

Artificial Intelligence

CS 165A

Apr 20, 2022

Instructor: Prof. Yu-Xiang Wang

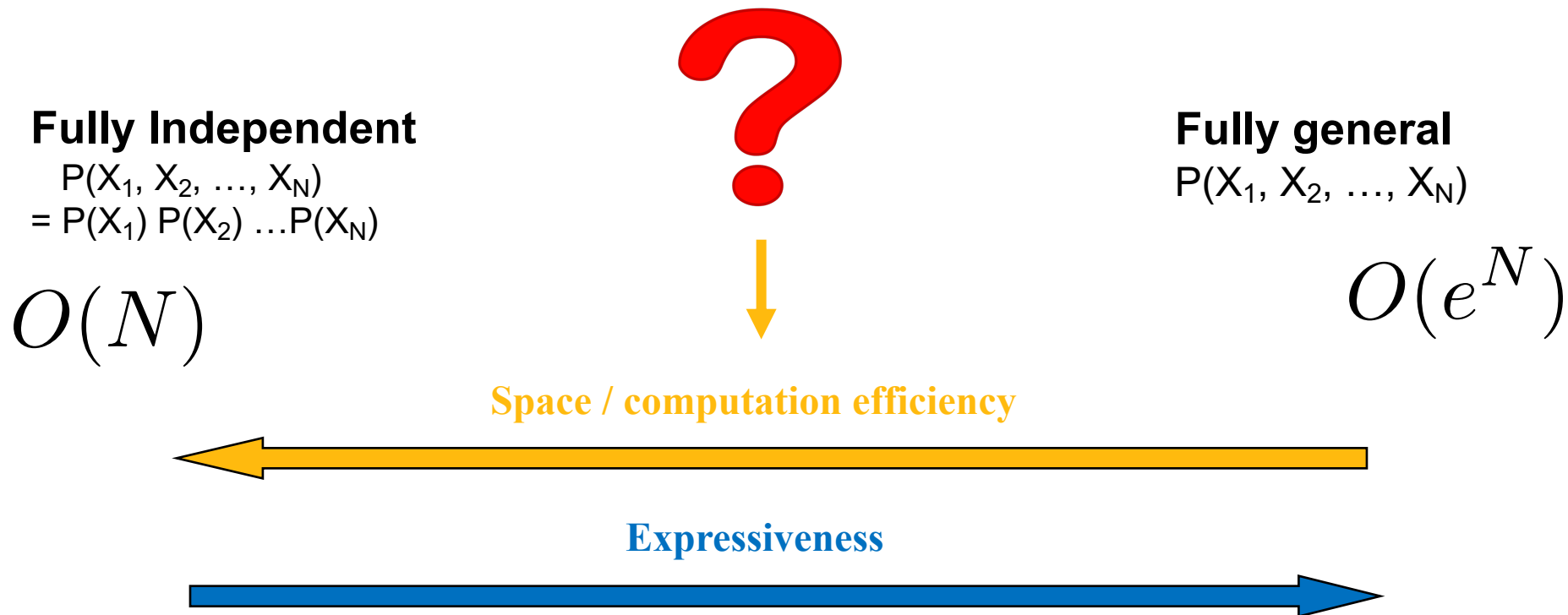
T
o
d
a
y

- Factorization and conditional independence
- Bayesian Network Examples
- Conditional Independence

Notes on Project 1

- Autograder updated.
 - Now the test cases should all make sense.
- Extra TA office hours this week
 - See the announcement on Ed Discussion
- You may come to me too! I have only 2 students who showed up in my OH this week.
- Great contribution to leaderboard from a few students. 95% now. Beating TA baseline at 86%.
 - Please keep it coming! Surprise me!

Recap: Modeling the world with joint probability distributions



Idea:

1. Independent groups of variables?
2. Conditional independences?

Recap: ChatGPT and Language Models

- Next word predictions vs joint probability distribution

$$P(X_1, \dots, X_N) = P(X_1)P(X_2|X_1)P(X_3|X_{1:2}) \cdot \dots \cdot P(X_N|X_{1:N-1})$$

- Instead one models

$$P_{\theta}(X_t | X_{t-\tau}, X_{t-\tau+1}, \dots, X_{t-1})$$

Modelling ideas:

- Bounded memory (conditional independence!)
- Make it time-invariant (parameter sharing)
- Make it parametric (use / learn features of words)

This lecture

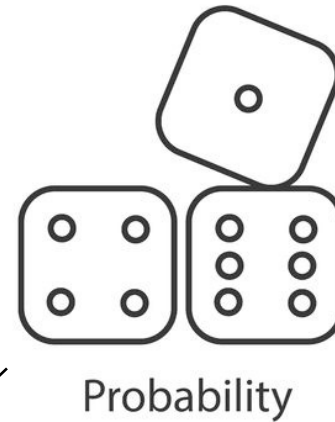
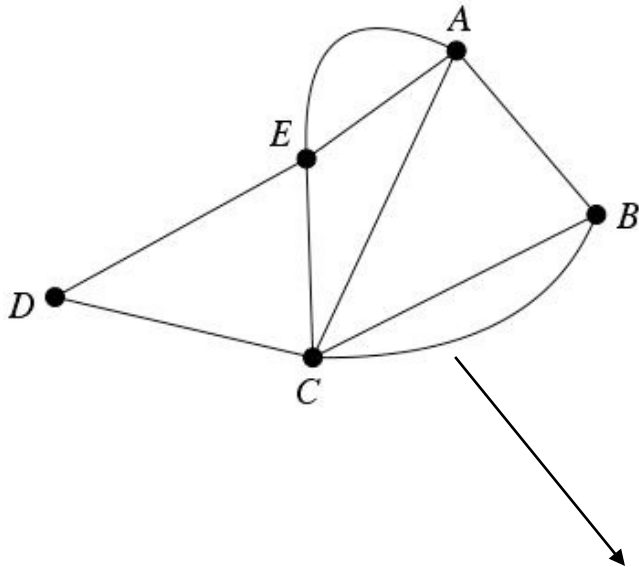
- Probabilistic Graphical Models / Bayesian Network
- Conditional independence
- Reading off Conditional Independences from a Bayesian Network
 - d-separation
 - Bayes Ball algorithm
 - Markov Blanket

When given a problem with many variables.

[CS165A Lecture attendance,
Project 1, Project 2, Project 3,
Readings, Ed Discussion, Final Grade,
Weather, Job Offer]

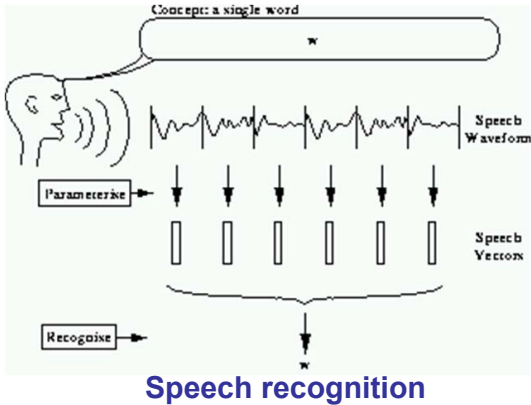
How do we know **which conditional independence(s)**
to include in the joint distribution?

Graphical models come out of the marriage of graph theory and probability theory

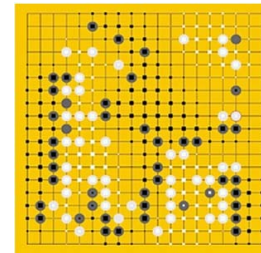


Directed Graph => Bayesian Networks / Belief Networks
Undirected Graph => Markov Random Fields

Used as a modeling tool. Many applications!



