y)$$

Yes!

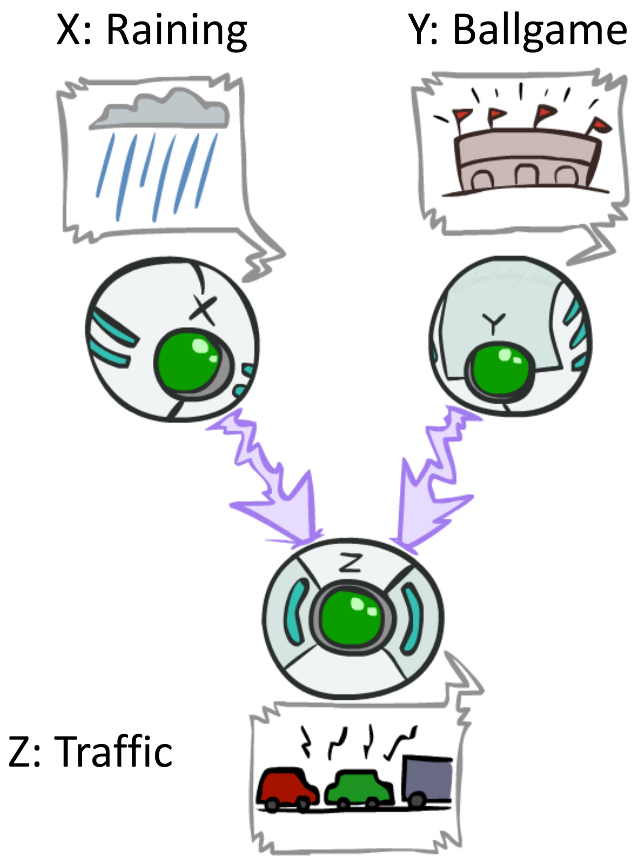
- Observing the cause blocks influence between effects.

Common Effect



$X \perp\!\!\!\perp Y$

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



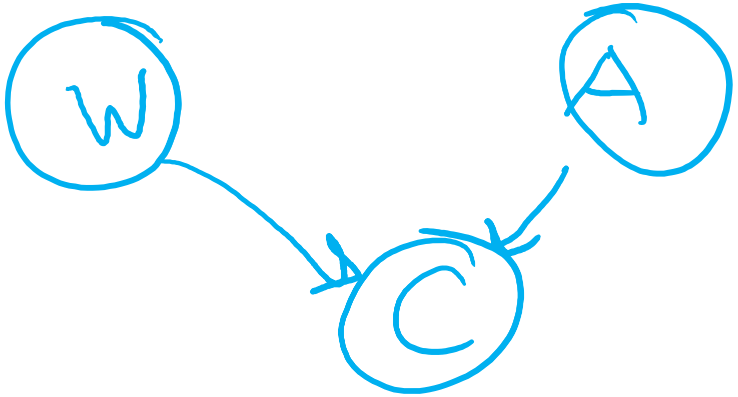
- Are X and Y independent?
  - *Yes*: the ballgame and the rain cause traffic, but they are not correlated
  - Still need to prove they must be (try it!)
- Are X and Y independent given Z?
  - *No*: seeing traffic puts the rain and the ballgame in competition as explanation.
- This is backwards from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

Practice

NO $W \perp\!\!\!\perp A \mid C$

Consider the case where you show up to Claire's office hours and she isn't there

- Draw a Bayes Net capturing the causal relationship between **Claire** being in the room (yes/no), Claire being abducted by **aliens** (yes/no), and your **watch** having the wrong time (yes/no)



