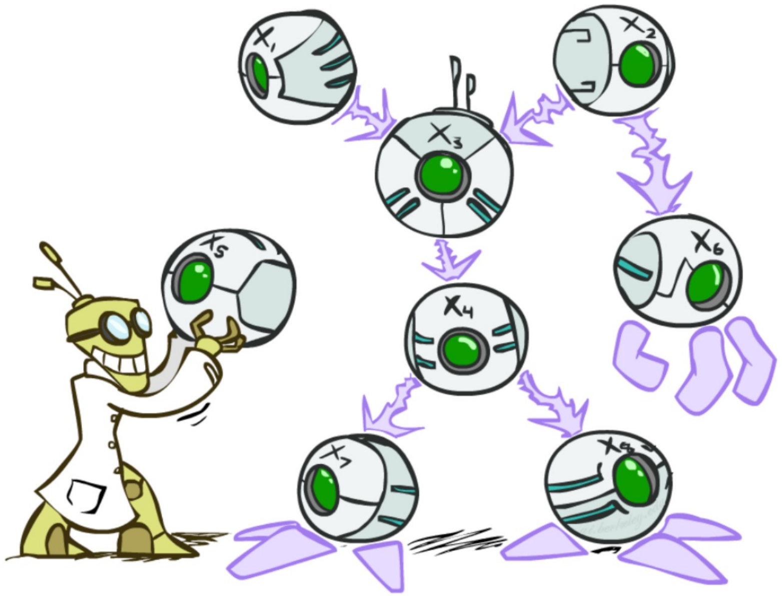
# Bayes Nets

CSE 473: Introduction to Artificial Intelligence



**James Weichert** (jpw@cs)
**cs.uw.edu/473**
acknowledgements



**Welcome!**
*What should the CSE 473 logo/mascot be?*
sli.do **#cse473**

**Mid-Quarter Feedback Survey :)**

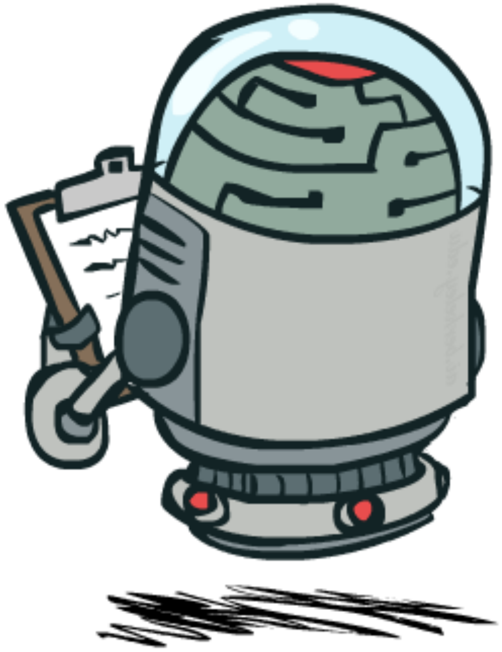
# Administrivia

## ANNOUNCEMENTS

- **Test 1** scores released, regrade requests due tomorrow
- Complete the **Mid-Quarter Feedback** survey!

## ASSIGNMENTS

- **Homework 2 due** tomorrow (7.23)
- **Project 3 due** next Thursday (7.30)

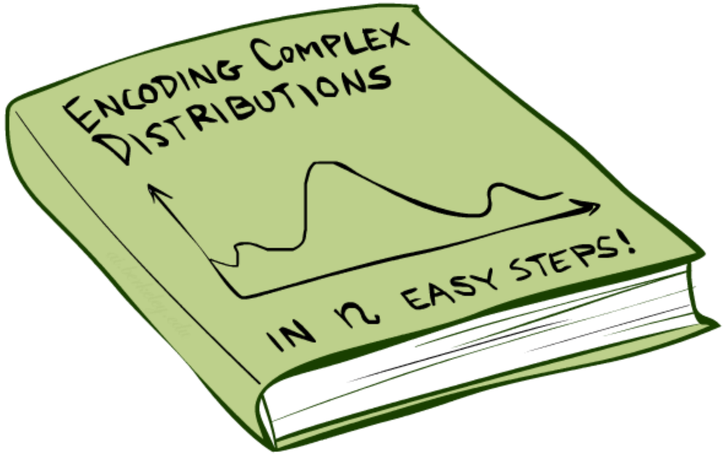


# Probability Recap

CONTEXT

- Laws of Probability
  - 'Worlds' $\omega$ in universe $\Omega$
  - $0 \leq P(\omega) \leq 1$
  - $\sum_{\omega \in \Omega} P(\omega) = 1$
  - $P(A) = \sum_{\omega \in A} P(\omega)$
- Random Variable $X$
  - $X(\omega)$ maps outcomes $\omega$ to range $x \epsilon X$
  - $P(X = x)$ is a deterministic function from $x \in X$ to $[0,1]$
- Distributions
  - **Joint Distribution** $P(X, Y)$ gives total probability for each $x, y$ combination
  - **Marginal Distribution** $P(X) = \sum_{y \in Y} P(X, Y = y)$
  - **Conditional Distribution** $P(X|Y) = \frac{P(X,Y)}{P(Y)}$
- Product Rule
  - $P(X, Y) = P(X|Y) \cdot P(Y)$
  - Chain Rule: $P(X_1, \dots, X_n) = \prod_i P(X_i \,|X_1, \dots, X_{i-1})$
- Bayes Rule
  - $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$

# *How to manage complex probability distributions?*

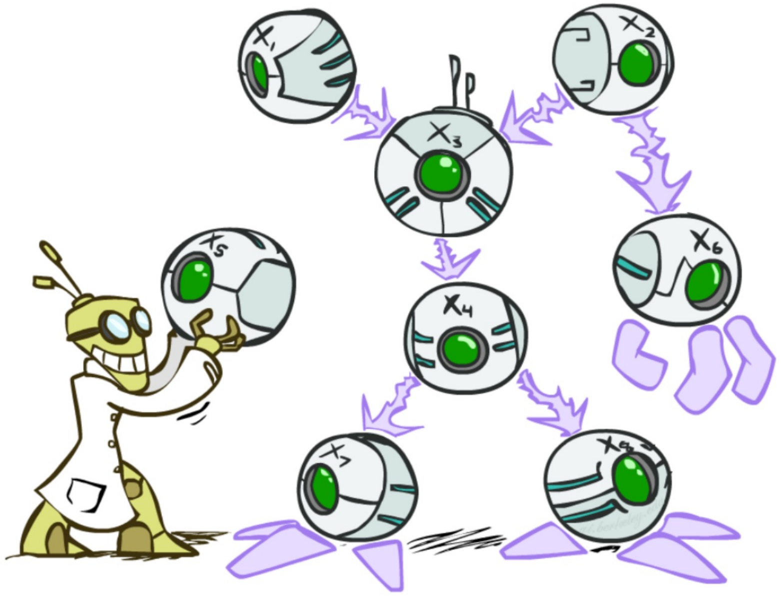


# Bayesian Networks

FOUNDATIONS

**Bayes Nets** provide a technique describing **complex joint distributions** using **simple conditional distributions**

- Subset of a general class of graphical models
- Also called "belief networks"