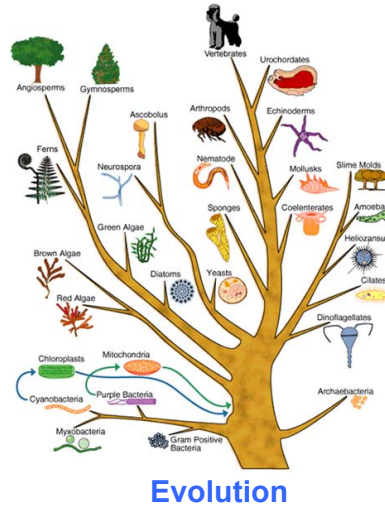
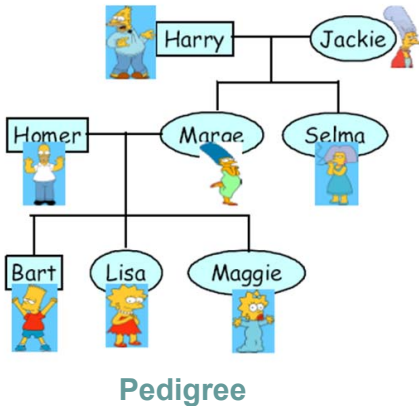
Computer vision



Games



Robotic control

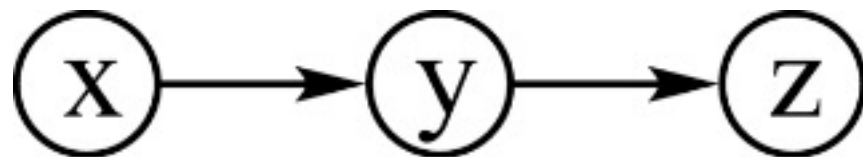


Planning

© Eric Xing @ CMU, 2005-2014

Two ways to think about Graphical Models

- A particular factorization of a joint distribution
 - $P(X,Y,Z) = P(X) P(Y|X) P(Z|Y)$
- A collection of conditional independences
 - $\{ X \perp Z \mid Y, \dots \}$



Represented using a graph!

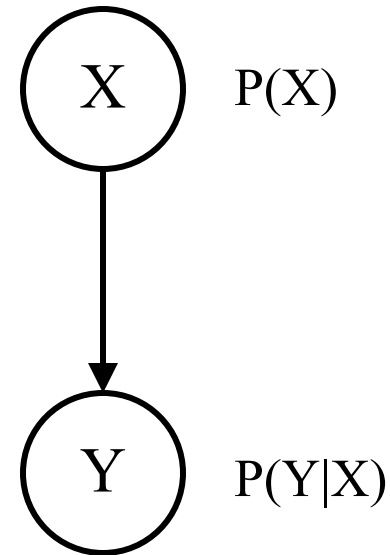
Bayesian Networks (BayesNets)

a.k.a. directed graphical models, Belief nets etc.

- Bayesian Network
 - A data structure (depicted as a graph) that represents the dependence among variables and allows us to concisely specify the joint probability distribution
 - The graph itself is known as an “influence diagram”
- A BayesNet is a **directed acyclic graph** where:
 - The nodes represent the set of random variables (one node per random variable)
 - Arcs between nodes represent *influence*, or *causality*
 - A link from node X to node Y means that X “directly influences” Y
 - Each node has a *conditional probability table* (CPT) that defines **P(node | parents)**

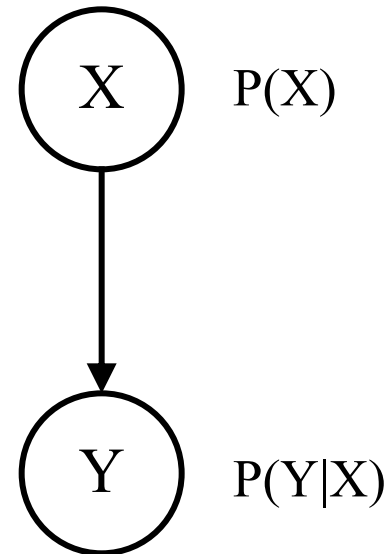
Example

- Random variables X and Y
 - X – It is raining
 - Y – The grass is wet
- X has an *effect* on Y
Or, Y is a *symptom* of X
- Draw two nodes and link them
- Define the CPT for each node
 - $P(X)$ and $P(Y | X)$
- Typical use: we observe Y and we want to query $P(X | Y)$
 - Y is an *evidence variable*
 - X is a *query variable*



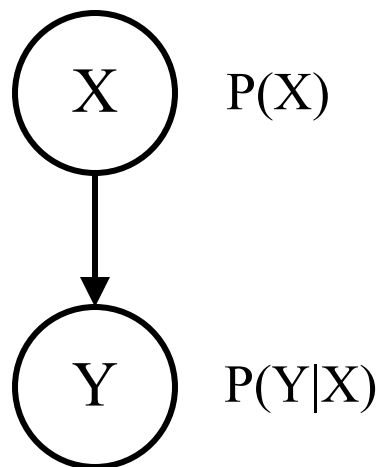
We can write everything we want as a function of the CPTs. Try it!

- What is $P(X | Y)$?
 - Given that we know the CPTs of each node in the graph

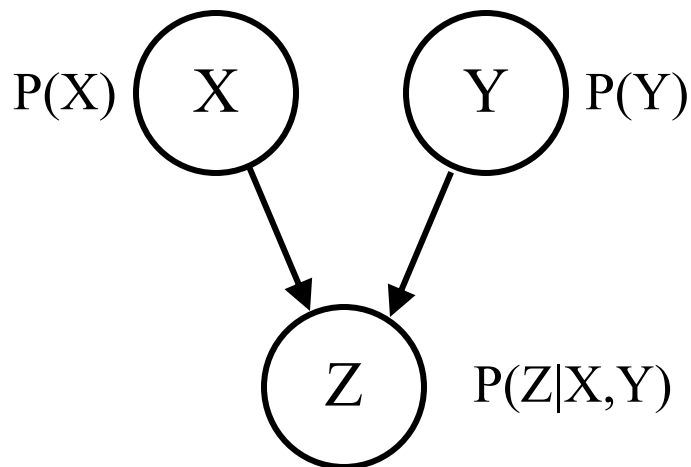


BayesNets represent the joint probability

- The joint probability function can be calculated directly from the network
 - It's the product of the CPTs of all the nodes
 - $P(\text{var}_1, \dots, \text{var}_N) = \prod_i P(\text{var}_i | \text{Parents}(\text{var}_i))$



$$P(X, Y) = P(X) P(Y|X)$$



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

Three steps in modelling with BayesNet

1. Choose variables in the environments, represent them as nodes.
2. Connect the variables by inspecting the “direct influence”: cause-effect
3. Fill in the probabilities in the CPTs.

Example: Modelling with BayesNet

I'm at work and my neighbor John called to say my home alarm is ringing, but my neighbor Mary didn't call. The alarm is sometimes triggered by minor earthquakes. Was there a burglar at my house?

