Introduction to Artificial Intelligence

Unit # 11

Course Outline

- Overview of Artificial Intelligence ✓
- State Space Representation ✓
- Search Techniques ✓
- Machine Learning ✓
- Logic ✓
- Probabilistic Reasoning/Bayesian Networks
  - Knowledge Elicitation
  - Inference in BNs
- Evolutionary Computation
- Miscellaneous Topics (depending upon the availability of time)
  - Computer Vision
  - Introduction to Robotics
  - Reinforcement Learning

Acknowledgement

- The slides of this lecture have been taken from the lecture slides of CS307 – “Introduction to Artificial Intelligence by Dr. Sajjad Haider.

Recap

- Joint Distribution: A joint probability gives us a full picture about how random variables are related.
- Marginal Distribution: Marginalization lets us to focus one aspect of the picture.
- Prior probability: Belief about a hypothesis h before obtaining observations.
  - Example: Suppose 10% of people suffer from Hepatitis B. A doctor’s prior probability about a new patient suffering from Hepatitis B is 0.1.
- Posterior probability: Belief about a hypothesis after obtaining observations.

Bayes Theorem

- $P(A \mid B) = \dfrac{P(B \mid A)\, P(A)}{P(B)}$

  $= \dfrac{P(B \mid A)\, P(A)}{P(B|A)P(A)+P(B|\neg A)P(\neg A)}$
- P(A) is the prior probability and P(A | B) is the posterior probability.
- Suppose events A_1, A_2, ……., A_k are mutually exclusive and exhaustive; i.e., exactly one of the events must occur. Then for any event B:

  $P(A_i \mid B) = \dfrac{P(B \mid A_i)\, P(A_i)}{\sum P(B \mid A_i)\, P(A_i)}$

Probabilistic Reasoning

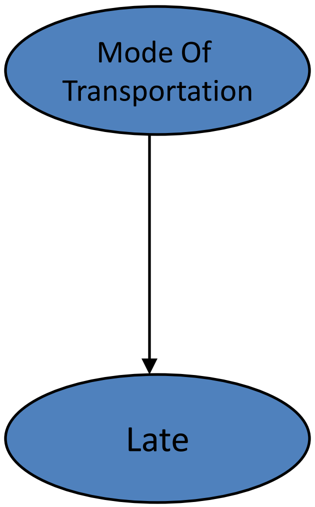
- Logic is used to represent and reason knowledge about the world with facts and rules.
  - bird(tweety).
  - fly(X) :- bird(X).
- In many cases, certain rules are hard to find.
- Probability theory is the most popular way of representing uncertainty.
  - lung_cancer(X) :- smoking(X). is replaced by
  - P(Lung Cancer | Smoking) = 0.6

Bayes Theorem Example

- Suppose that Bob can decide to go to work by one of three modes of transportation, car, bus, or commuter train. Because of high traffic, if he decides to go by car, there is a 50% chance he will be late. If he goes by bus, which has special reserved lanes but is sometimes overcrowded, the probability of being late is only 20%. The commuter train is almost never late, with a probability of only 1%, but is more expensive than the bus.
- Suppose that Bob is late one day, and his boss wishes to estimate the probability that he drove to work that day by car. Since he does not know which mode of transportation Bob usually uses, he gives a prior probability of 1/3 to each of the three possibilities. What is the boss' estimate of the probability that Bob drove to work?

Bayes Theorem Example

- Suppose that a coworker of Bob's knows that he almost always takes the commuter train to work, never takes the bus, but sometimes, 10% of the time, takes the car. What is the coworker's probability that Bob drove to work that day, given that he was late?



Earthquake Example (Pearl)

- You have a new burglar alarm installed.
- It is reliable about detecting burglary, but responds to minor earthquakes.
- Two neighbors (John, Mary) promises to call you at work when they hear the alarm.
  - John always call when hears alarm, but confuses alarm with phone ringing (and calls then also)
  - Mary likes loud music and sometimes misses alarm
- Given evidence about who has and hasn't called, estimate the probability of a burglary.