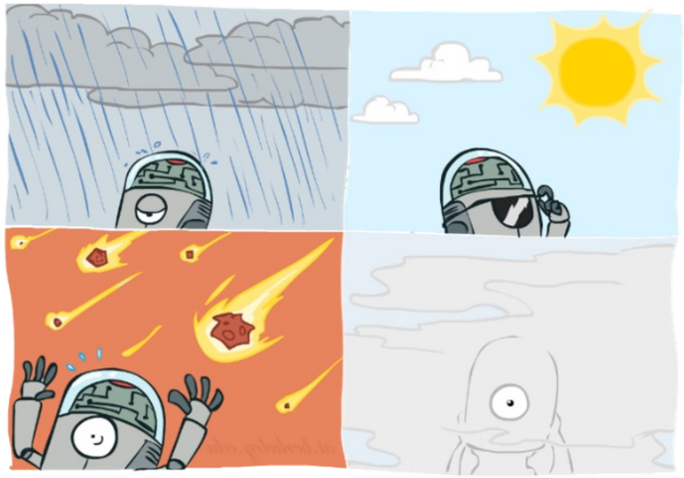


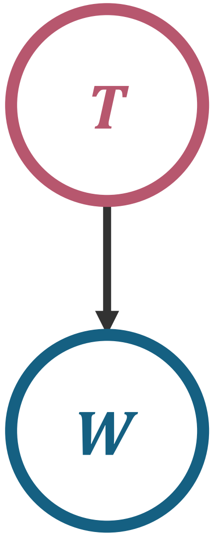
- Use local causality / conditional independence
  - The world is composed of many variables
  - Each interacting locally with a few others
- Expressivity vs. Efficiency
  - While conditional probabilities are less expressive (comprehensive) than joint probabilities, they are faster and more space-efficient to work with

# Graphical Models

DEFINITION

- Nodes: Variables
  - Assigned (observed)
  - Unassigned (unobserved)
- Arcs: Interactions
  - Indicate "direct influence" between variables
  - Formally: encode conditional (in)dependence



