Artificial Intelligence

CS 165A

Jan 22, 2019

Instructor: Prof. Yu-Xiang Wang

Today

→ Probabilistic Reasoning /
Bayesian networks (Ch 14)

Announcement

- Machine Problem 1 will be posted tonight. (we are still working on the hand in process. Stay tuned…)

- 20% grade reduction every day for both homework and MP submission delays.

- **[Important]** You will need a COE account to submit your code
  - Create on here: https://accounts.engr.ucsb.edu/create
  - Or contact the helpdesk: help@engineering.ucsb.edu

Machine Problem 1 : AI Movie Critic

- Movie reviews -> Positive or Negative?

"Paul W.S. Mother Effing Anderson. Sick of the guy. He's the posterboy for rambunctiously stupid Hollywood movies that focus on "cool" action scenes and a completely disregarded plotline. Have you guys seen the Resident Evil movies? ...Freakin horrible. None of them have any merits of anything that is good about filmmaking other then the fact that it made it into theaters. Please Anderson... stop making movies. This is one of the worst movies to come out in this decade. Don't watch this garbage."

- You will be given a data set with users rating movies in raw text, along with labels on whether they are positive or negative.

- You must write a program that automatically classifies these text instances, including ones that are not in the data set (on our test machines).

Machine Problem 1 : AI Movie Critic

- You are supposed to implement at least two models:
  - Gaussian Naïve Bayes
  - Multinomial Naïve Bayes

- And also implement at least two feature extractor:
  - TF-IDF
  - Bag of Words

- There will be a leaderboard.
  - Explore more advanced classifiers /feature extractors (within a runtime time / space limit)
  - Get bonus points!

- **See the released MP1 doc tonight for details.**

Today

- Quick review of the last lecture
- d-separation and Markov Blankets
  - Reading conditional independences from a DAG
- Undirected Graphical models

Question

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

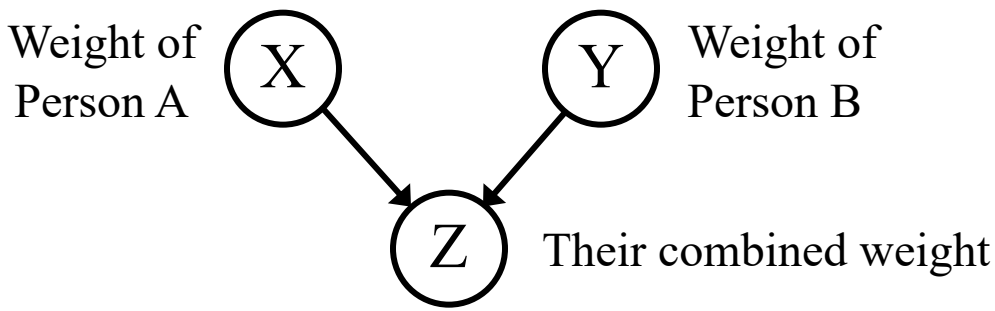
I.e., if $P(X, Y) = P(X)\,P(Y)$ [i.e., if $P(X|Y) = P(X)$], can we conclude that
$P(X \mid Y, Z) = P(X \mid Z)$?

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
$P(X \mid Y, Z) = P(X \mid Z)$?

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



$$P(X \mid Y) = P(X)$$

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

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

Belief nets

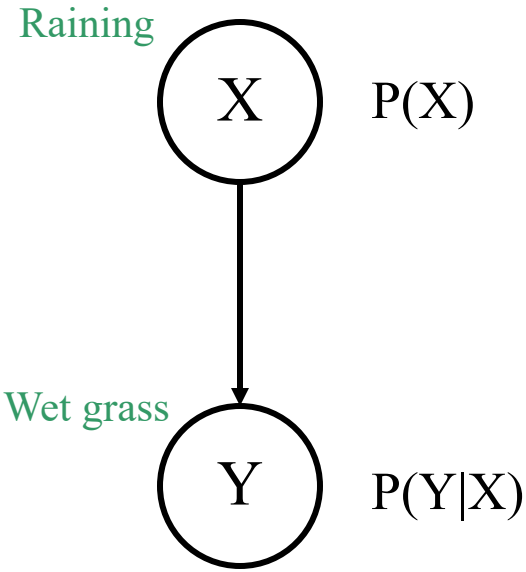
- General assumptions
  - A DAG is a reasonable representation of the influences among the variables
    - Leaves of the DAG have no direct influence on other variables
  - Implies a set of conditional independences
    - Makes Bayesian inference computationally more tractable.
  - The CPTs are relatively easy to state
    - Many can be estimated as a **empirical distribution** (a standard pattern – just specify the parameters) or as a deterministic node (direct function – logical or numerical combination – of parents)

What are belief nets for?

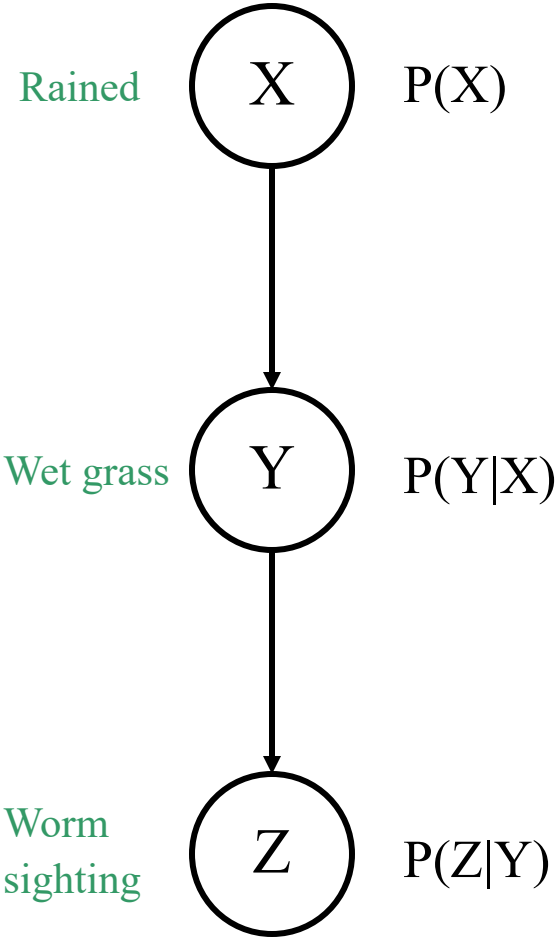
- Given the structure, we can now pose queries:
  - Typically: **P(Query | Evidence)** or **P(Cause | Symptoms)**
  - $P(X_1 \mid X_4, X_5)$
  - P(Earthquake | JohnCalls)
  - P(Burglary | JohnCalls, ¬MaryCalls)

Query variable

Evidence variables



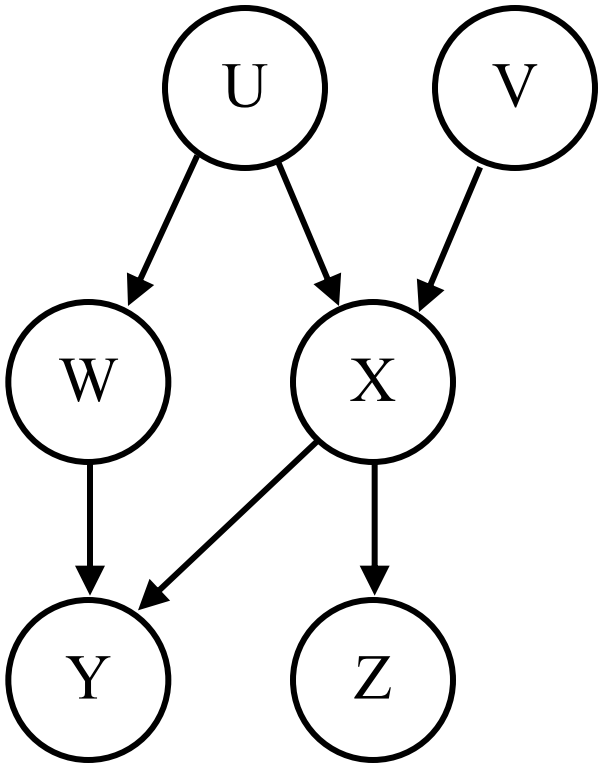
ASK P(X|Y)



ASK P(X|Z)

Quiz

1. What is the joint probability distribution of the random variables described by this belief net?
   - I.e., what is P(U, V, W, X, Y, Z)?