- Random (boolean) variables:
 - JohnCalls, MaryCalls, Earthquake, Burglar, Alarm
- The belief net shows the causal links
- This defines the joint probability
 - $P(\text{JohnCalls}, \text{MaryCalls}, \text{Earthquake}, \text{Burglar}, \text{Alarm})$
- What do we want to know?

$$P(\mathbf{B} \mid \mathbf{J}, \neg \mathbf{M})$$

How should we connect the nodes? (3 min discussion)

Burglary

Earthquake

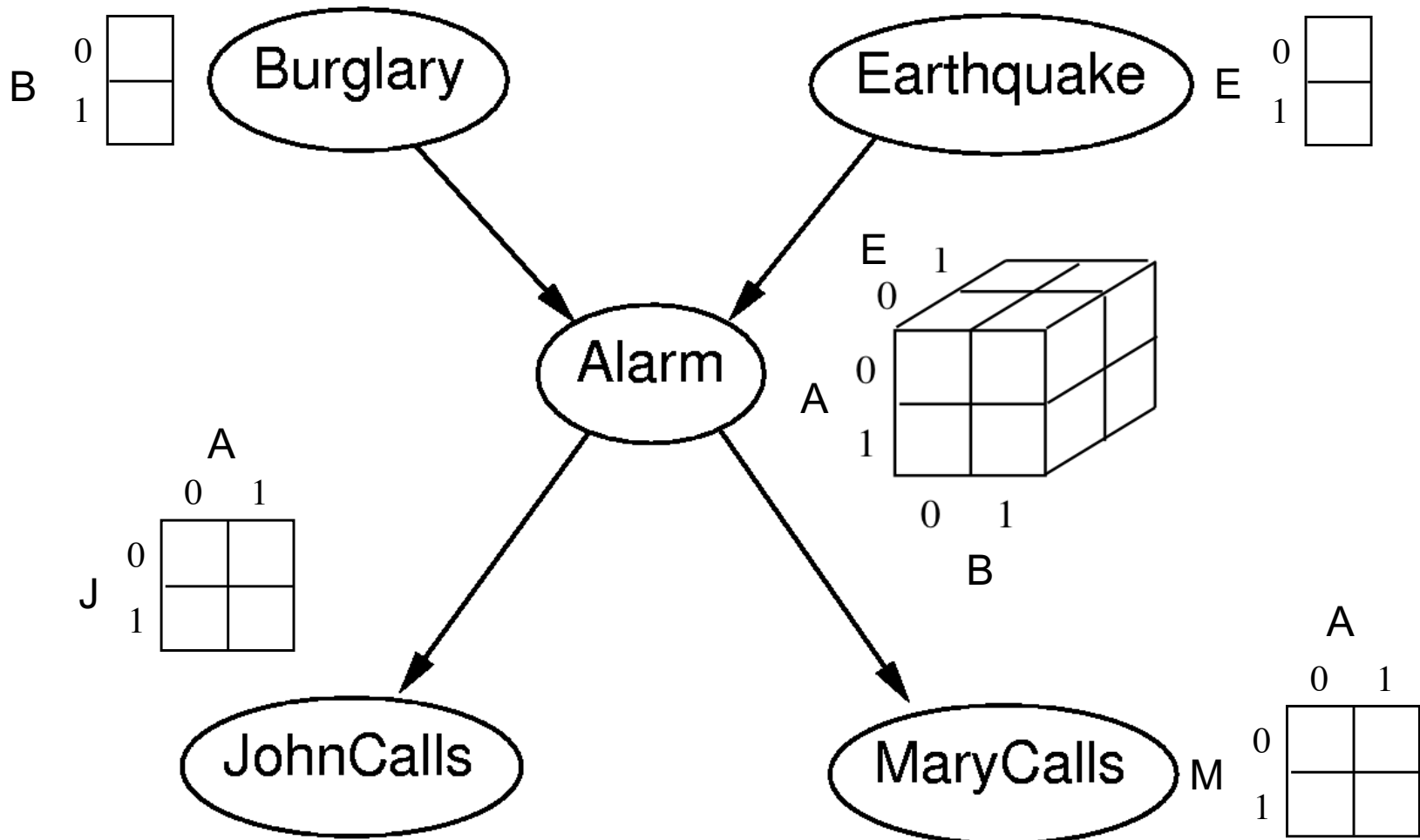
Alarm

JohnCalls

MaryCalls

Links and CPTs?

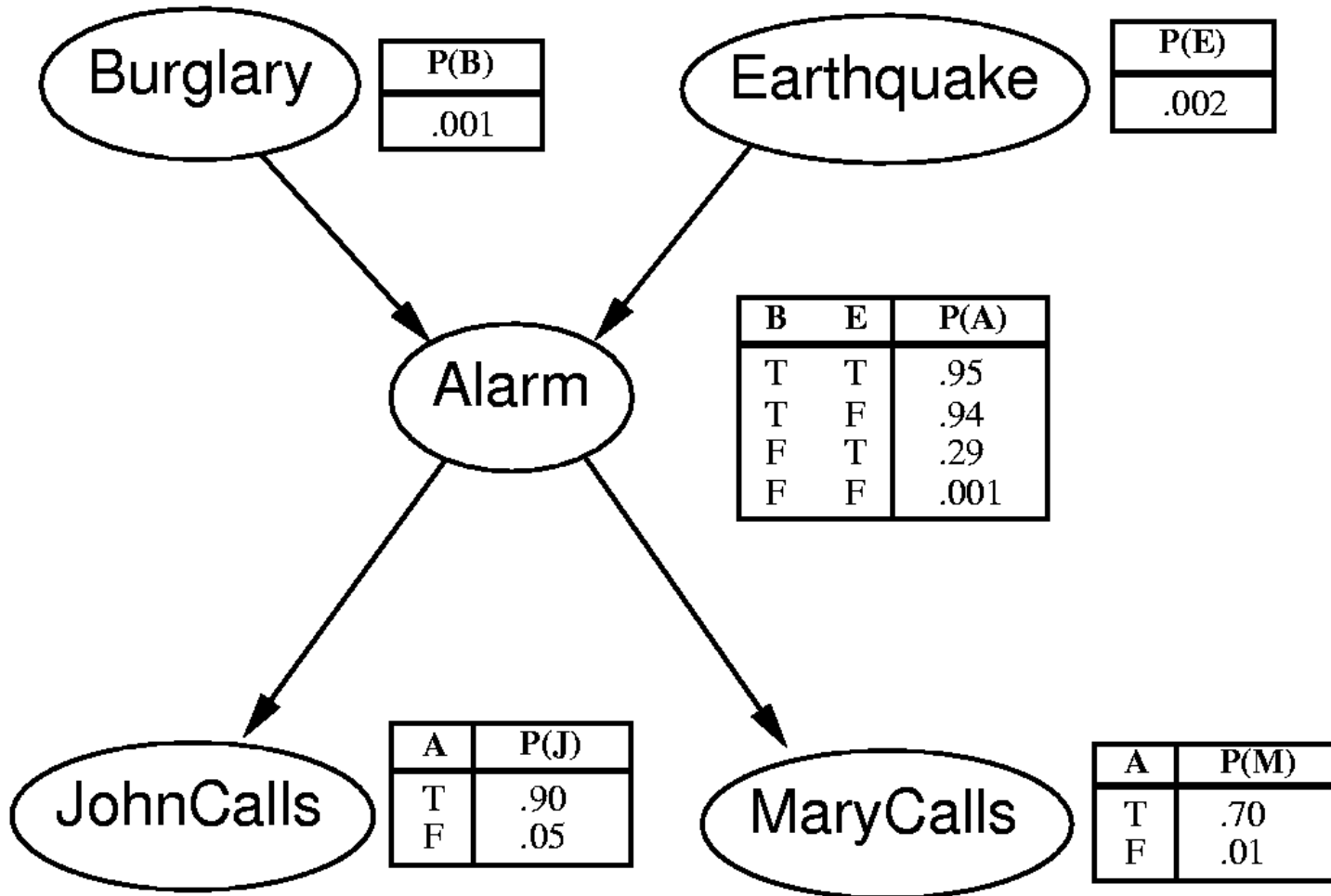
What are the CPTs? What are their dimensions?



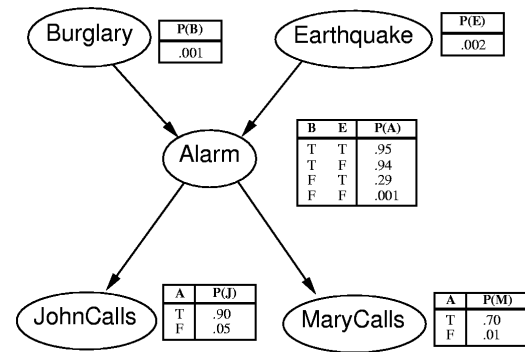
Question: How to fill values into these CPTs?