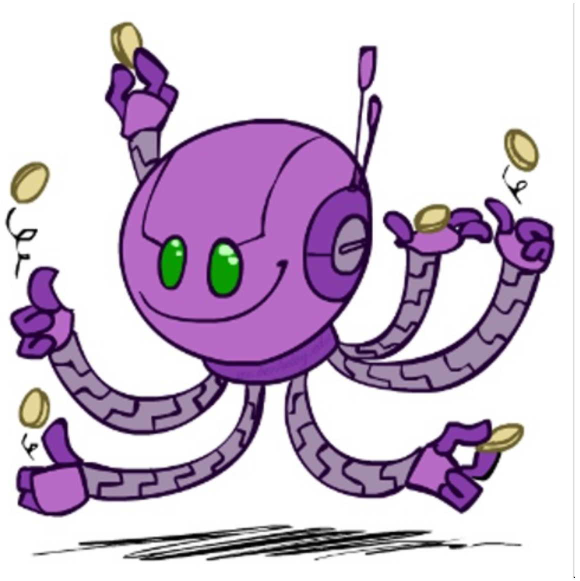


# Coin Flips

EXAMPLE

$N$ independent coin flips

- No interactions between variables $\Longrightarrow$ absolute independence

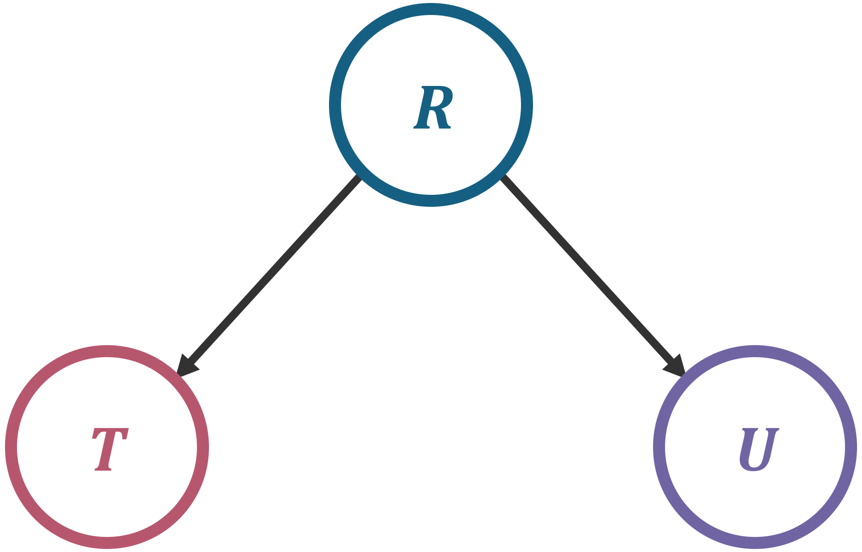



# Traffic

EXAMPLE

## Variables

- $R$ it rains
- $T$ there's traffic
- $U$ I'm holding my umbrella

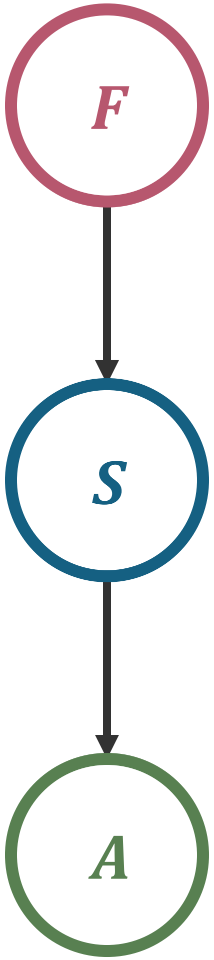


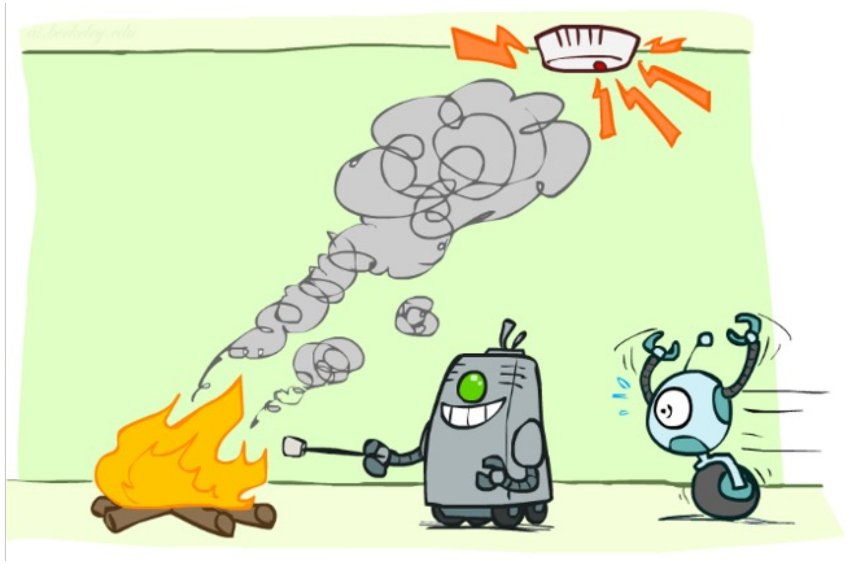

# Smoke Alarm

EXAMPLE

## Variables

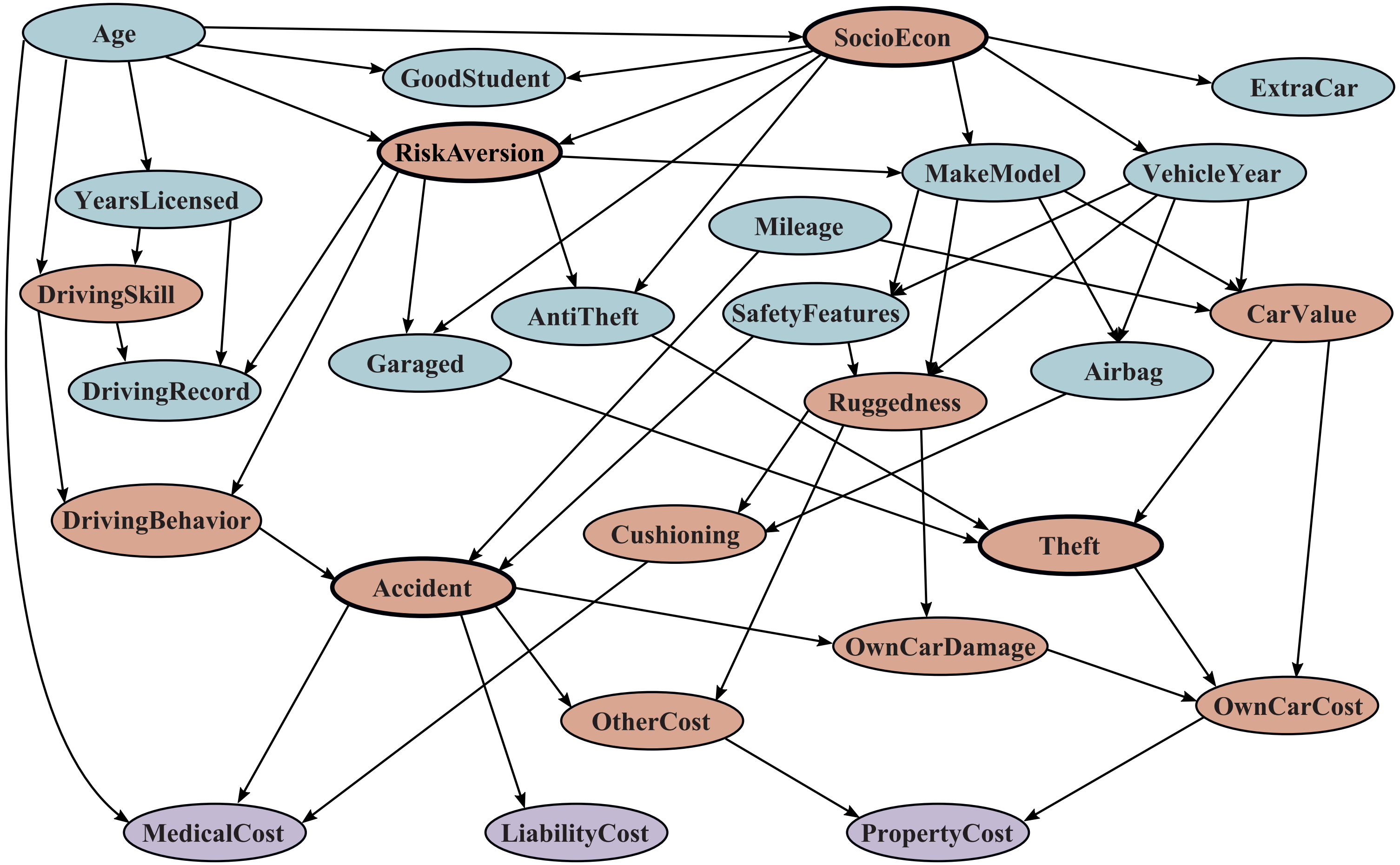
- $F$ there is a fire
- $S$ there is smoke
- $A$ alarm sounds



# Car Insurance

EXAMPLE



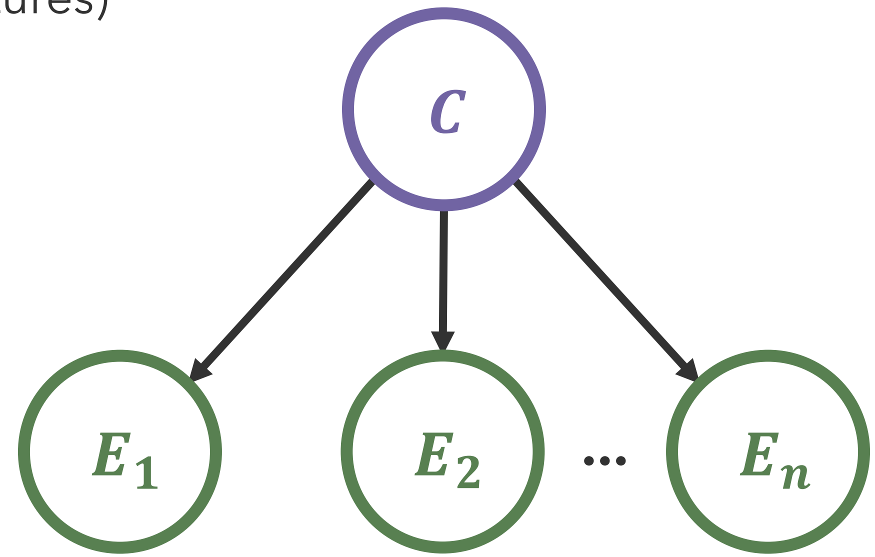
# Naïve Bayes

EXAMPLE

## Variables

- $C$ class we want to predict
- $E_1, \ldots, E_n$ evidence (or features)

"**Naïve**" because of the simplifying assumption that evidence variables are **conditionally independent** given $C$



# Questions?

live and on sli.do #cse473

# Syntax and Semantics

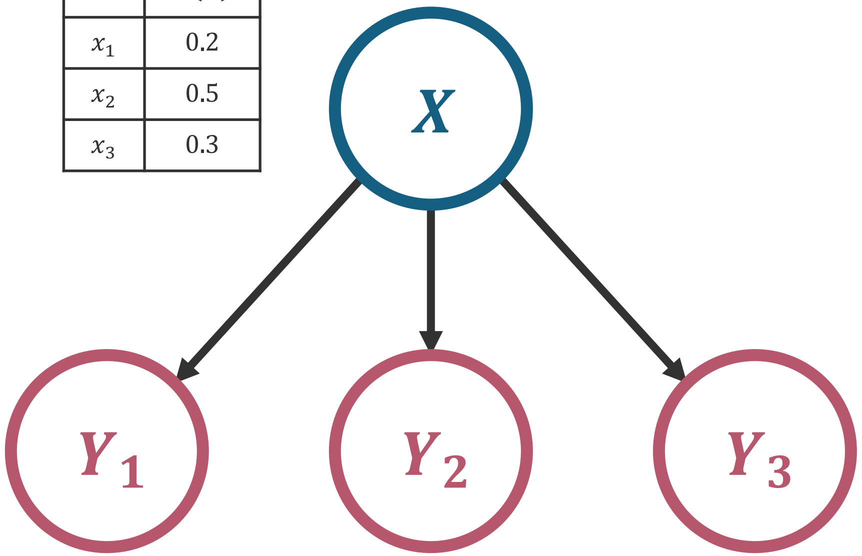
FOUNDATIONS

# Bayes Net Syntax

DEFINITION

- Set of nodes, one per variable $X_i$
- Directed, acyclic graph
- Conditional distribution for each node given parent
  - Conditional Probability Table (CPT)

| $X$ | $P(X)$ |
|---|---|
| $x_1$ | 0.2 |
| $x_2$ | 0.5 |
| $x_3$ | 0.3 |




| | $P(Y_1\|X)$ | | |
|---|---|---|---|
| $X$ | $Y_1 = a$ | $Y_1 = b$ | $Y_1 = c$ |
| $x_1$ | 0.2 | 0.6 | 0.2 |
| $x_2$ | 0.1 | 0.0 | 0.9 |
| $x_3$ | 0.4 | 0.3 | 0.3 |