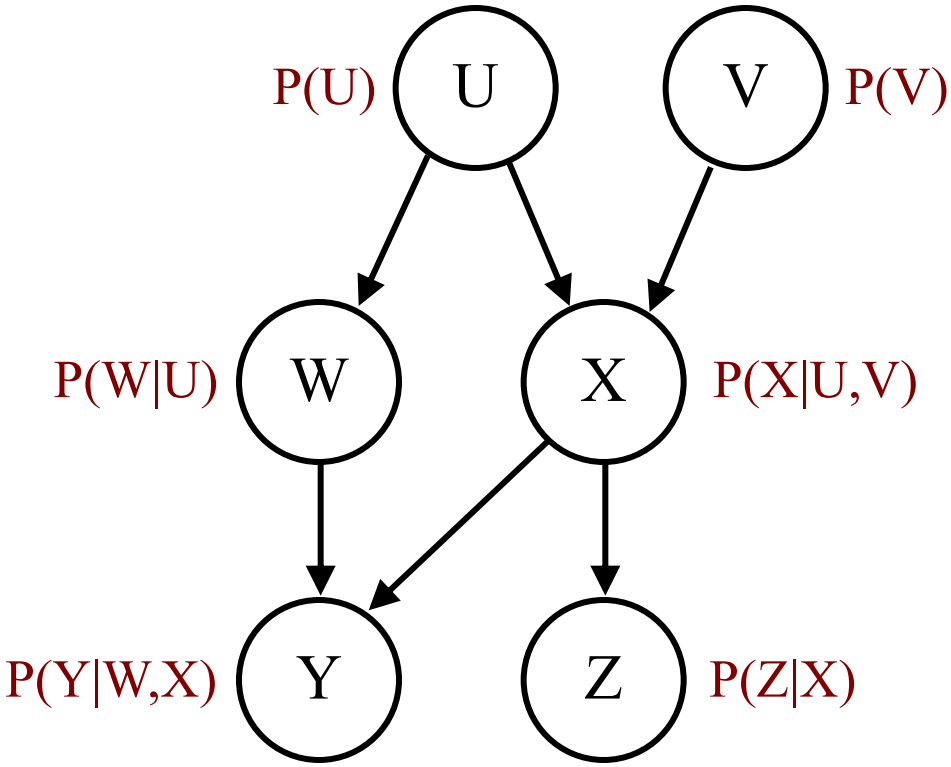


2. Variables W and X are
   a) Independent
   b) Independent given U
   c) Independent given Y

   (choose one)

3. If you know the CPTs, is it possible to compute P(Z | U)? P(U | Z)?

Review



Given this Bayesian network:

1. What are the CPTs?
2. What is the joint probability distribution of all the variables?
3. How would we calculate P(X | W, Y, Z)?

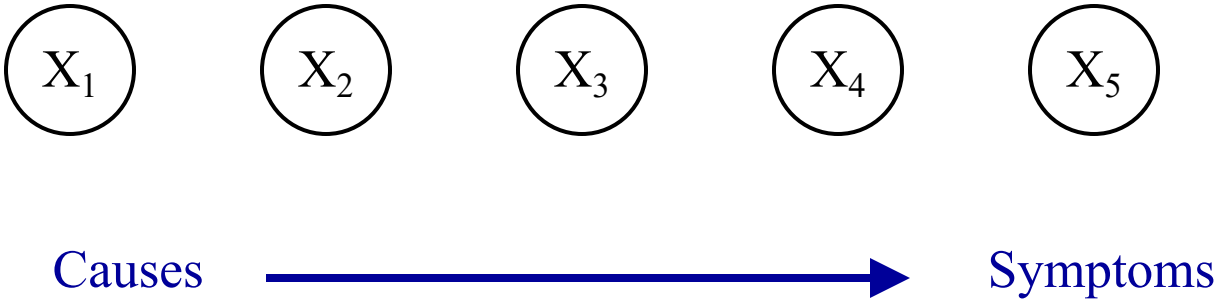
P(U,V,W,X,Y,Z) = product of the CPTs

= P(U) P(V) P(W|U) P(X|U,V) P(Y|W,X) P(Z|X)

How to construct a belief net

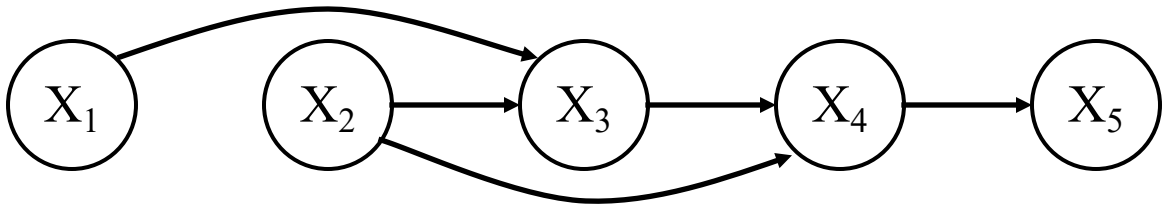
- Choose the random variables that describe the domain
  - These will be the nodes of the graph

- Choose a left-to-right ordering of the variables that indicates a general order of influence
  - "Root causes" to the left, symptoms to the right



How to construct a belief net (cont.)

- Draw arcs from left to right to indicate "direct influence" (causality) among variables
  - May have to reorder some nodes



- Define the conditional probability table (CPT) for each node
  - P(node | parents)