Ans: Specify by hands. Learn from data (e.g., MLE).

Example



Joint probability? $P(J, \neg M, A, B, \neg E)$?



Calculate $P(J, \neg M, A, B, \neg E)$

Read the joint pf from the graph:

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

Plug in the desired values:

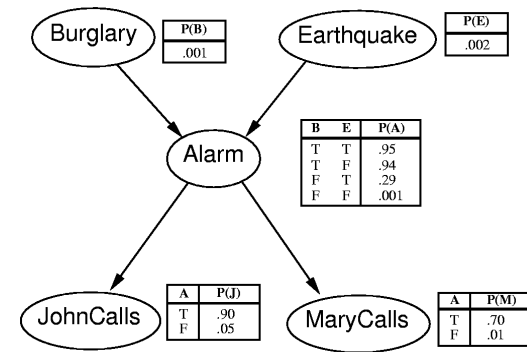
$$\begin{aligned} P(J, \neg M, A, B, \neg E) &= P(B) P(\neg E) P(A|B, \neg E) P(J|A) P(\neg M|A) \\ &= 0.001 * 0.998 * 0.94 * 0.9 * 0.3 \\ &= 0.0002532924 \end{aligned}$$

How about $P(B | J, \neg M)$?

Remember, this means $P(B=\text{true} | J=\text{true}, M=\text{false})$

Calculate $P(B | J, \neg M)$

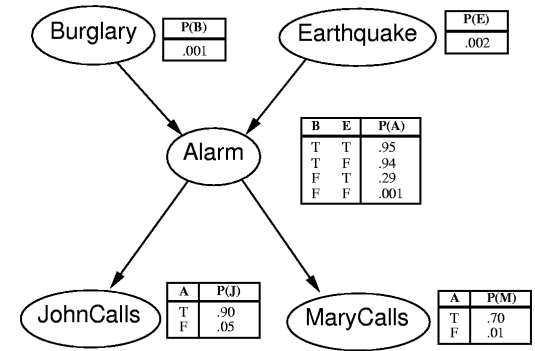
$$P(B | J, \neg M) = \frac{P(B, J, \neg M)}{P(J, \neg M)}$$



By marginalization:

$$\begin{aligned}
 & \sum_i \sum_j P(J, \neg M, A_i, B, E_j) \\
 = & \frac{\sum_i \sum_j \sum_k P(J, \neg M, A_i, B_j, E_k)}{\sum_i \sum_j \sum_k P(B_j)P(E_k)P(A_i | B_j, E_k)P(J | A_i)P(\neg M | A_i)}
 \end{aligned}$$

Variable elimination algorithm



$$P(B | J, \neg M) = \frac{P(B, J, \neg M)}{P(J, \neg M)}$$

$$= \frac{\sum_i \sum_j P(B)P(E_j)P(A_i | B, E_j)P(J | A_i)P(\neg M | A_i)}{\sum_i \sum_j \sum_k P(B_j)P(E_k)P(A_i | B_j, E_k)P(J | A_i)P(\neg M | A_i)}$$

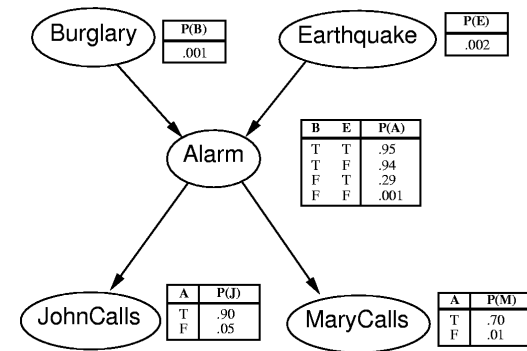
*Exchange the order of summation and product

Quick checkpoint

- Bayesian Network as a modelling tool
- By inspecting the cause-effect relationships, we can draw directed edges based on our domain knowledge
- The product of the CPTs give the joint distribution
 - We can calculate $P(A | B)$ for any A and B
 - The factorization makes it computationally more tractable

What else can we get?

Example: Conditional Independence



- Conditional independence is seen here
 - $P(\text{JohnCalls} \mid \text{MaryCalls}, \text{Alarm}, \text{Earthquake}, \text{Burglary}) = P(\text{JohnCalls} \mid \text{Alarm})$
 - So JohnCalls is independent of MaryCalls, Earthquake, and Burglary, given Alarm
- Does this mean that an earthquake or a burglary do not influence whether or not John calls?
 - No, but the influence is already accounted for in the Alarm variable
 - JohnCalls is conditionally independent of Earthquake, but not marginally independent of it

***This conclusion is independent to values of CPTs!**

Question

If X and Y are independent, are they therefore independent given any variable(s)?

I.e., if $P(X, Y) = P(X) P(Y)$ [i.e., if $P(X|Y) = P(X)$], can we conclude that

$$P(X | Y, Z) = P(X | Z)?$$

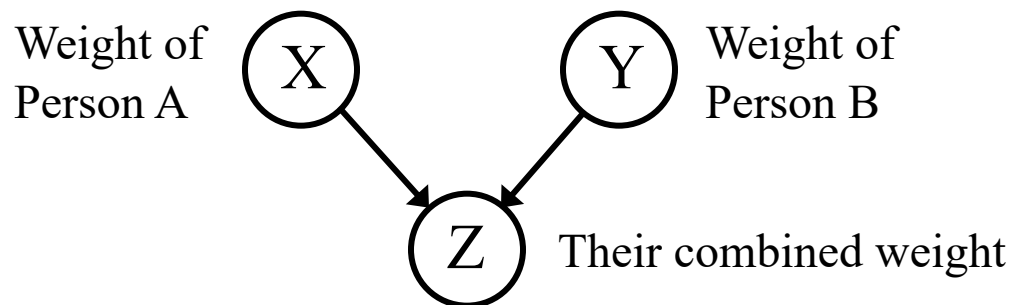
Question

If X and Y are independent, are they therefore independent given any variable(s)?

I.e., if $P(X, Y) = P(X) P(Y)$ [i.e., if $P(X|Y) = P(X)$], can we conclude that

$$P(X | Y, Z) = P(X | Z)?$$

The answer is **no**, and here's a counter example:



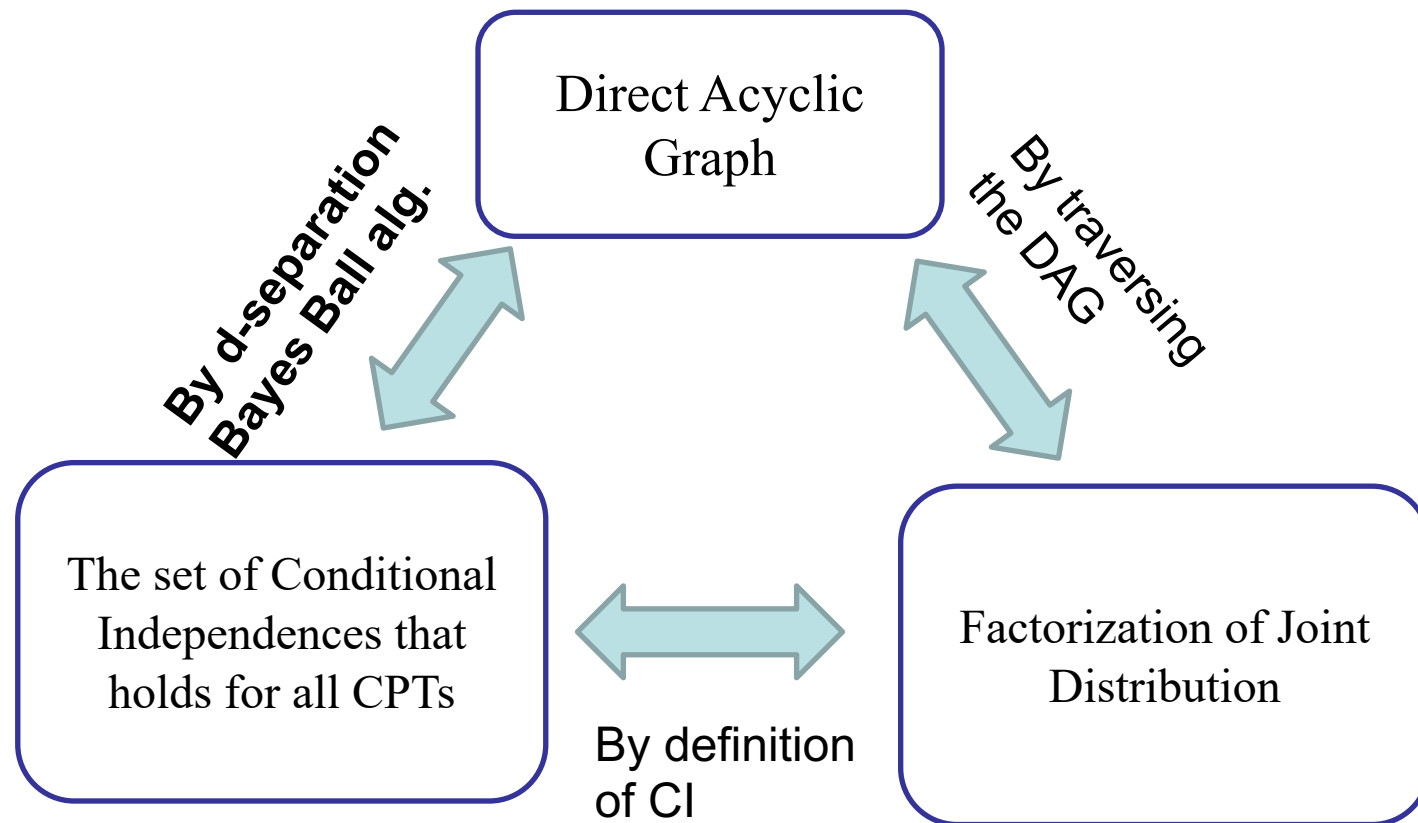
$$P(X | Y) = P(X)$$
$$P(X | Y, Z) \neq P(X | Z)$$

Note: Even though Z is a deterministic function of X and Y , it is still a random variable with a probability distribution

***Again: This conclusion is independent to values of CPTs!**

Big question: Is there a general way that we can answer questions about conditional independences by just inspecting the graphs?

- Turns out the answer is “Yes!”



Intuition: the graph and the edges controls the information flow, if there is no path that the information can flow from one-node to another, we say these two nodes are independent..

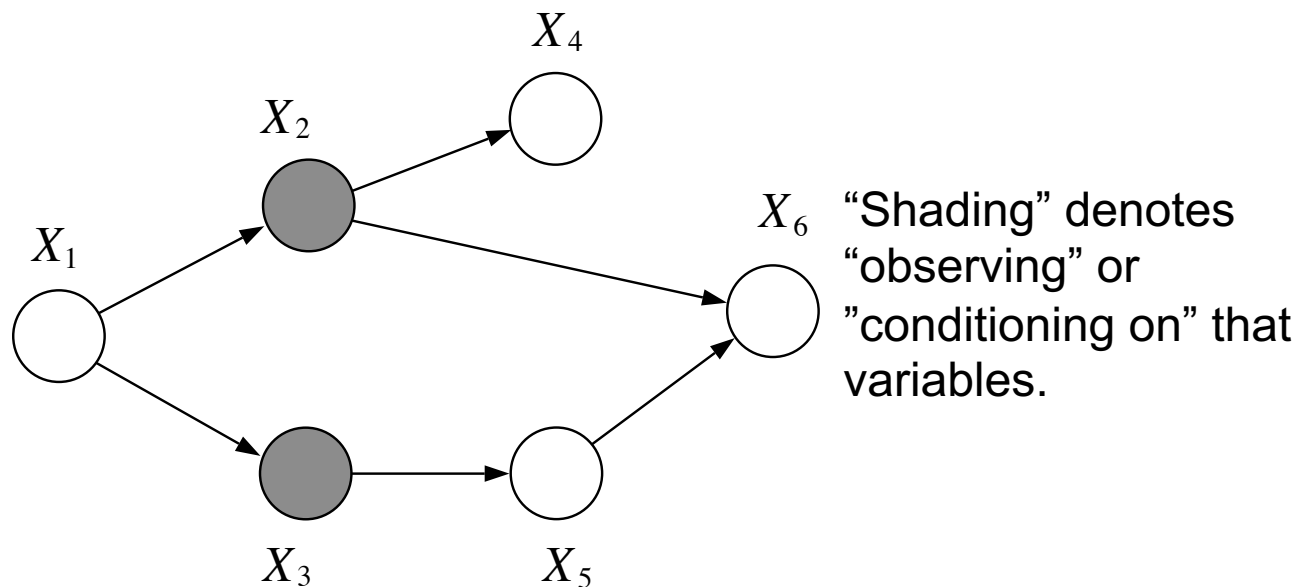
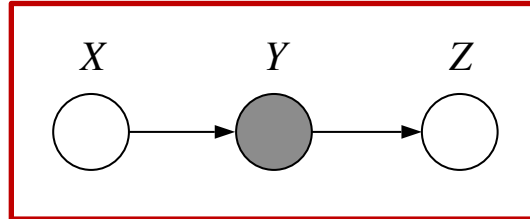
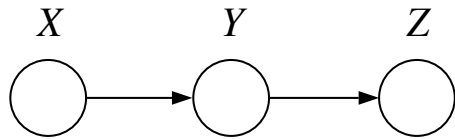


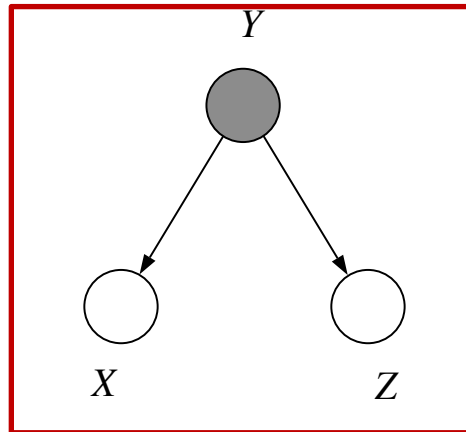
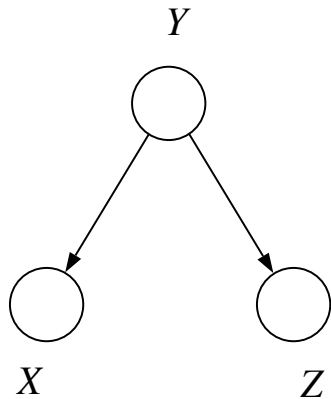
Figure 2.3: The nodes X_2 and X_3 separate X_1 from X_6 .

d-separation in three canonical graphs



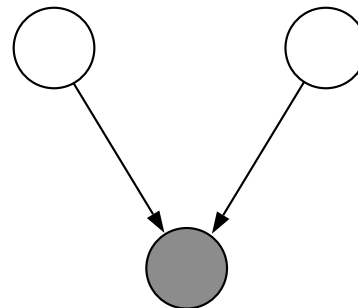
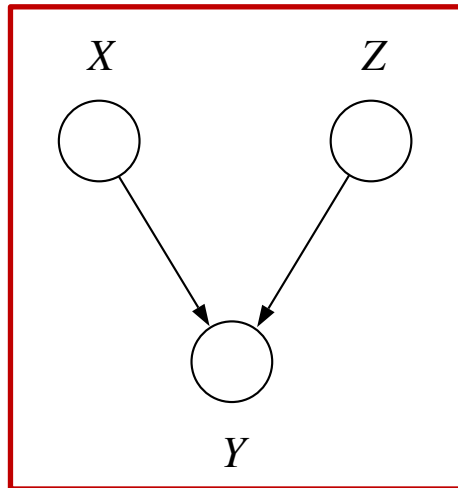
$$X \perp Z \mid Y$$

“**Chain:** X and Z are d-separated by the observation of Y.”



$$X \perp Z \mid Y$$

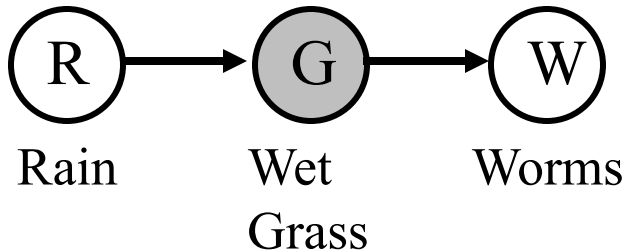
“**Fork:** X and Z are d-separated by the observation of Y.”