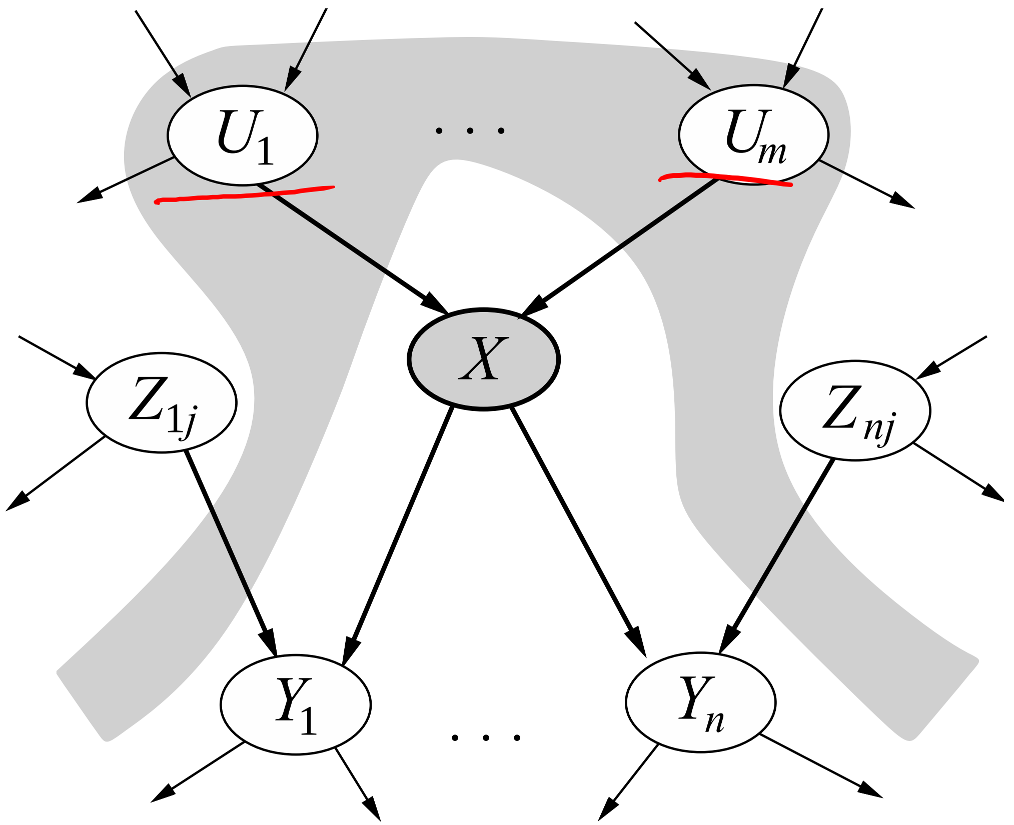
$W \perp\!\!\!\perp A$

$P(A \mid C) \neq P(A \mid W, C)$

- What statements about independence can you make?