$P(X_1)$

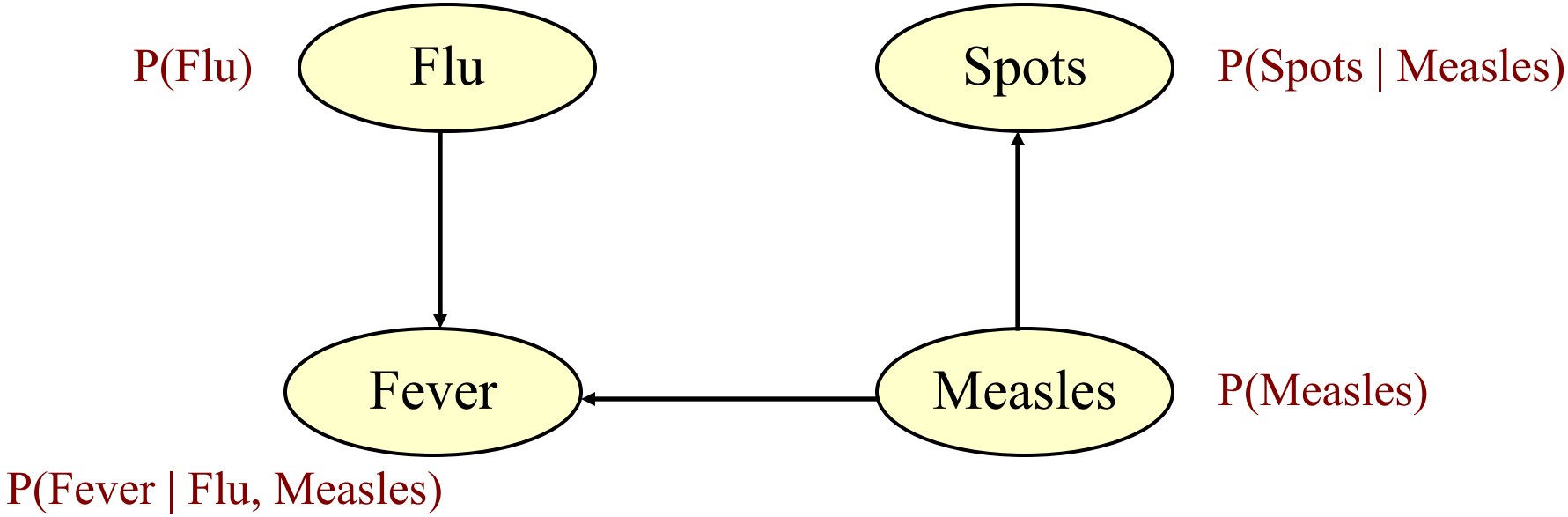
$P(X_2)$

$P(X_3 \mid X_1, X_2)$

$P(X_4 \mid X_2, X_3)$

$P(X_5 \mid X_4)$

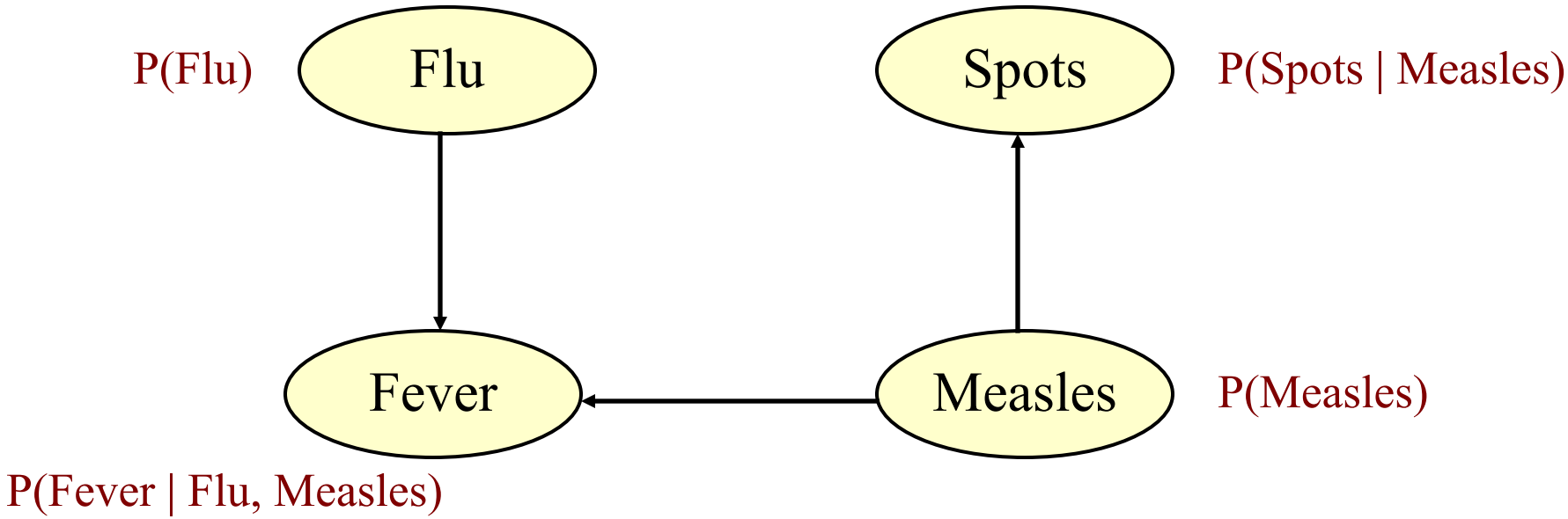
Example: Flu and measles



To create the belief net:

- Choose variables (evidence and query)
- Choose an ordering and create links (direct influences)
- Fill in probabilities (CPTs)

Example: Flu and measles



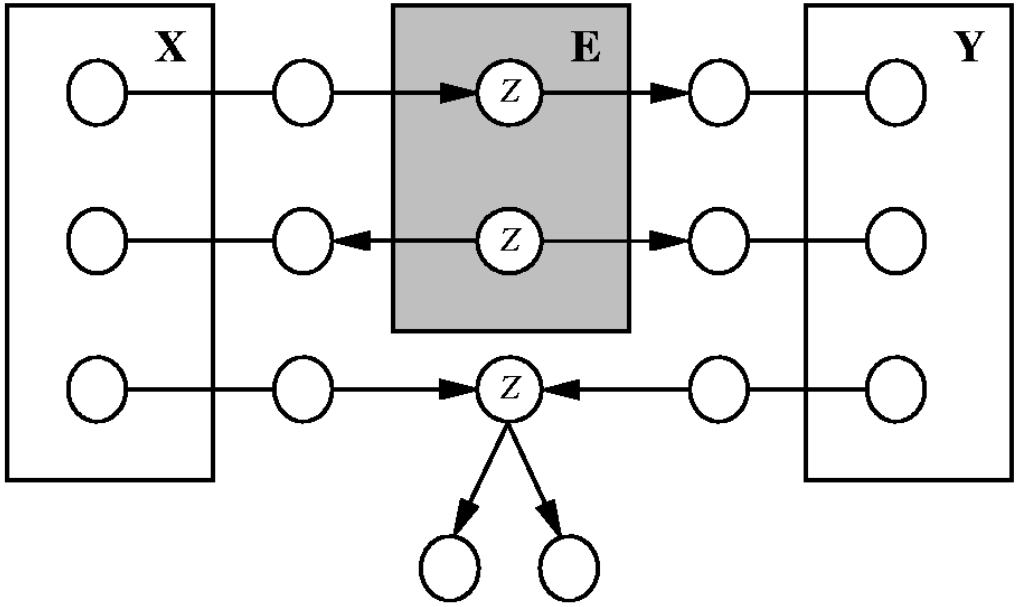
CPTs: P(Flu) = 0.01 P(Spots | Measles) = [0, 0.9]

P(Measles) = 0.001 P(Fever | Flu, Measles) = [0.01, 0.8, 0.9, 1.0]

Compute P(Flu | Fever) and P(Flu | Fever, Spots).
Are they equivalent?

Conditional Independence

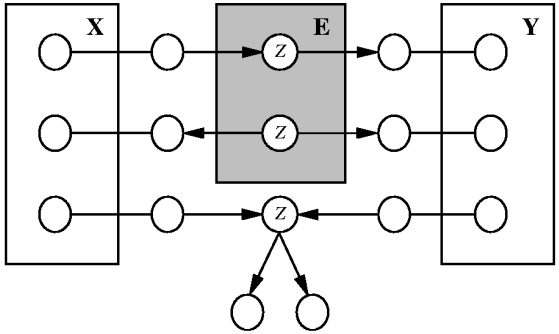
- Can we determine conditional independence of variables directly from the graph?
- A set of nodes *X* is independent of another set of nodes *Y*, given a set of (evidence) nodes *E*, if every path from *X* to *Y* is **d-separated**, or blocked, by *E*



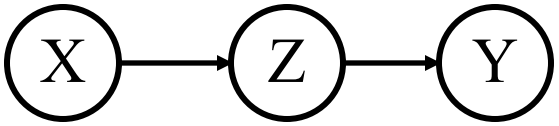
The set of nodes *E* <u>d-separates</u> sets *X* and *Y*

3 ways to block paths from *X* to *Y*, given *E*

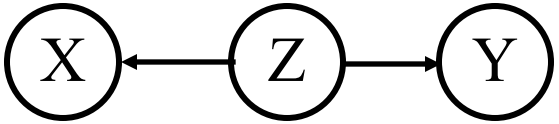
Examples



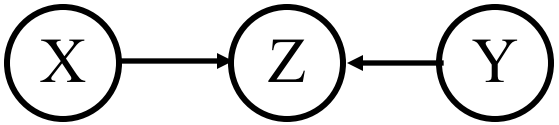
X ind. of Y? X ind. of Y given Z?