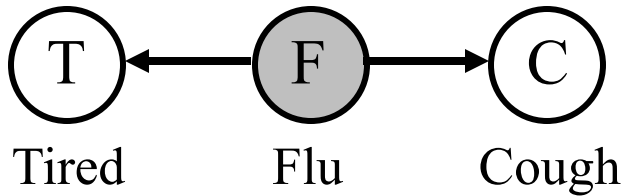
$$X \perp Z$$

“**Collider:** X and Z are d-separated by NOT observing Y **nor any descendants** of Y.”

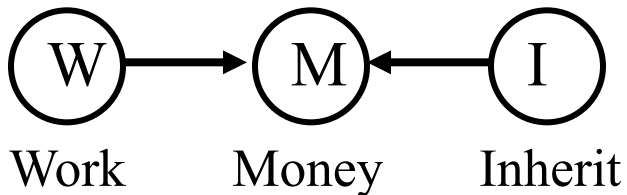
Examples



$$P(W \mid R, G) = P(W \mid G)$$



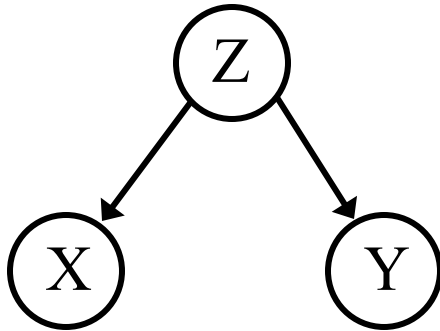
$$P(T \mid C, F) = P(T \mid F)$$



$$P(W \mid I, M) \neq P(W \mid M)$$

$$P(W \mid I) = P(W)$$

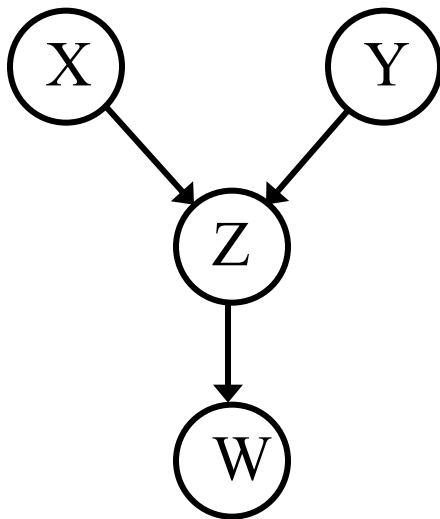
Examples



X – wet grass
Y – rainbow
Z – rain

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

Are X and Y ind.? Are X and Y cond. ind. given...?



X – rain
Y – sprinkler
Z – wet grass
W – worms

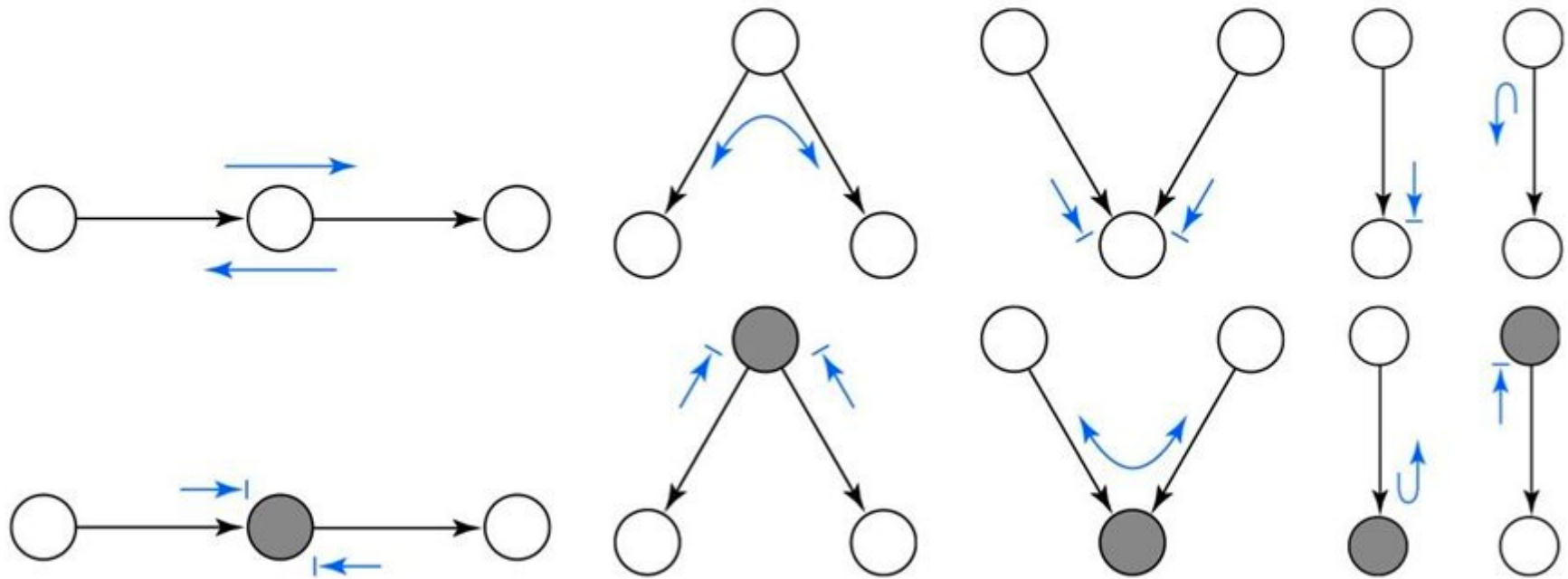
$$P(X, Y) = P(X) P(Y)$$
$$P(X | Y, Z) \neq P(X | Z)$$
$$P(X | Y, W) \neq P(X | W)$$

The Bayes Ball algorithm

- Let \mathbf{X} , \mathbf{Y} , \mathbf{Z} be “*groups*” of nodes / set / subgraphs.
- Shade nodes in \mathbf{Y}
- Place a “ball” at each node in \mathbf{X}
- Bounce balls around the graph according to **rules**
- If no ball reaches any node in \mathbf{Z} , then declare

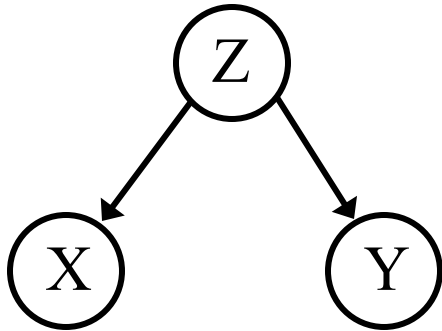
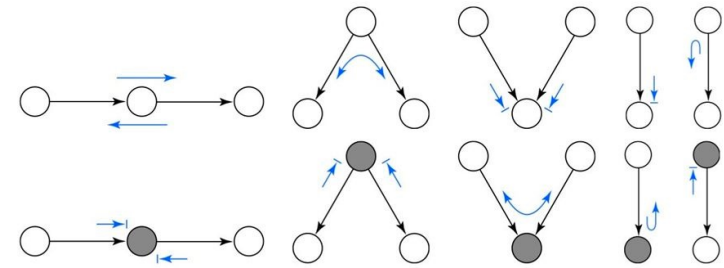
$$\mathbf{X} \perp \mathbf{Z} \mid \mathbf{Y}$$