Conditional Independence Semantics

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

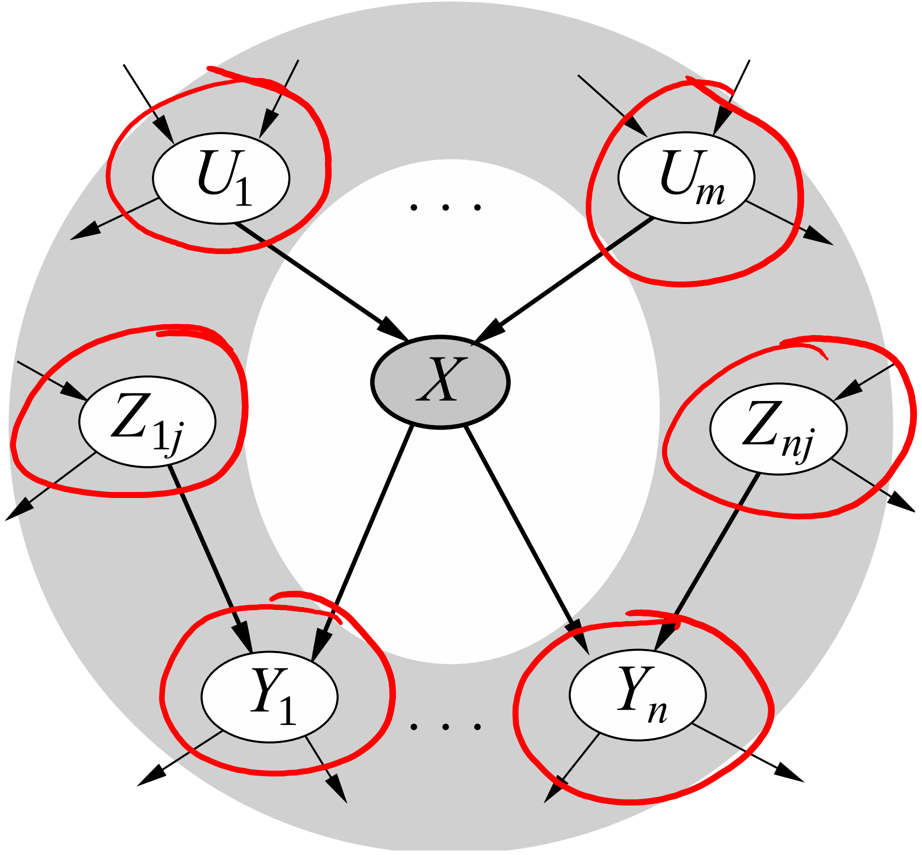


$X \perp\!\!\!\perp$ | Parents

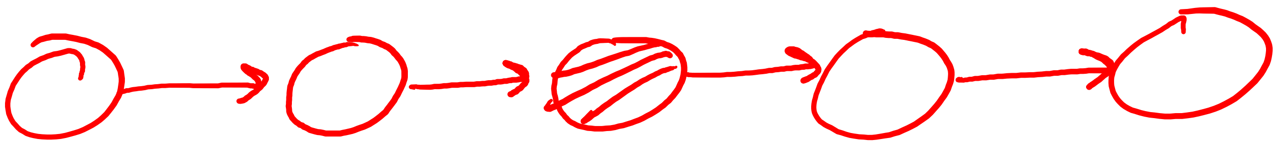
Markov blanket

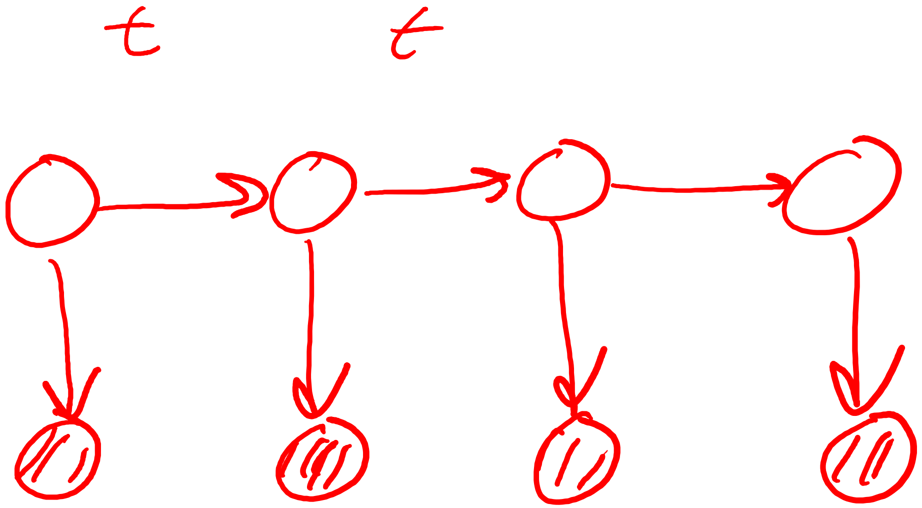
A variable's Markov blanket consists of parents, children, children's other parents

Every variable is conditionally independent of all other variables given its Markov blanket