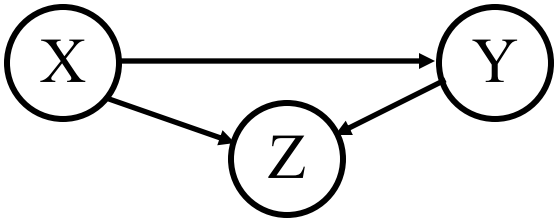
X ind. of Y? X ind. of Y given Z?



X ind. of Y? X ind. of Y given Z?



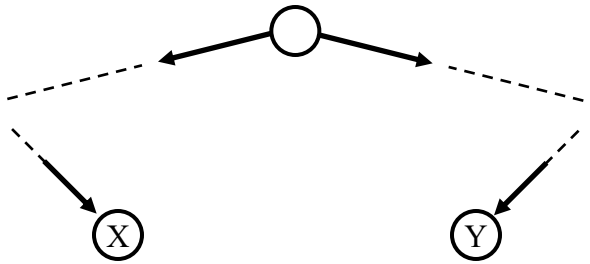
X ind. of Y? X ind. of Y given Z?



X ind. of Y? X ind. of Y given Z?

Independence (again)

- Variables X and Y are *independent* if and only if
  - P(X, Y) = P(X) P(Y)
  - P(X | Y) = P(X)
  - P(Y | X) = P(Y)
- We can determine independence of variables in a belief net directly from the graph
  - Variables ***X*** and ***Y*** are independent if they share no common ancestry
    - I.e., the set of { X, parents of X, grandparents of X, … } has a null intersection with the set of {Y, parents of Y, grandparents of Y, … }



X, Y dependent

Conditional Independence

- X and Y are *(conditionally) independent given E* iff
  - $P(X \mid Y, E) = P(X \mid E)$
  - $P(Y \mid X, E) = P(Y \mid E)$

- $\{X_1,\ldots,X_n\}$ and $\{Y_1,\ldots,Y_m\}$ are conditionally independent given $\{E_1,\ldots,E_k\}$ iff
  - $P(X_1,\ldots,X_n \mid Y_1, \ldots, Y_m, E_1, \ldots,E_k) = P(X_1,\ldots,X_n \mid E_1, \ldots,E_k)$
  - $P(Y_1, \ldots, Y_m \mid X_1,\ldots,X_n, E_1, \ldots,E_k) = P(Y_1, \ldots, Y_m \mid E_1, \ldots,E_k)$

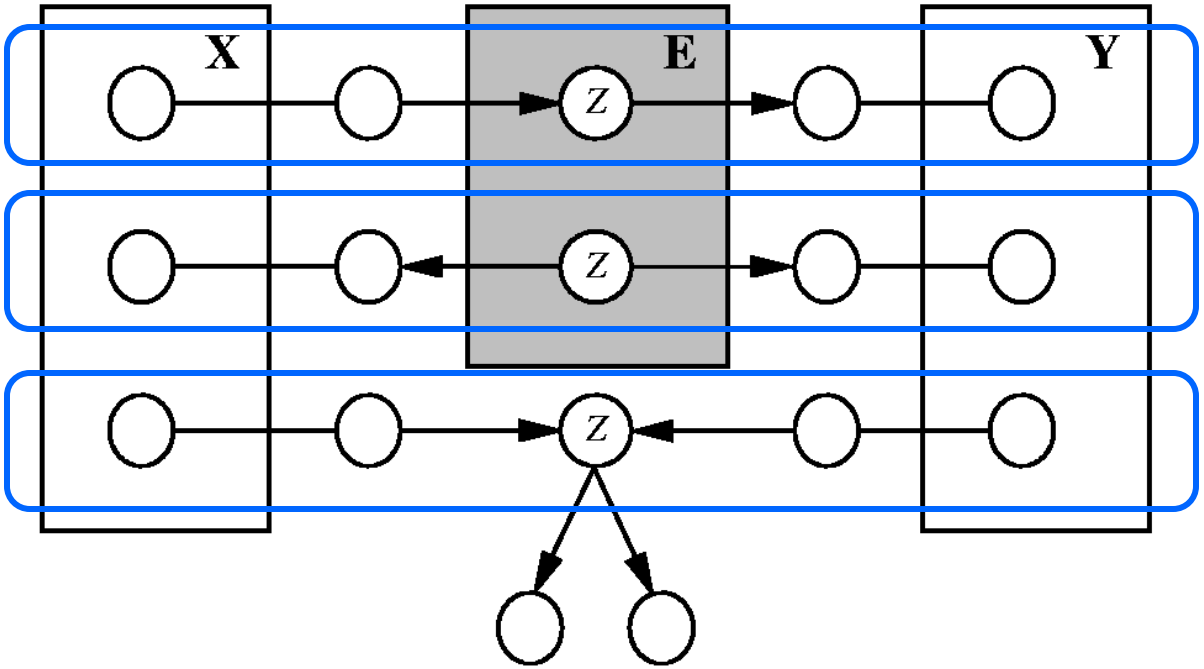
- We can determine conditional independence of variables (and sets of variables) in a belief net directly from the graph

How to determine conditional independence

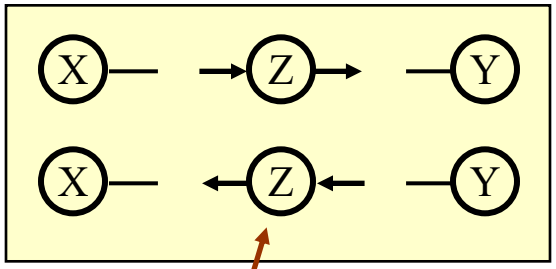
- A set of nodes ***X*** is independent of another set of nodes ***Y***, given a set of (evidence) nodes ***E***, if **every** path from X_i to Y_j is **d-separated**, or blocked
  - The set of nodes *E* <u>d-separates</u> sets *X* and *Y*

- The textbook (p. 499) mentions the *Markov blanket*, which is the same general concept

- There are three ways to **block** a path from X_i to Y_j

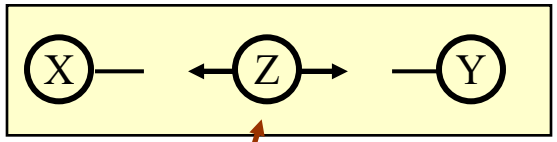


#1



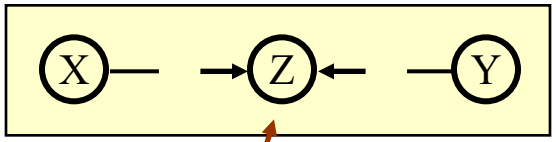
The variable Z is in E

#2



The variable Z is in E

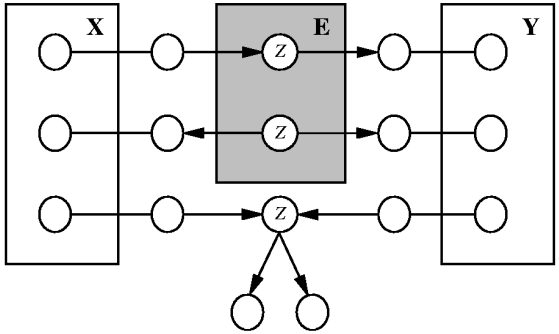
#3

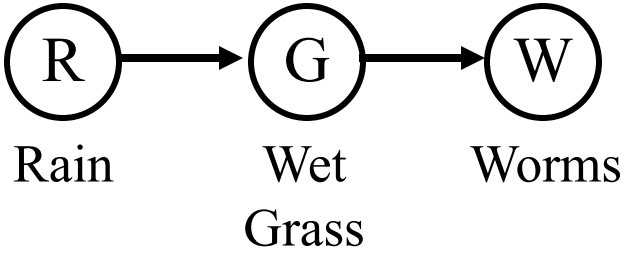


This variable is not in E!