The Ten Rules of Bayes Ball Algorithm



Please read [Jordan PGM Ch. 2.1]
to learn more about the Bayes Ball algorithm

Examples (revisited using Bayes-ball alg)

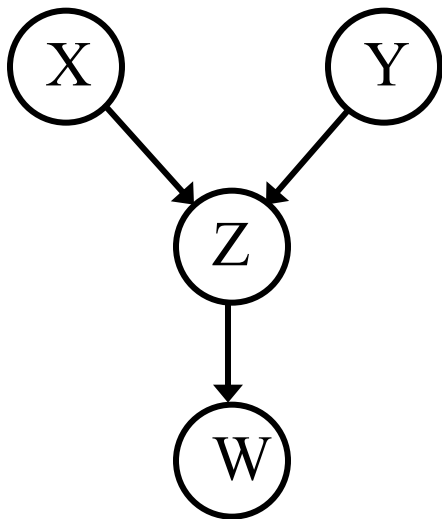


X – wet grass
 Y – rainbow
 Z – rain

$$P(X, Y) \neq P(X) P(Y)$$

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

Are X and Y ind.? Are X and Y cond. ind. given...?



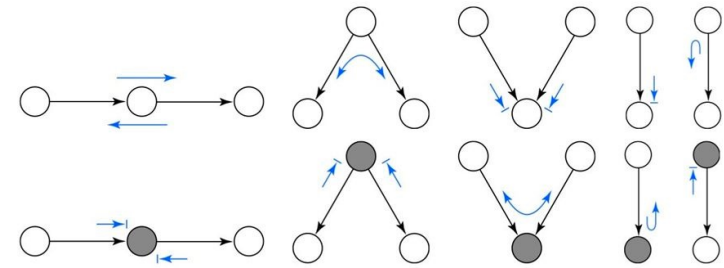
X – rain
 Y – sprinkler
 Z – wet grass
 W – worms

$$P(X, Y) = P(X) P(Y)$$

$$P(X | Y, Z) \neq P(X | Z)$$

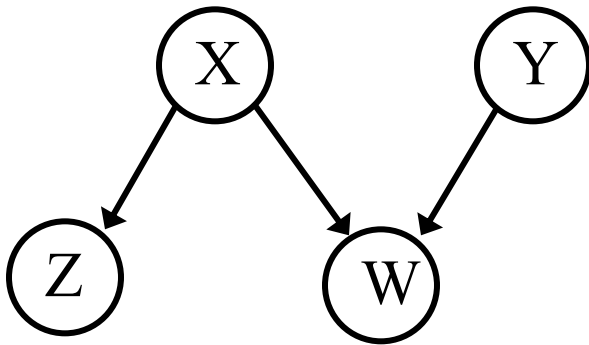
$$P(X | Y, W) \neq P(X | W)$$

Examples (3 min work)

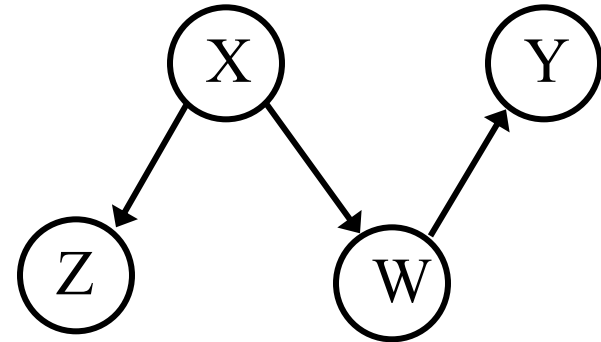


Are X and Y independent?

Are X and Y conditionally independent given Z?

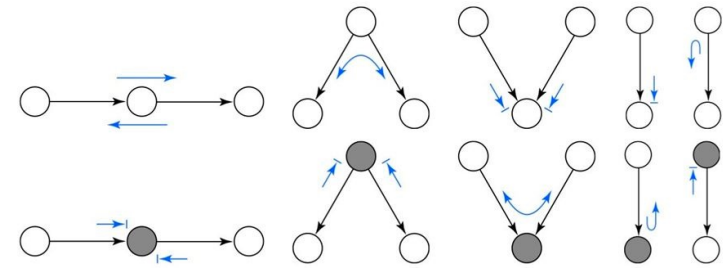


X – rain
 Y – sprinkler
 Z – rainbow
 W – wet grass



X – rain
 Y – sprinkler
 Z – rainbow
 W – wet grass

Conditional Independence



- Where are conditional independences here?

Radio and Ignition, given Battery?

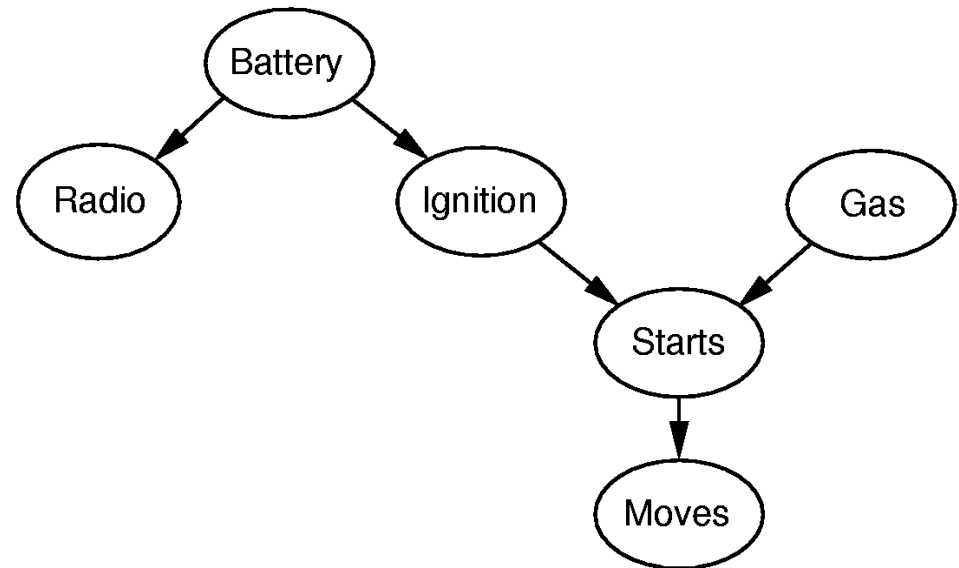
Radio and Starts, given Ignition?

Gas and Radio, given Battery?

Gas and Radio, given Starts?

Gas and Radio, given nil?