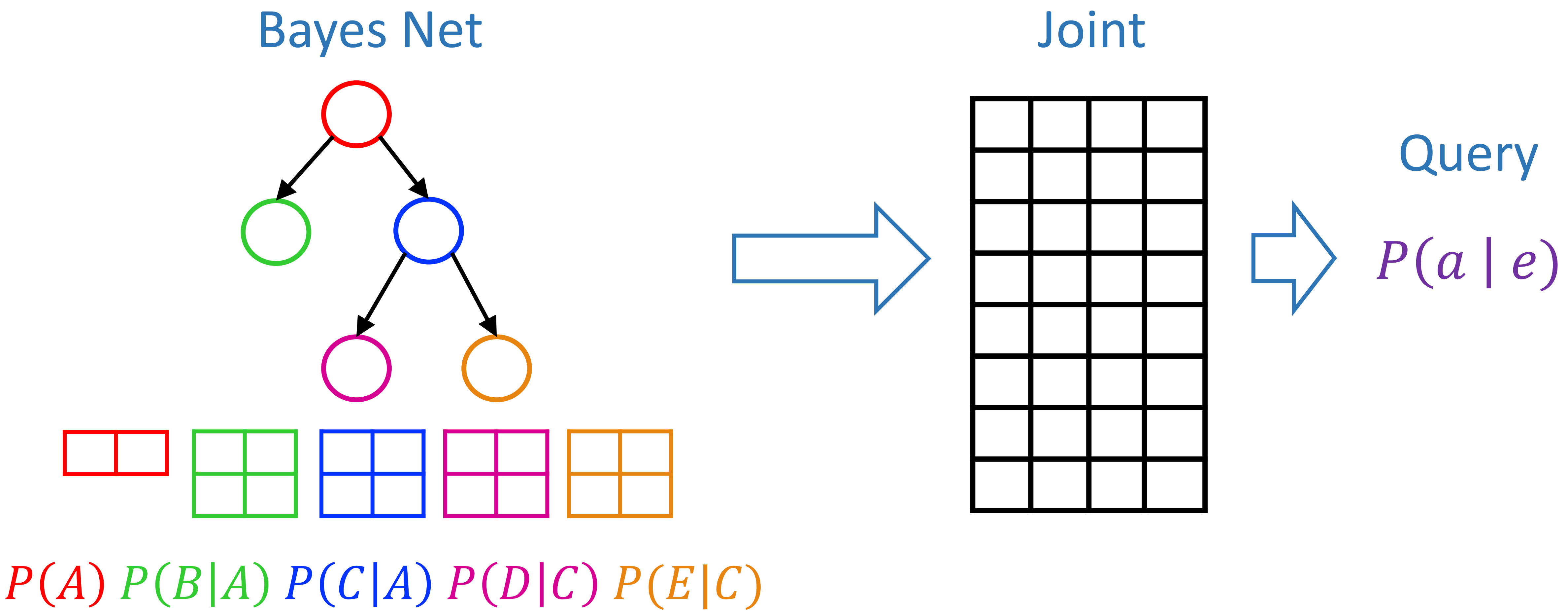


Markov Structures





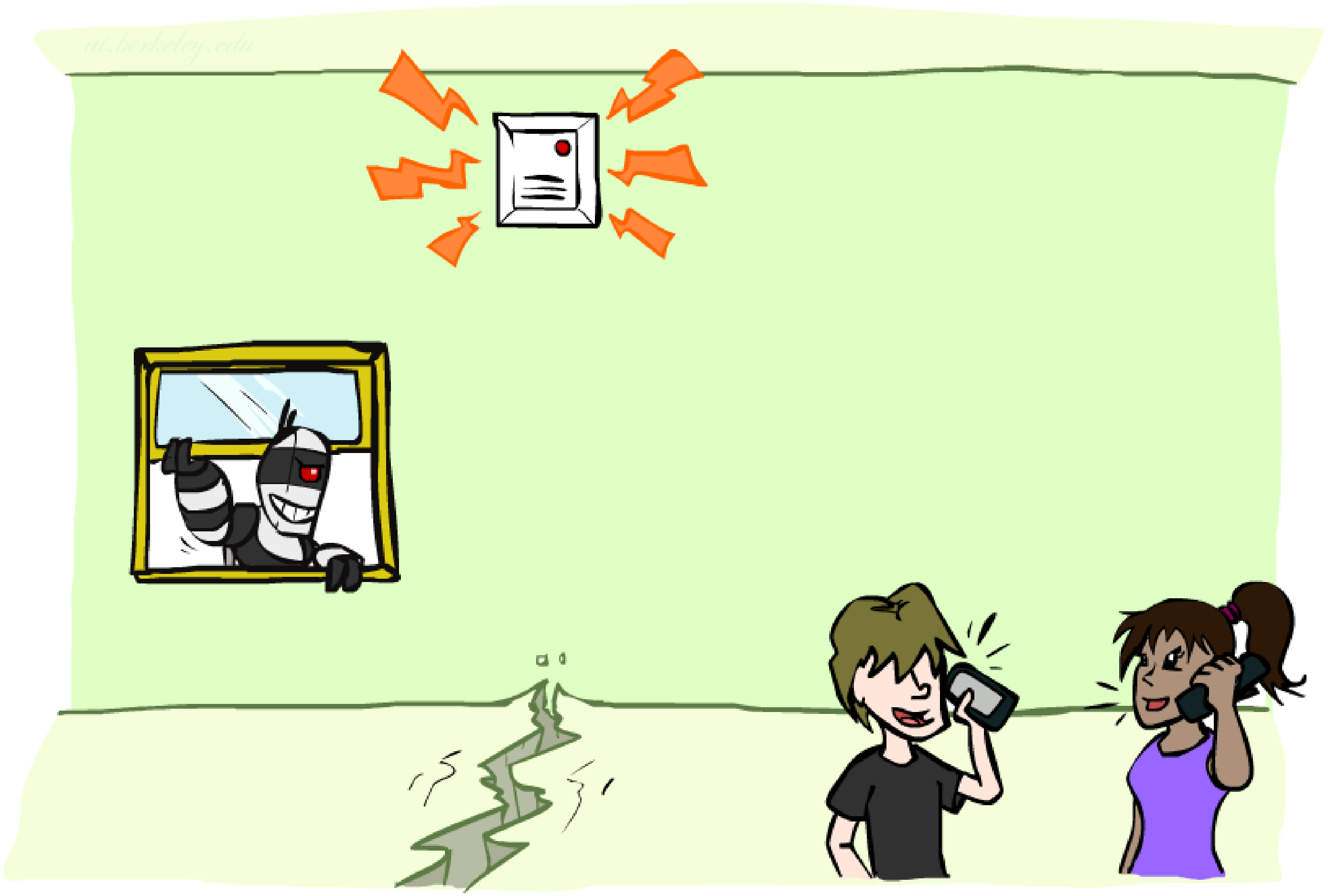
Answer Any Query from Condition Probability Tables