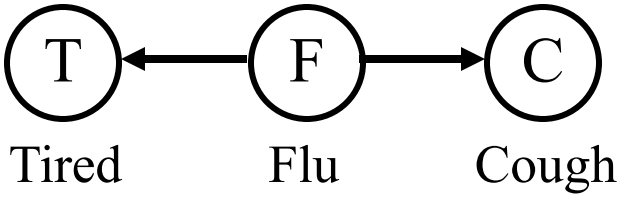
(Nor are its descendents)

Examples

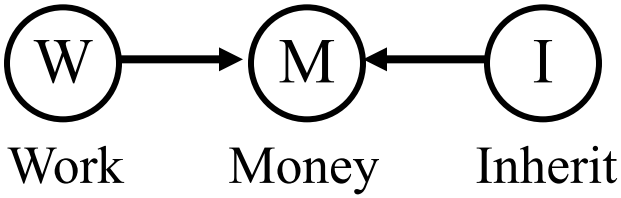




$P(W \mid R, G) = P(W \mid G)$



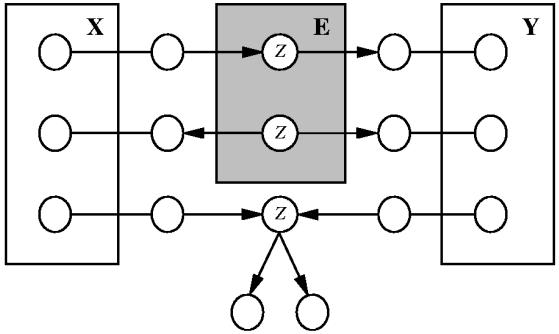
$P(T \mid C, F) = P(T \mid F)$



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

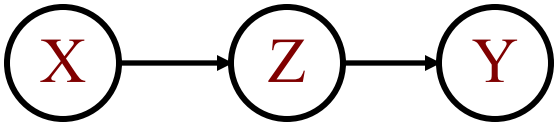
$P(W \mid I) = P(W)$

Examples



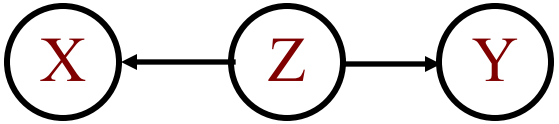
X ind. of Y? Yes

X ind. of Y given Z? Yes



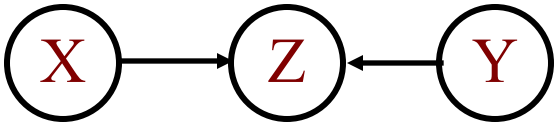
X ind. of Y? No

X ind. of Y given Z? Yes



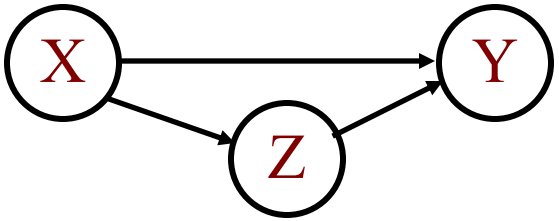
X ind. of Y? No

X ind. of Y given Z? Yes



X ind. of Y? Yes

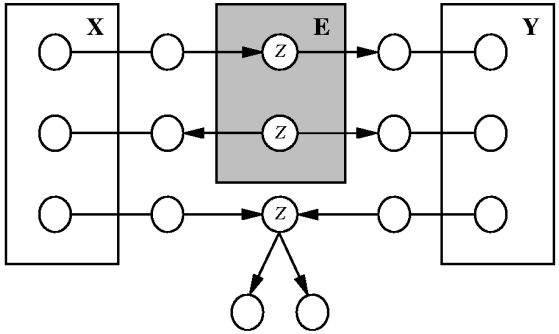
X ind. of Y given Z? No

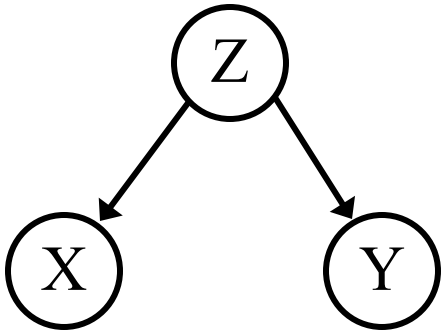


X ind. of Y? No

X ind. of Y given Z? No

Examples (cont.)





X – wet grass

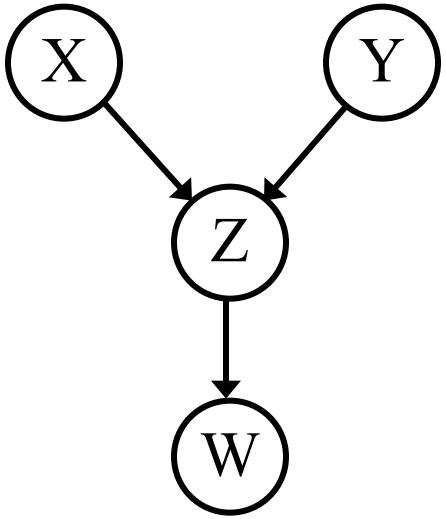
Y – rainbow

Z – rain

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

$P(X \mid Y, Z) = P(X \mid Z)$

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



X – rain

Y – sprinkler

Z – wet grass

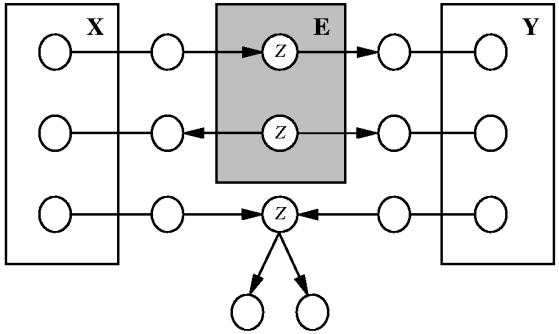
W – worms

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

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

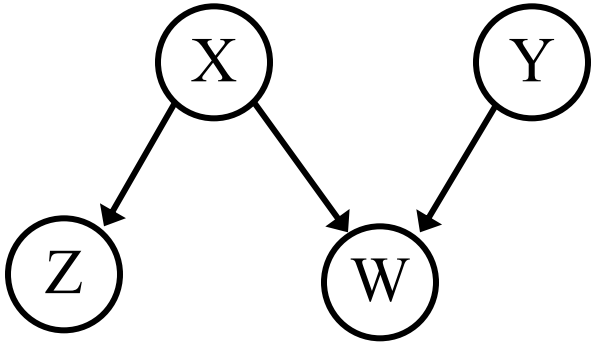
$P(X \mid Y, W) \neq P(X \mid W)$

Examples



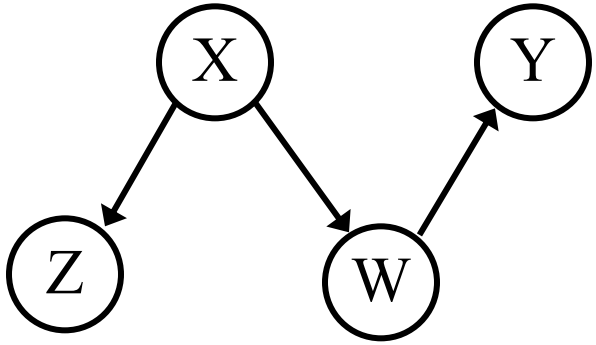
Are X and Y independent?