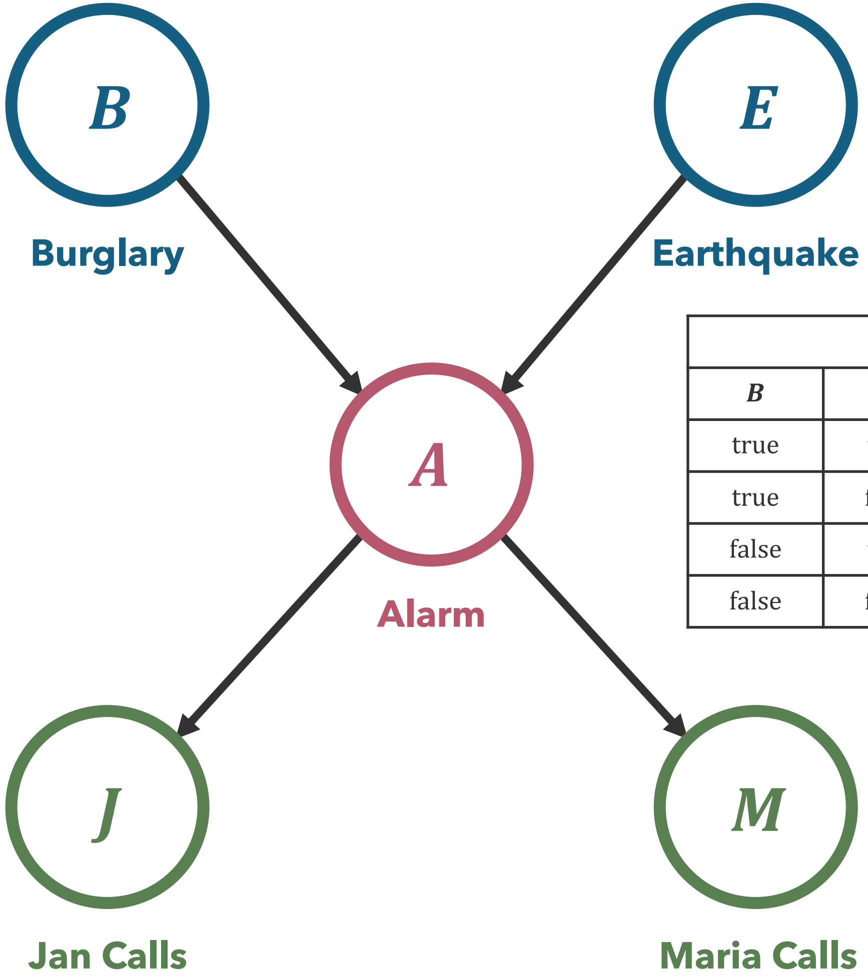
| | $P(Y_3\|X)$ | | |
|---|---|---|---|
| $X$ | $Y_3 = d$ | $Y_3 = e$ | $Y_3 = f$ |
| $x_1$ | 0.0 | 0.1 | 0.9 |
| $x_2$ | 0.25 | 0.25 | 0.5 |
| $x_3$ | 0.8 | 0.1 | 0.1 |

**Bayes Net = Graph + Local Probability Tables**

# Alarm Bayes Net

EXAMPLE

| $B$ | $P(B)$ |
|---|---|
| true | 0.001 |
| false | 0.999 |

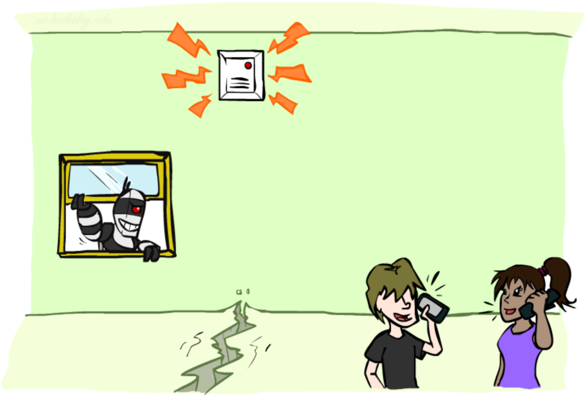



| $E$ | $P(E)$ |
|---|---|
| true | 0.002 |
| false | 0.998 |

| | | $P(A\|B,E)$ | |
|---|---|---|---|
| $B$ | $E$ | true | false |
| true | true | 0.95 | 0.05 |
| true | false | 0.94 | 0.06 |
| false | true | 0.29 | 0.71 |
| false | false | 0.001 | 0.999 |

| | $P(J\|A)$ | |
|---|---|---|
| $A$ | true | false |
| true | 0.9 | 0.1 |
| false | 0.05 | 0.95 |

| | $P(M\|A)$ | |
|---|---|---|
| $A$ | true | false |
| true | 0.7 | 0.3 |
| false | 0.01 | 0.99 |



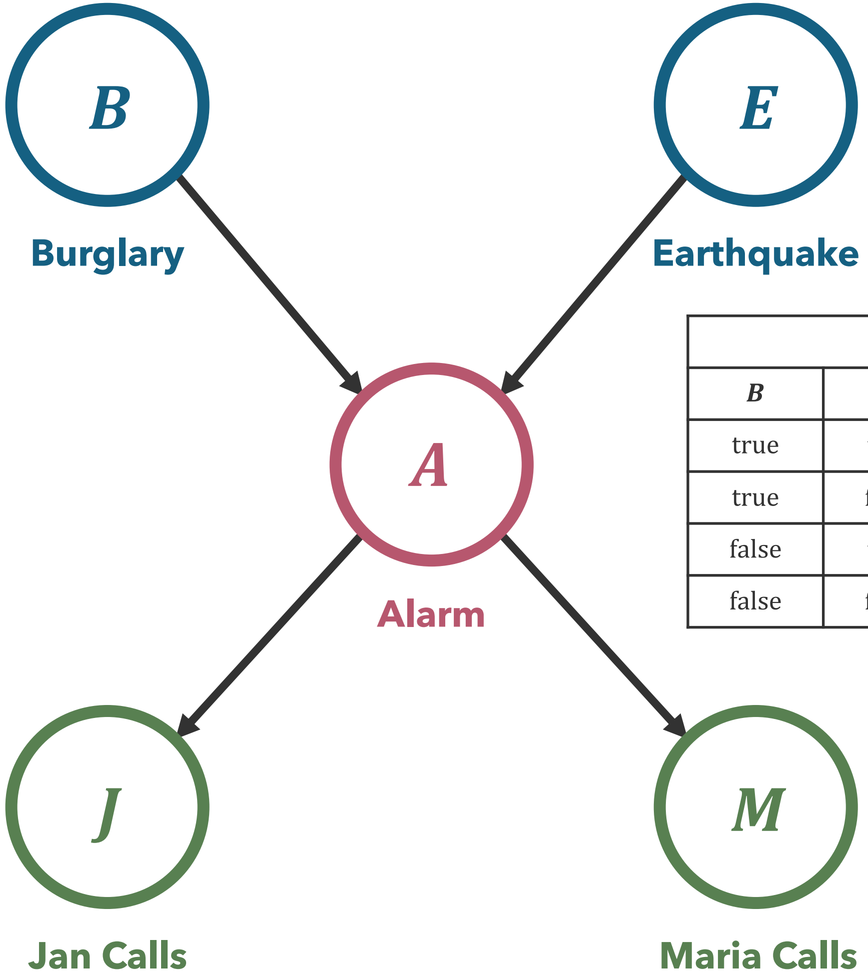
**Factor Size** per CPT:

$$= d_c \prod_{i \in \text{parents}(X_c)} d_i$$

# Alarm Bayes Net Practice

**EXAMPLE**

| $B$ | $P(B)$ |
|---|---|
| true | 0.001 |
| false | 0.999 |

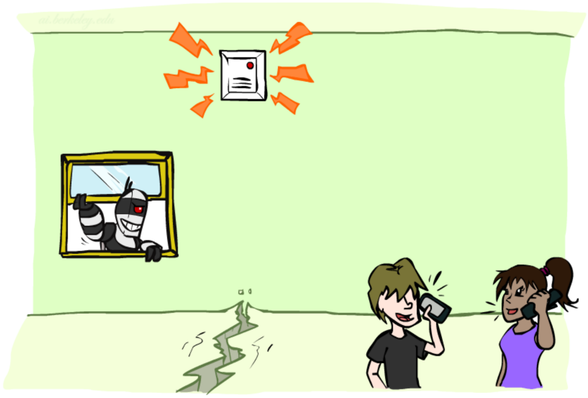



| $E$ | $P(E)$ |
|---|---|
| true | 0.002 |
| false | 0.998 |

| | | $P(A \mid B, E)$ | |
|---|---|---|---|
| $B$ | $E$ | true | false |
| true | true | 0.95 | 0.05 |
| true | false | 0.94 | 0.06 |
| false | true | 0.29 | 0.71 |
| false | false | 0.001 | 0.999 |

| | $P(J \mid A)$ | |
|---|---|---|
| $A$ | true | false |
| true | 0.9 | 0.1 |
| false | 0.05 | 0.95 |

| | $P(M \mid A)$ | |
|---|---|---|
| $A$ | true | false |
| true | 0.7 | 0.3 |
| false | 0.01 | 0.99 |

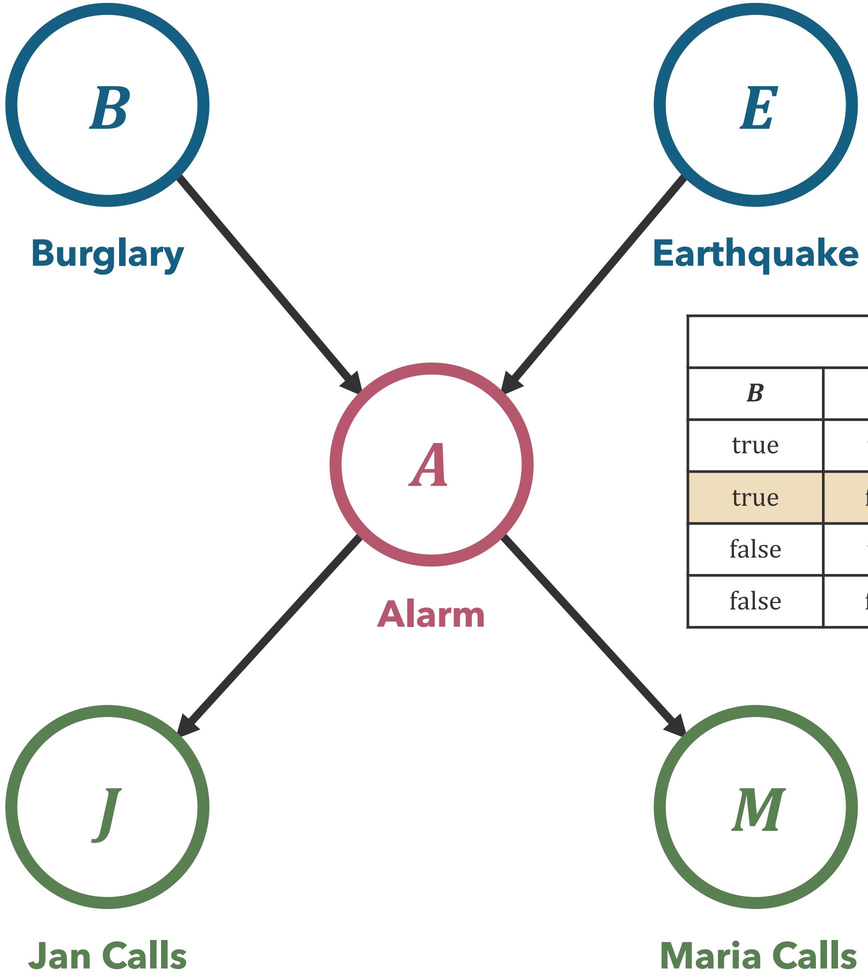


Calculate $P(b, \neg e, a, \neg j, \neg m)$

# Alarm Bayes Net Practice (1)

EXAMPLE

| $B$ | $P(B)$ |
|---|---|
| true | 0.001 |
| false | 0.999 |

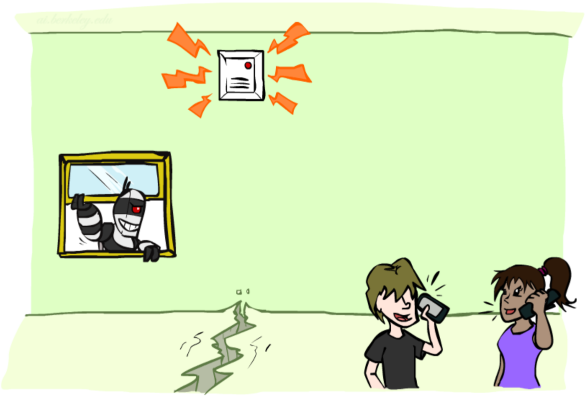



| $E$ | $P(E)$ |
|---|---|
| true | 0.002 |
| false | 0.998 |

| | | $P(A\|B,E)$ | |
|---|---|---|---|
| $B$ | $E$ | true | false |
| true | true | 0.95 | 0.05 |
| true | false | 0.94 | 0.06 |
| false | true | 0.29 | 0.71 |
| false | false | 0.001 | 0.999 |

| | $P(J\|A)$ | |
|---|---|---|
| $A$ | true | false |
| true | 0.9 | 0.1 |
| false | 0.05 | 0.95 |

| | $P(M\|A)$ | |
|---|---|---|
| $A$ | true | false |
| true | 0.7 | 0.3 |
| false | 0.01 | 0.99 |