Example: Alarm Network

Variables

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



Example: Alarm Network

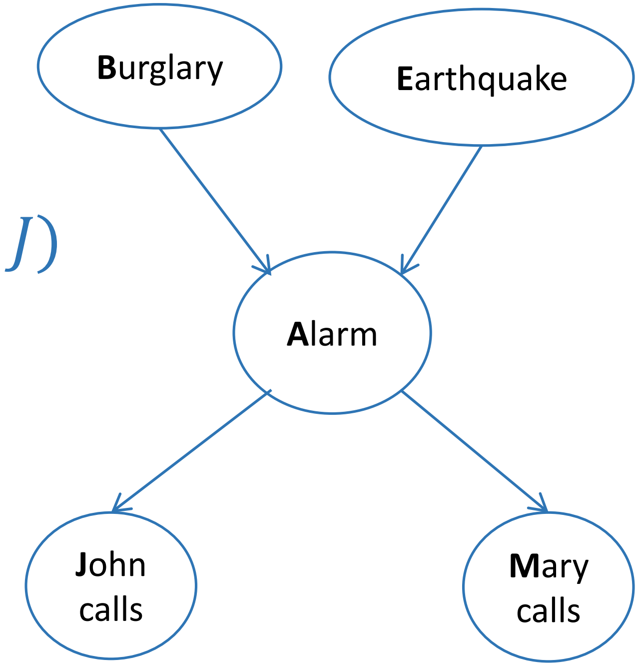
Joint distribution factorization example

Generic chain rule

- $P(X_1 \ldots X_2) = \prod_i P(X_i | X_1 \ldots X_{i-1})$

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

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



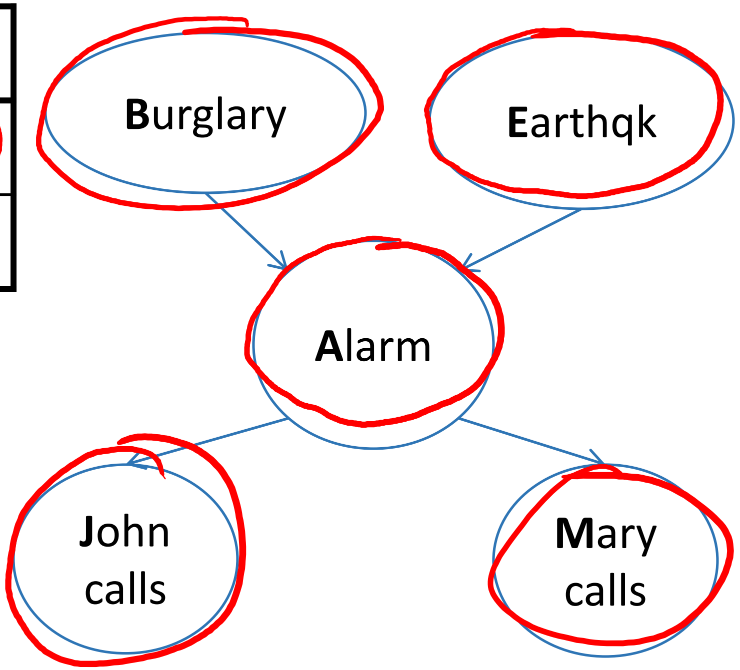
Bayes nets

- $P(X_1 \ldots X_2) = \prod_i P(X_i | Parents(X_i))$

Example: Alarm Network

$P(+b, -e, -a, -m, -j) =$ ___ ___ ___ ___ ___

B	P(B)
+b	0.001
-b	0.999



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

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