Are X and Y conditionally independent given Z?



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

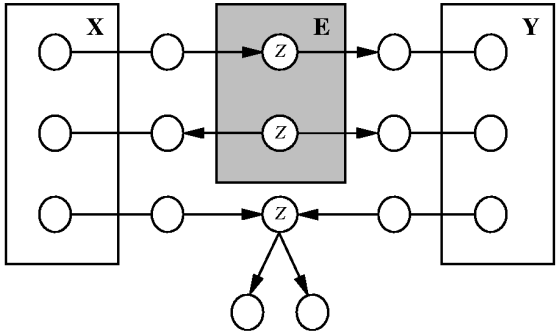
$P(X,Y) = P(X)\,P(Y)$ Yes
$P(X \mid Y, Z) = P(X \mid Z)$ Yes



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

$P(X,Y) \neq P(X)\,P(Y)$ No
$P(X \mid Y, Z) \neq P(X \mid Z)$ No

Conditional Independence



- Where are conditional independences here?

Radio and Ignition, given Battery?

Yes

Radio and Starts, given Ignition?

Yes

Gas and Radio, given Battery?

Yes

Gas and Radio, given Starts?

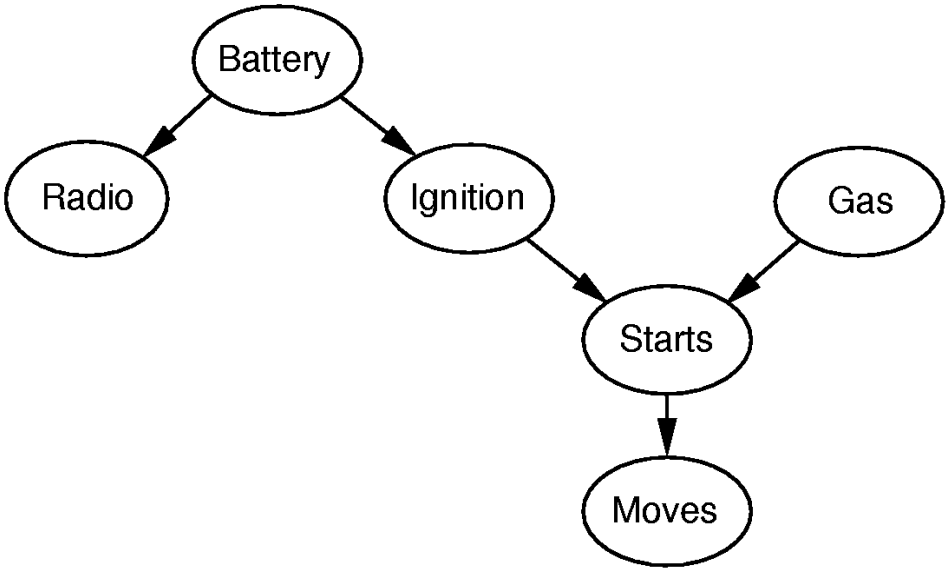
No

Gas and Radio, given nil?

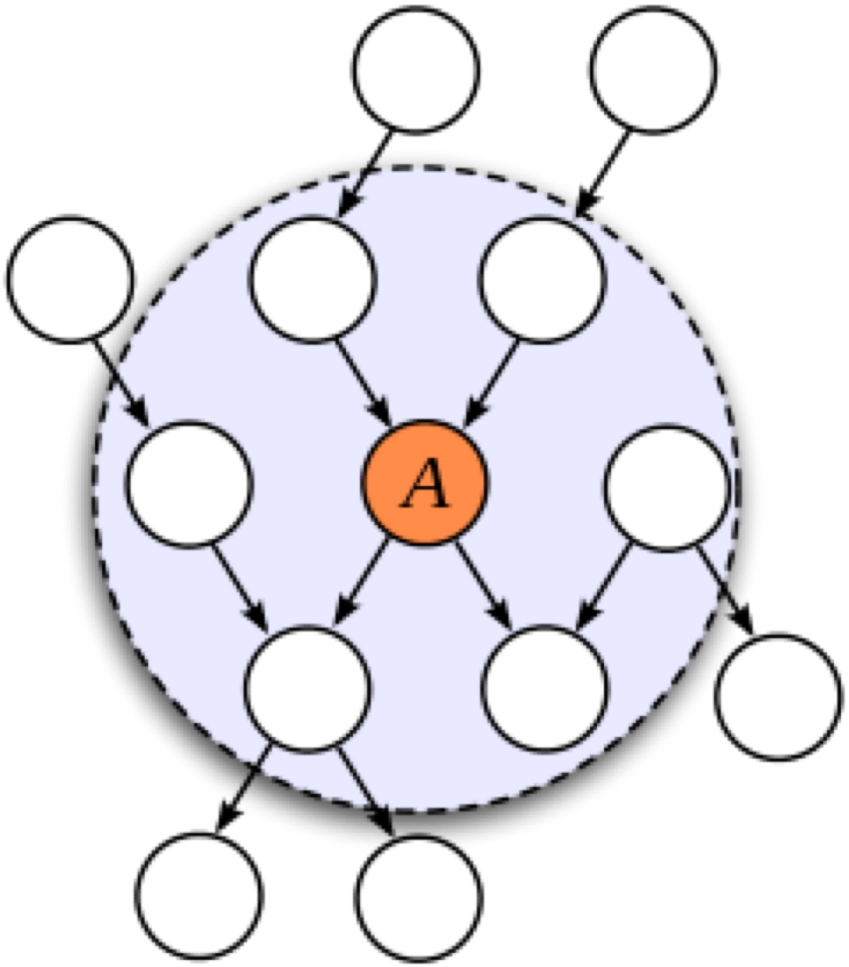
Yes

Gas and Battery, given Moves?

No (why?)



An alternative view: Markov Blankets



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

Why is this important?

- Helps the developer (or the user) verify the graph structure
  - Are these things really independent?
  - Do I need more/fewer arcs?

- Statistical tests for (Conditional) Independence
  - Hilbert-Schmidt Independence Criterion

- Gives hints about computational efficiencies

- Shows that you understand BNs…

Inference in belief nets

- We've seen how to compute any probability from the belief net
  - This is ***probabilistic inference***
    - P(Query | Evidence)
  - Since we know the joint probability, we can calculate anything via marginalization
    - P({red} | {green})

- However, things are usually not as simple as this
  - Structure is large or very complicated
  - Not all CPTs are known
  - **Calculation by marginalization is often infeasible**
  - Bayesian inference is NP hard!!

Inference in belief nets (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 probabilistic inference
  - Remember: AI is "solving exponential problems in polynomial time"
- Markov Chain Monte Carlo, Message Passing / Loopy Belief Propogation
  - Active area of research!
- We won't cover these probabilistic inference algorithms though….

Practical uses of Belief Nets

- Uses of belief nets:
  - Calculating probabilities of causes, given symptoms
  - Calculating probabilities of symptoms, given multiple causes
  - Calculating probabilities to be combined with an agent's utility function
  - Explaining the results of probabilistic inference to the user
  - Deciding which additional evidence variables should be observed in order to gain useful information
    - Is it worth the extra cost to observe additional evidence?
      - $P(X \mid Y)$ vs. $P(X \mid Y, Z)$
  - Performing sensitivity analysis to understand which aspects of the model affect the queries the most
    - How accurate must various evidence variables be? How will input errors affect the reasoning?