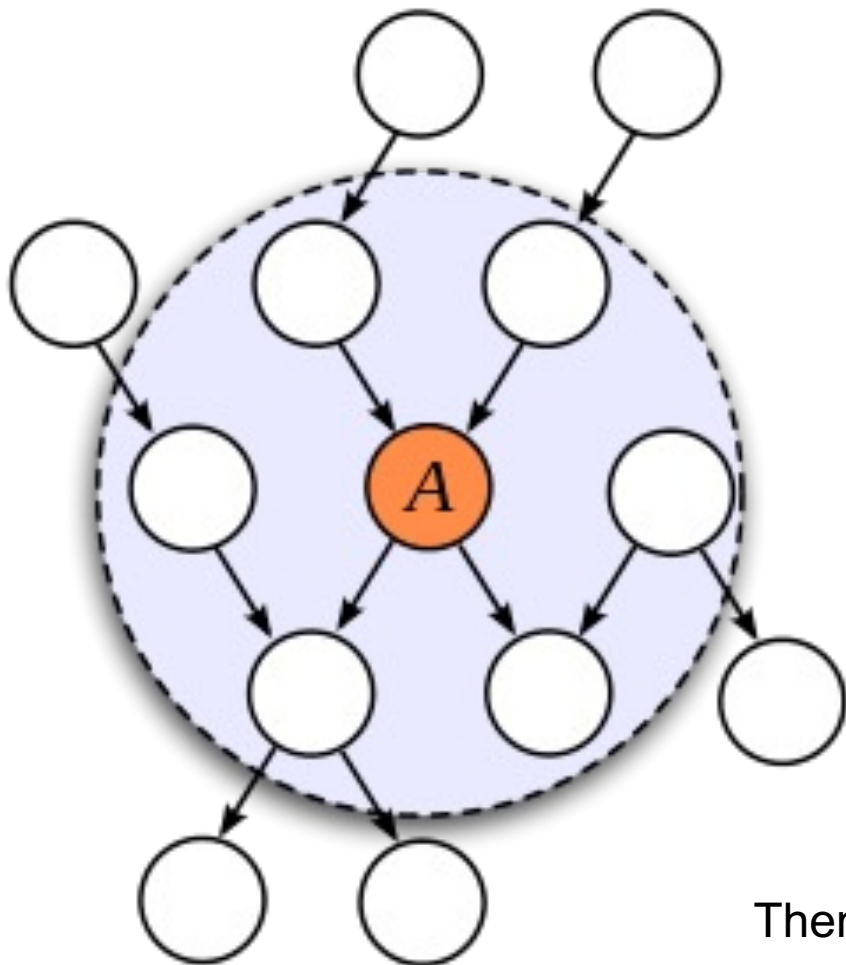
Gas and Battery, given Moves?



Quick checkpoint

- Reading conditional independences from the DAG itself.
- d-separation
 - Three canonical graphs: Chain, Fork, Collider
- Bayes ball algorithm for determining whether $\mathbf{X} \perp \mathbf{Z} \mid \mathbf{Y}$
 - Bounce the ball from any node in X by following the ten rules
 - If any ball reaches any node in Z, then return “False”
 - Otherwise, return “True”

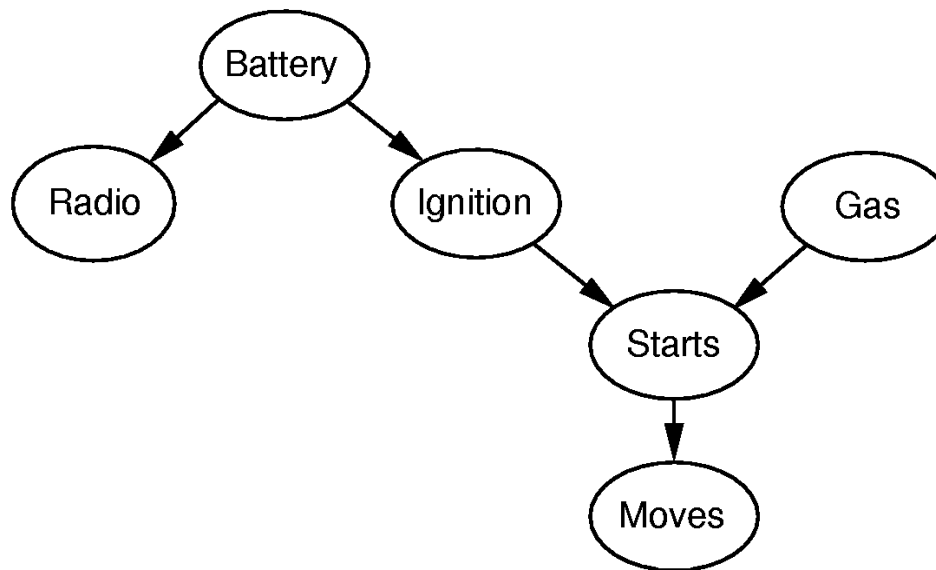
An alternative view: Markov Blankets



1. Parents
2. Children
3. Children's other parents

Then A is d-separated from everything else.

Example: Markov Blankets



- Question: What is the Markov Blanket of ...
 - “Ignition”:
 - “Starts”:

Why are conditional independences important?

- Helps the developer (or the user) verify the graph structure
 - Are these variables really independent?
 - Do I need more/fewer edges in the graphical model?
- Statistical tests for (Conditional) Independence
 - Hilbert-Schmidt Independence Criterion (not covered)
- Hints on computational efficiencies
- Shows that you understand BNs...

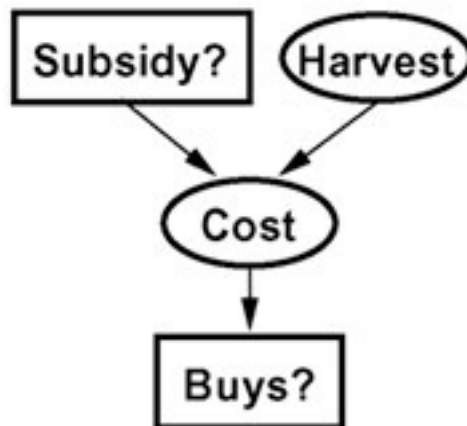
Inference in Bayesian networks

- We've seen how to compute any probability from the Bayesian network
 - This is *probabilistic inference*
 - $P(\text{Query} \mid \text{Evidence})$
 - Since we know the joint probability, we can calculate anything via marginalization
 - $P(\text{Query}, \text{Evidence}) / P(\text{Evidence})$
- However, things are usually not as simple as this
 - Structure is large or very complicated
 - **Calculation by marginalization is often intractable**
 - Bayesian inference is NP hard in space and time!!
 - (Details in AIMA Ch 13.4)

Inference in Bayesian networks (cont.)

- So in all but the most simple BNs, probabilistic inference is not really done just by marginalization
- Instead, there are practical algorithms for doing approximate probabilistic inference
 - Recall a similar argument in surrogate losses in ML
- Markov Chain Monte Carlo, Message Passing / Loopy Belief Propagation
 - Active area of research!
- We won't cover these probabilistic inference algorithms though.... (Read AIMA Ch 13.5)

One more thing: Continuous Variables?



- Dimension check: What are the shapes of the CPTs?
- Discretize? Very large CPT..
- Usually, we parametrize the conditional distribution.
 - e.g., $P(\text{Cost} \mid \text{Harvest}) = \text{Poisson}(\theta^T \text{Harvest})$

Summary of today's lecture

- Encode knowledge / structures using a DAG
 - How to check conditional independence algebraically by the factorizations?
 - How to read off conditional independences from a DAG
 - d-separation, Bayes Ball algorithm, Markov Blanket
 - Remarks on BN inferences and continuous variables
- (More examples, e.g., Hidden Markov Models, see AIMA 13.3)**

Additional resources about PGM

- Recommended: Ch.2 Jordan book. AIMA Ch. 12-13.
- More readings (if you need to use PGMs in the future):
 - Koller's PGM book: <https://www.amazon.com/Probabilistic-Graphical-Models-Daphne-Koller/dp/B007YXTT12>
 - Probabilistic programming: <http://probabilistic-programming.org/wiki/Home>
- Software for PGMs and modeling and inference:
 - Stan: <https://mc-stan.org/>
 - JAGS: <http://mcmc-jags.sourceforge.net/>

Upcoming lectures

- Lecture 7: Problem solving by search
 - Lecture 8: Search algorithms
 - Lecture 9: Minimax search and game playing
 - Lecture 10: Midterm review.
-
- Recommended readings on search:
 - AIMA Ch 3, Ch 5.1-5.3.