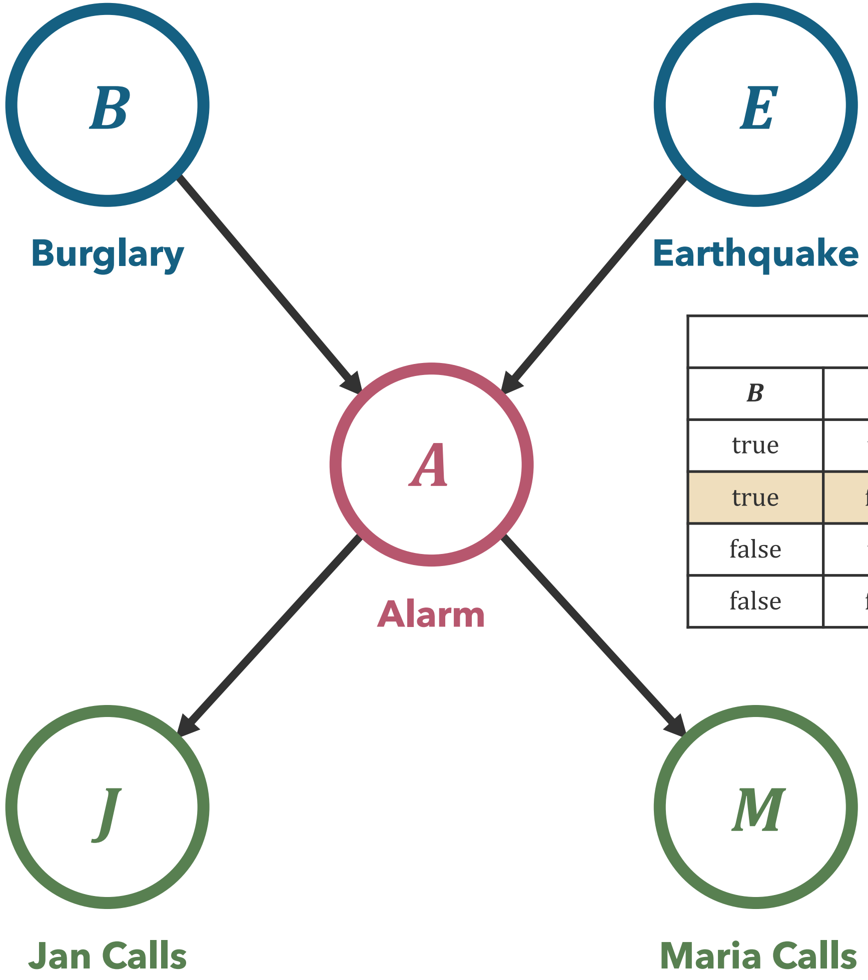


Calculate $P(b, \neg e, a, \neg j, \neg m)$

# Alarm Bayes Net Practice (2)

EXAMPLE

| $B$ | $P(B)$ |
|---|---|
| true | 0.001 |
| false | 0.999 |

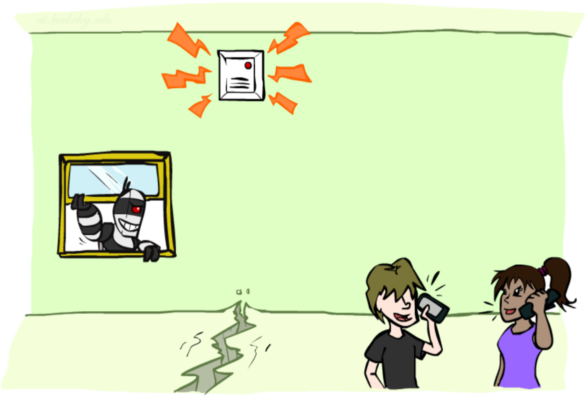



| $E$ | $P(E)$ |
|---|---|
| true | 0.002 |
| false | 0.998 |

| | | $P(A\|B,E)$ | |
|---|---|---|---|
| $B$ | $E$ | true | false |
| true | true | 0.95 | 0.05 |
| true | false | 0.94 | 0.06 |
| false | true | 0.29 | 0.71 |
| false | false | 0.001 | 0.999 |

| | $P(J\|A)$ | |
|---|---|---|
| $A$ | true | false |
| true | 0.9 | 0.1 |
| false | 0.05 | 0.95 |

| | $P(M\|A)$ | |
|---|---|---|
| $A$ | true | false |
| true | 0.7 | 0.3 |
| false | 0.01 | 0.99 |



Calculate $\boldsymbol{P(b, \neg e, a, \neg j, \neg m)}$

$$= P(b) \cdot P(\neg e) \cdots$$

$$\cdot P(a|b, \neg e) \cdots$$

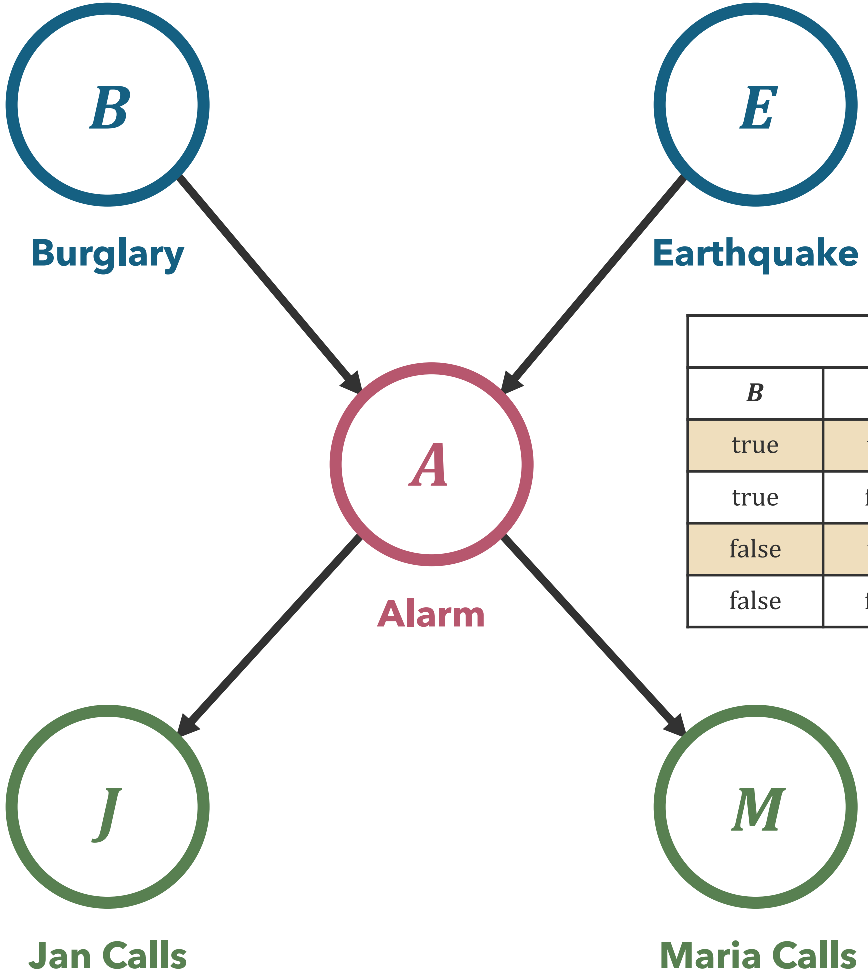
$$\cdot P(\neg j|a) \cdot P(\neg m|a) \cdots$$

$$= 0.001 \cdot 0.998 \cdot 0.94 \cdot 0.1 \cdot 0.3 = 0.000028$$

# Alarm Bayes Net Practice (3)

EXAMPLE

| $B$ | $P(B)$ |
|---|---|
| true | 0.001 |
| false | 0.999 |

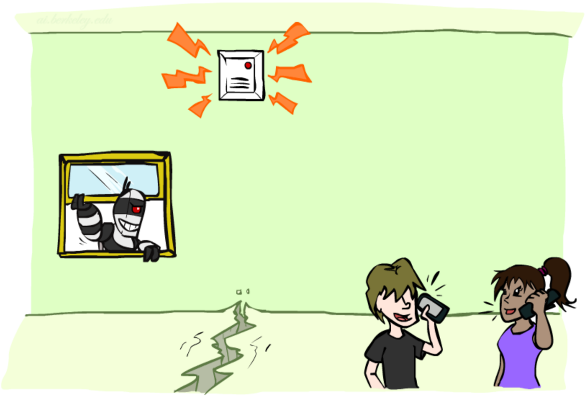



| $E$ | $P(E)$ |
|---|---|
| true | 0.002 |
| false | 0.998 |

| | | $P(A\|B,E)$ | |
|---|---|---|---|
| $B$ | $E$ | true | false |
| true | true | 0.95 | 0.05 |
| true | false | 0.94 | 0.06 |
| false | true | 0.29 | 0.71 |
| false | false | 0.001 | 0.999 |

| | $P(J\|A)$ | |
|---|---|---|
| $A$ | true | false |
| true | 0.9 | 0.1 |
| false | 0.05 | 0.95 |

| | $P(M\|A)$ | |
|---|---|---|
| $A$ | true | false |
| true | 0.7 | 0.3 |
| false | 0.01 | 0.99 |



What about $P(B,e,a,j,m)$?