Belief Networks

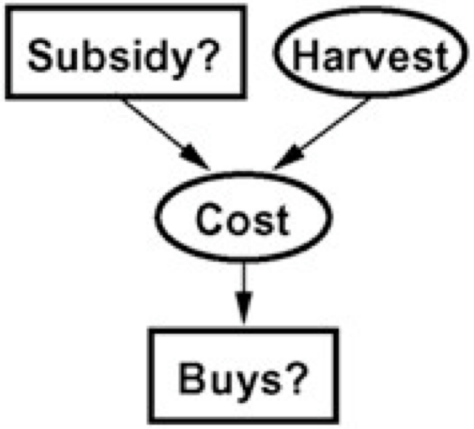
- There are lots of success stories for Bayesian networks
  - Typically for diagnostic inference
    - Business, science, medicine, HCI, databases…
  - Some systems outperform the experts

- Unlike logic-based systems, getting it <u>exactly</u> right isn't always critical

- Given good tools, domain experts (not just computer scientists!) can create good links and fill in reasonable probabilities
  - There are several BN software systems available

One more thing: Continuous Variables?

- CPT will no longer be tables, but functions.



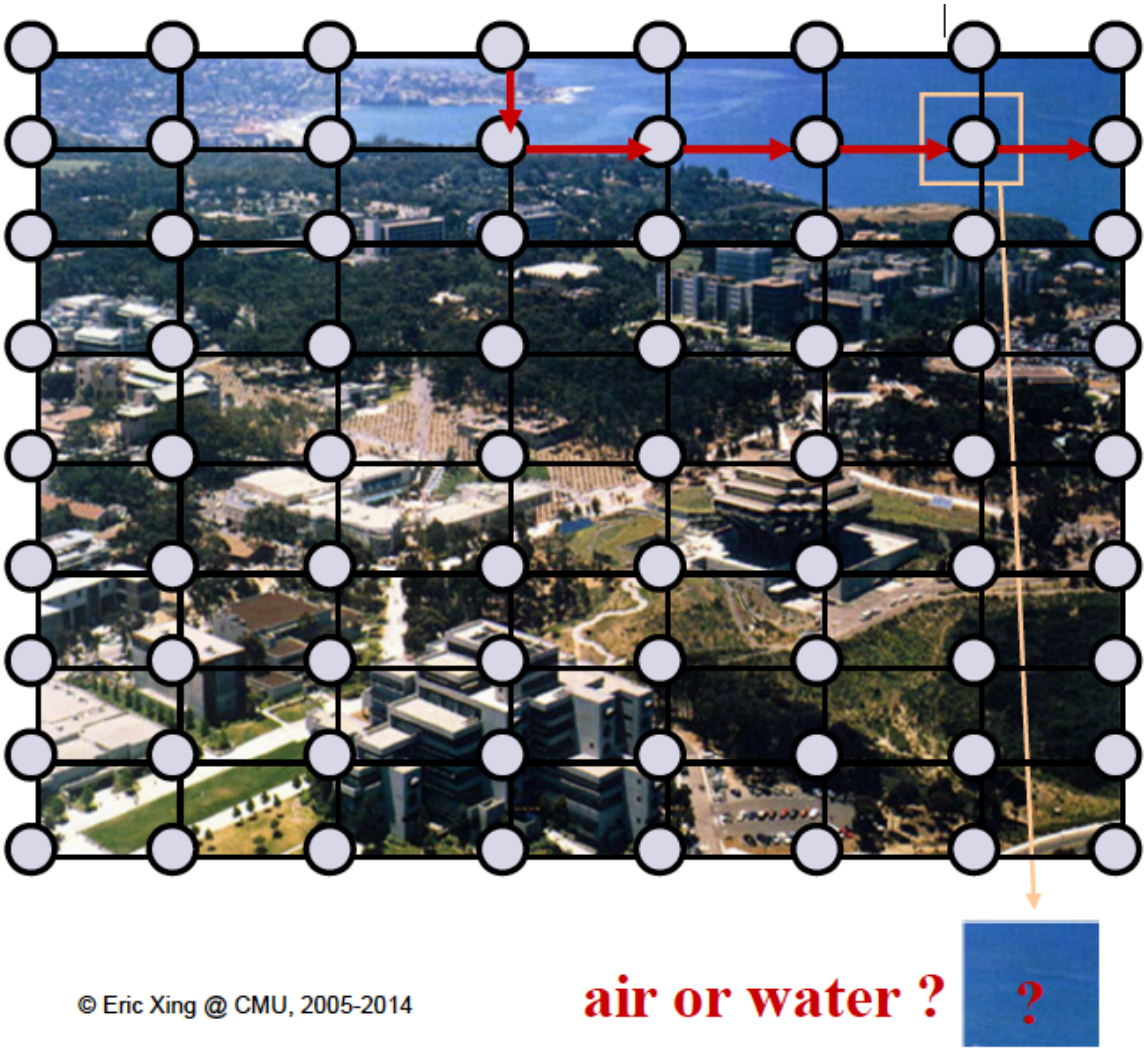
- Discretize? Very large CPT..
- Usually, we parametrize the conditional distribution.
  - e.g., P(Cost | Harvest) = Poisson($< \theta$, Harvest$>$)

Undirected Graphical models **(a.k.a. Markov Random Field)**

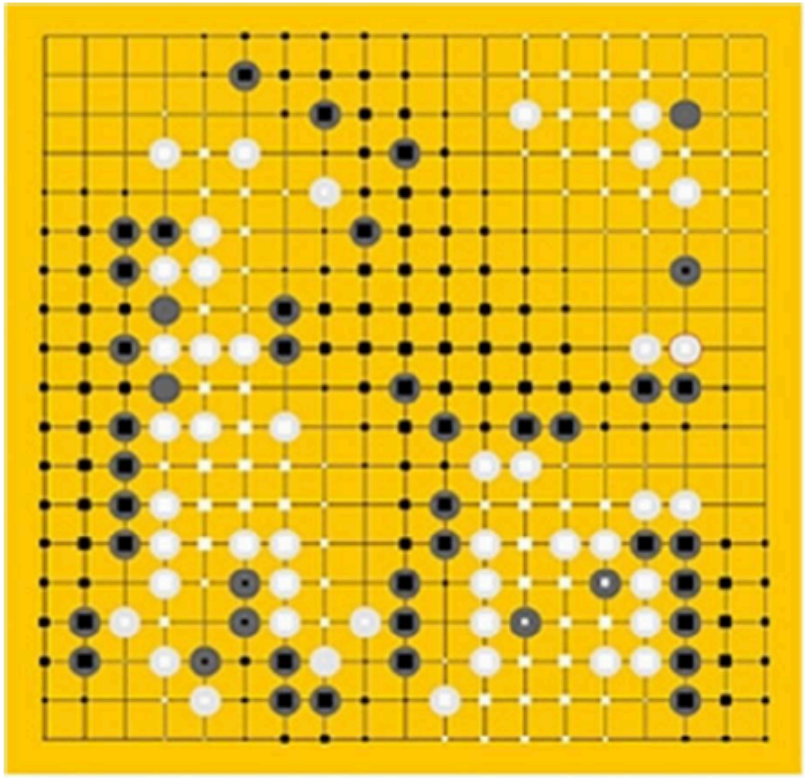
- Directed graph
  - Twitters / Instagrams / Emails
  - NOT necessarily DAG!

- Undirected graph: Facebook

An example for image segmentation.