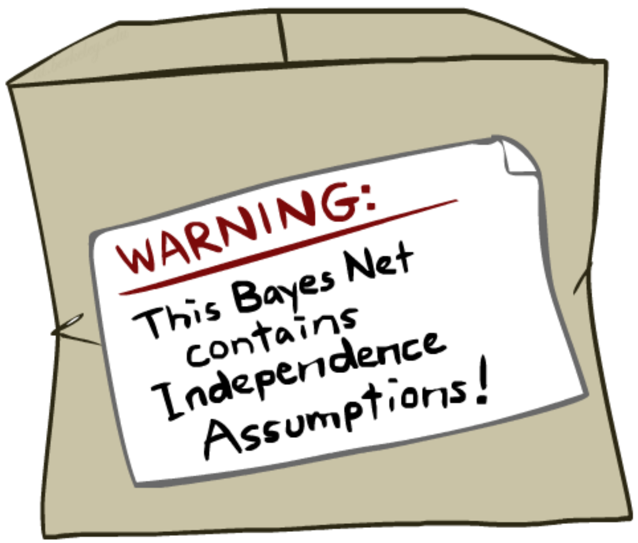
# Questions?

live and on sli.do #cse473

# Sparse Bayes Nets

DEFINITION

- Suppose
  - $n$ variables
  - Maximum range size is $d$
  - Maximum number of parents is $k$
- Full Joint Distribution: $O(d^n)$
- Bayes Net: $O(n \cdot d^k)$
  - Linear scaling with $n$ as long as causal structure is local
- Often $O(n \cdot d^k) \ll O(d^n)$



# Bayes Net Semantics

DEFINITION

Bayes nets encode joint distributions as products of conditional distributions on each variable

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i | \text{parents}(X_i))$$

- Compare with chain rule: $P(X_1, \dots, X_n) = \prod_i P(X_i | X_1, \dots, X_{i-1})$
  - Assume $X_1, \dots, X_n$ sorted in topological order, so $\text{parents}(X_i) \subseteq X_1, \dots, X_{i-1}$
  - Therefore, bayes net asserts the conditional independence:

$$P(X_i | X_1, \dots, X_{i-1}) = P(X_i | \text{parents}(X_i))$$

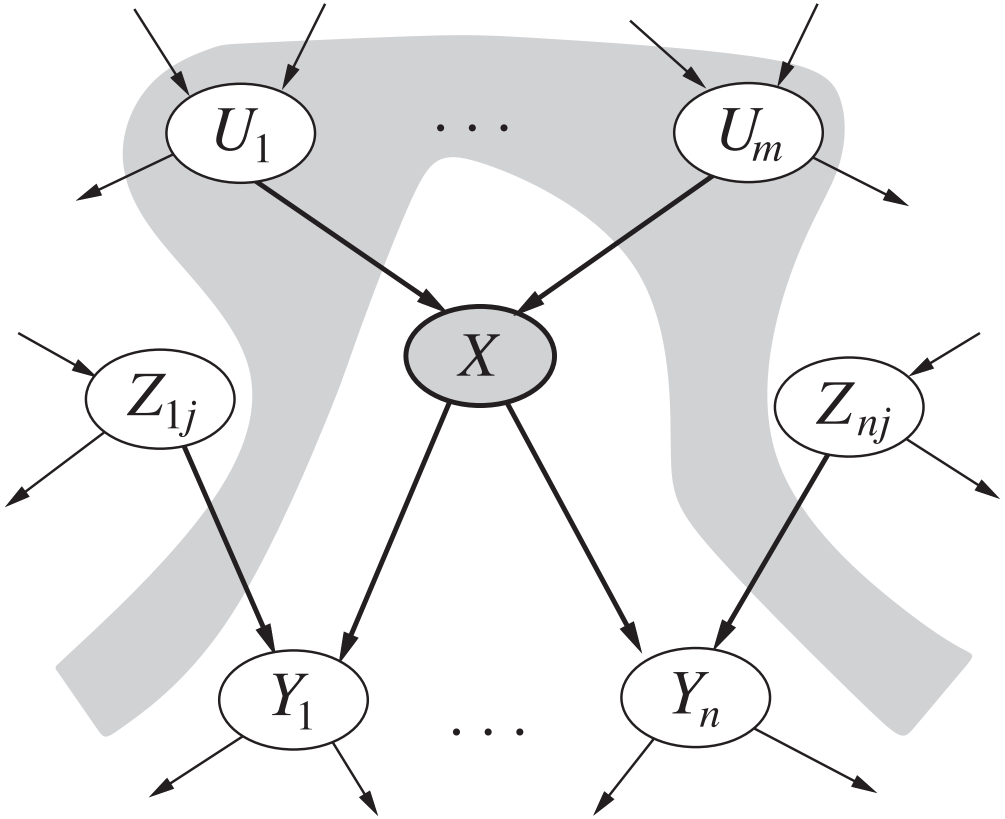
# Questions?

live and on sli.do #cse473

# Conditional Independence

FOUNDATIONS

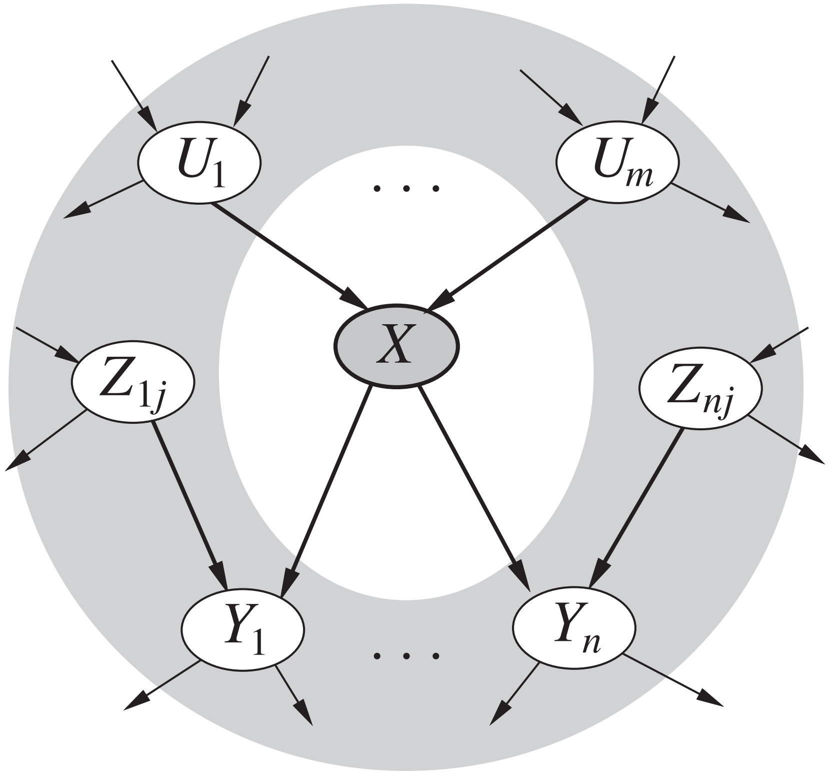
Every variable is conditionally independent of its non-descendants, given its parents



# Markov Blanket

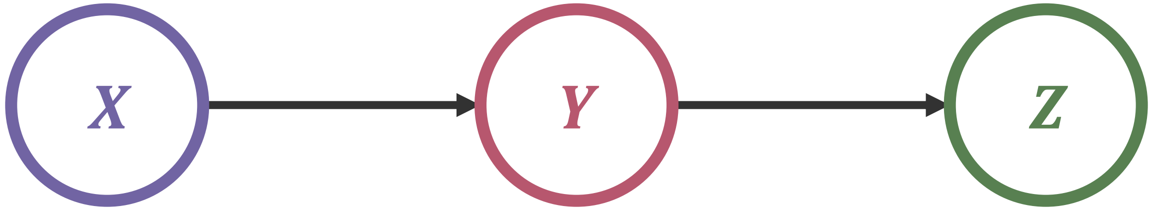
FOUNDATIONS

- Compare with Markov Assumption
  - "The future is independent of the past, given the present"
- Similar Formulation
  - **Markov blanket** consists of a node's parents, children, and children's other parents
  - Every variable is conditionally independent of all other variables, given its Markov blanket



# Questions?

live and on sli.do #cse473




$X \perp Z \mid Y$

*(not absolute independence)*

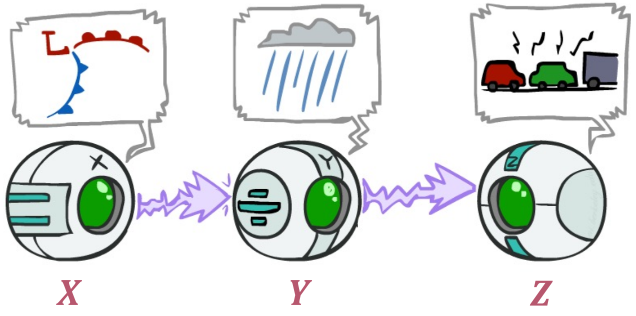
**$X$ independent of $Z$?**

# *absolute independence?*

# Node Triples

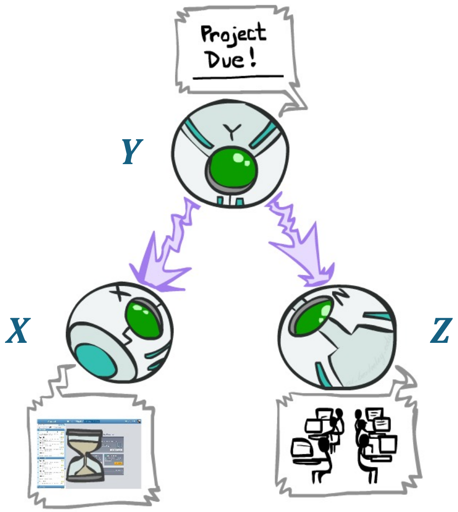
CONTEXT

*How can three nodes be arranged?*



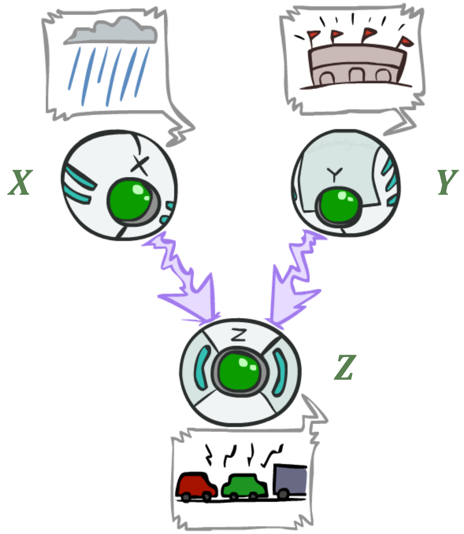

**Causal Chain**

$$P(X,Y,Z) = P(X) \cdot P(Y|X) \cdot P(Z|Y)$$




**Common Cause**

$$P(X,Y,Z) = P(Y) \cdot P(X|Y) \cdot P(Z|Y)$$




**Common Effect**

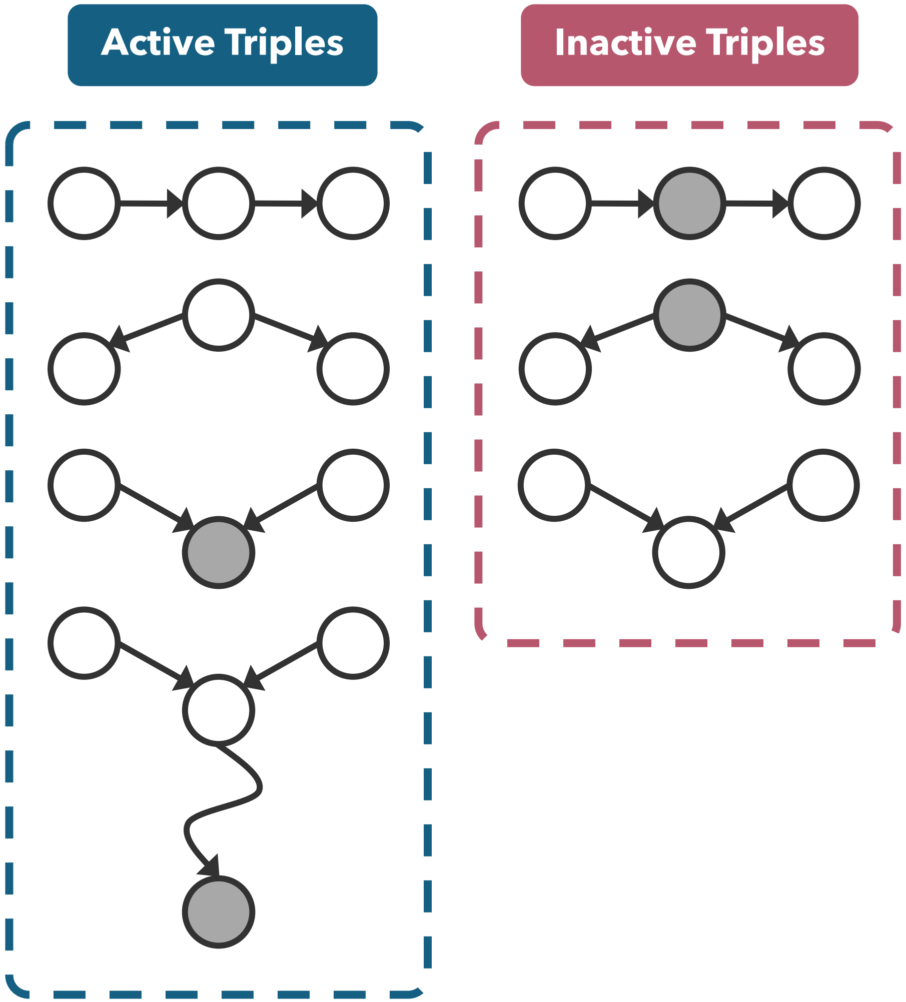
$$P(X,Y,Z) = P(X) \cdot P(Y) \cdot P(Z|X,Y)$$

# D-Separation

DEFINITION

**D-separation** is an algorithm for checking (conditional) independence of two nodes

- **Query**: $X_i \perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$ ?
- **Test**:
  - Check all *undirected* paths between $X_i$ and $X_j$:
    - If any paths are **active**: $X_i \not\perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$
    - If all paths are **inactive**: $X_i \perp X_j \mid \{X_{k_1}, \dots, X_{k_n}\}$



# Questions?

live and on sli.do #cse473

# that's it for today!

**SUMMARY**

- Implications of independence and conditional independence
- Bayes nets encode joint distributions by tracking conditional probabilities

**UPCOMING**

- How to compute conditional probabilities of queries given evidence?

**REMINDERS**

- Complete **Practice Problem 11**
- Fill out the **Mid-Quarter Feedback Survey**
- Do **Homework 2** (due 7.23)
- Do **Project 3** (due 7.30)