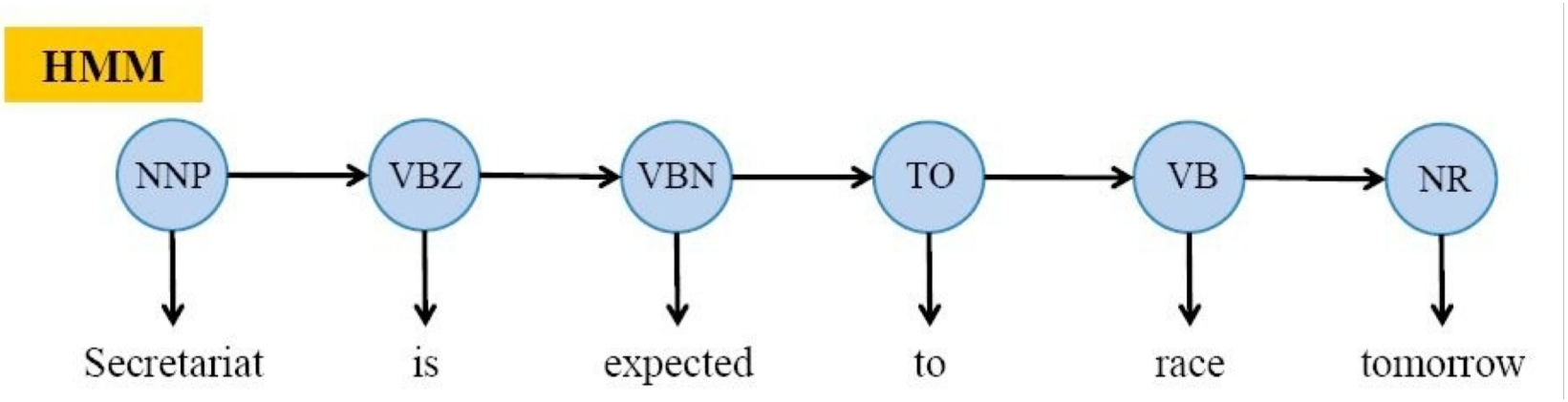


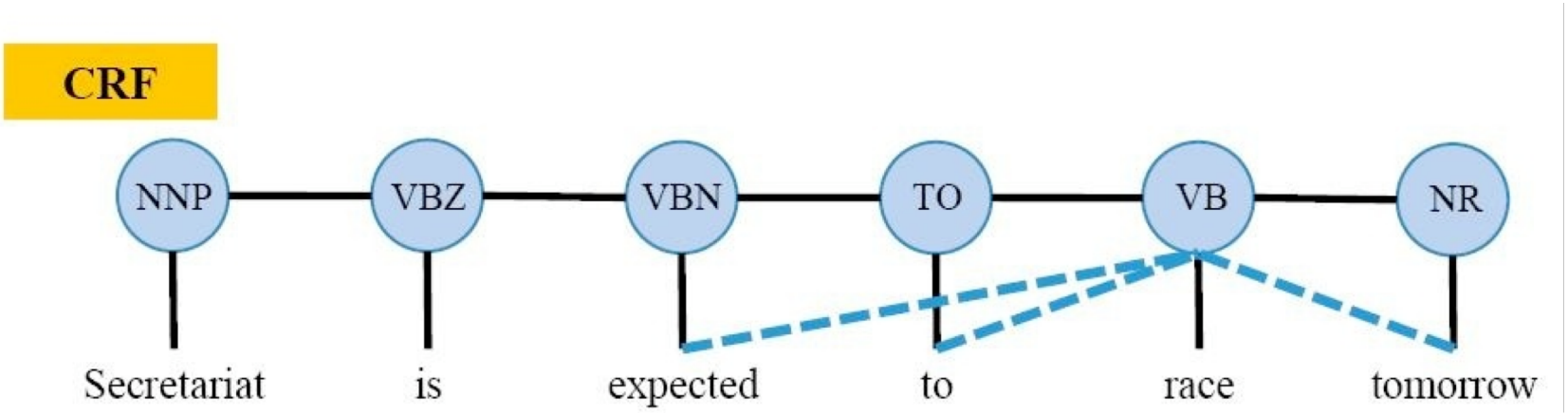
An example for modelling Go!



This is the middle position of a Go game. Overlaid is the estimate for the probability of becoming black or white for every intersection. Large squares mean the probability is higher.

Sometimes using BayesNet or MRF is a matter of choice.





(From quora answer by Pravankumar Reddy)

What did we learn about BayesNet?

- Factorize the joint distributions into the product of CPTs!
- Computational barriers and how to overcome it
  - By assuming conditional independences
- How to read off conditional independences from a DAG

For Undirected Graphical Models

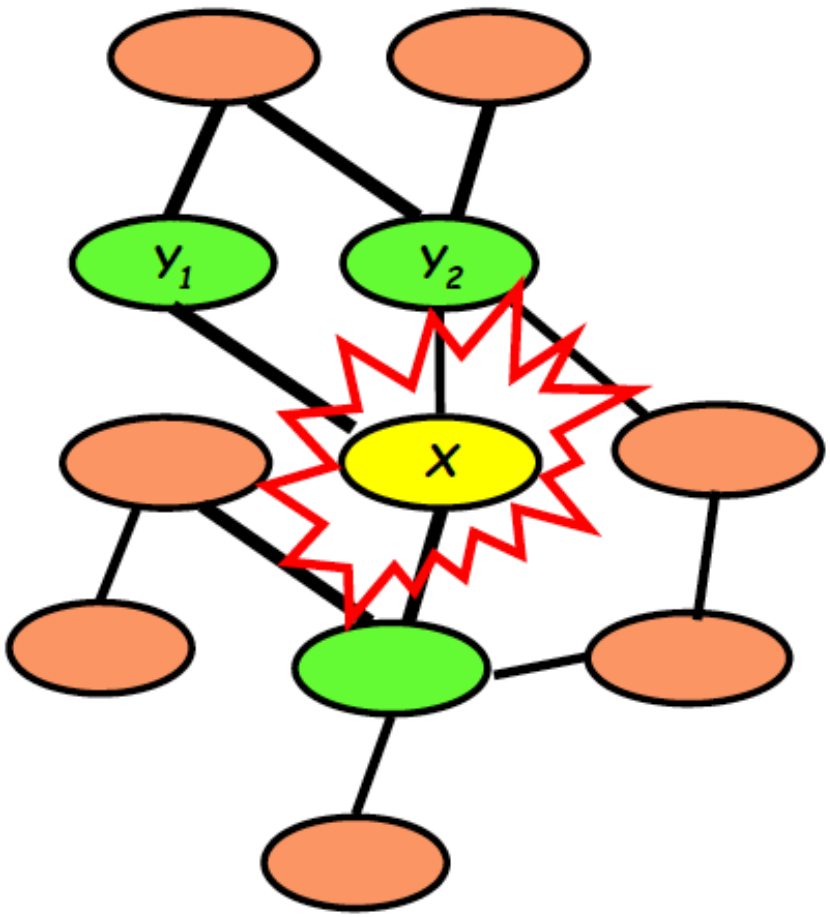
- Factorize the joint distributions into the product of ~~CPTs~~ **Potential Functions (not probabilities!)**!

$$P(x_1,\ldots,x_n) = \frac{1}{Z}\prod_{c\in C}\psi_c(\mathbf{x}_c) \quad \text{where} \quad Z = \sum_{x_1,\ldots,x_n}\prod_{c\in C}\psi_c(\mathbf{x}_c)$$

Partition function

- Computational barriers and how to overcome it
  - By assuming **conditional independences**
- How to read off conditional independences from a ~~DAG~~ **Undirected Graph**

Markov Blankets in MRF



(Illustration from Eric Xing)

Thank You

(More examples in this week's Discussion Class!)