Cause and Effect Relationships

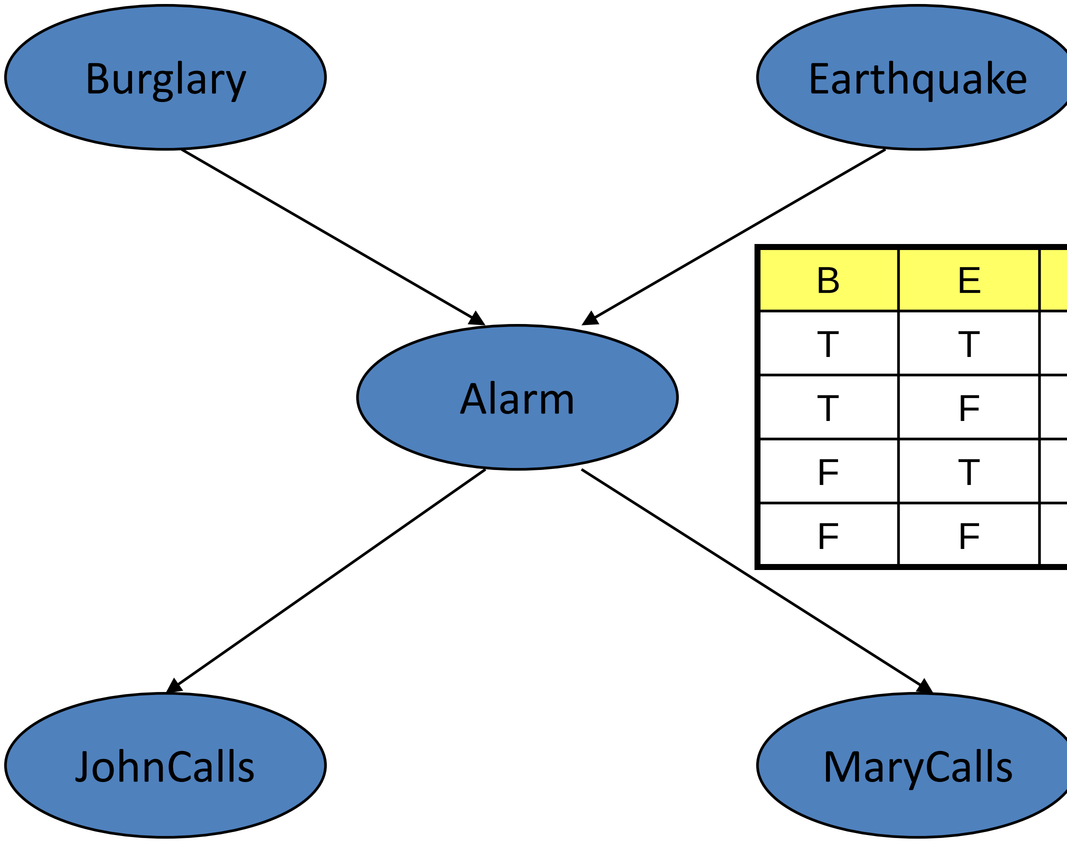
- Burglary → Alarm Goes Off
- Earthquake → Alarm Goes Off
- Alarm Goes Off → John Calls
- Alarm Goes Off → Mary Calls

- Number of Variables
  - 5
- Number of Links
  - 4

Earthquake Example (Pearl)

P(B)
0.1

P(E)
0.02



B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

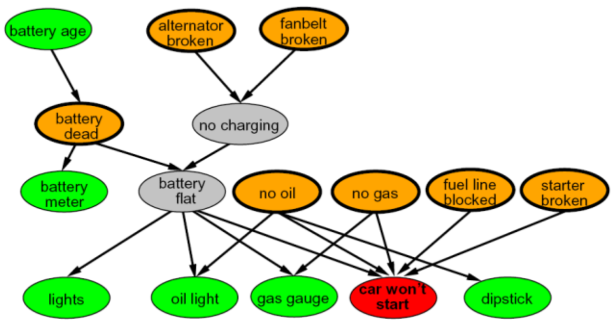
A	P(J|A)
T	0.9
F	0.05

A	P(M|A)
T	0.7
F	0.01

Graphical Models

- Models are descriptions of how (a portion of) the world works

- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables



- What do we do with probabilistic models?
  - We need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)

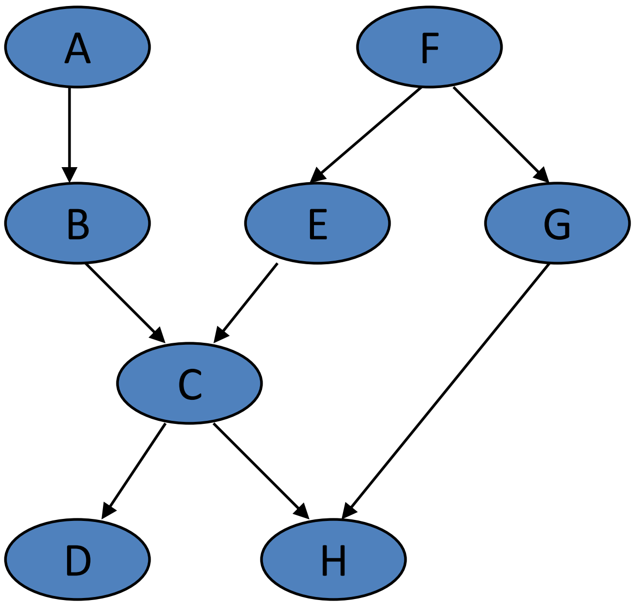
Bayesian Networks

- Two problems with using full joint distributions for probabilistic models:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to estimate anything empirically about more than a few variables at a time

- Bayesian Networks are a technique for describing complex joint distributions using a bunch of simple, local distributions
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions

Bayesian Networks

- A problem domain is modeled by a list of variables $X_1, X_2, \ldots, X_n$.
- Knowledge about the problem domain is represented by a joint probability $P(X_1, X_2, \ldots, X_n)$.
- General probability distribution of 8 variables with 2 states each has $2^8 = 256$ possible values and $2^8 - 1$ probabilities need to be specified.
  - Exponential model size
  - Knowledge acquisition difficult
  - Exponential storange and inference
- A Bayes net is a graphical representation of the joint distribution.
- Assumes that each node is conditionally independent of all its non-descendants given its parents.
- Product of all conditional probabilities is the joint probability of all variables.
  - $P(X_1, X_2, \ldots, X_n = \prod P(X_i \mid parents(X_i))$



18 probabilities are required to specify the joint distribution

Bayesian Networks

- A BN is a Directed Acyclic Graph (DAG) in which:
  - A set of random variables makes up the nodes in the network.
  - A set of directed links or arrows connects pairs of nodes.
  - Each node has a conditional probability table that quantifies the effects the parents have on the node.
- The intuitive meaning of an arrow from a parent to a child is that the parent directly influences the child.
- The direction of this influence is often taken to represent casual influence.
- These influences are quantified by conditional probabilities.

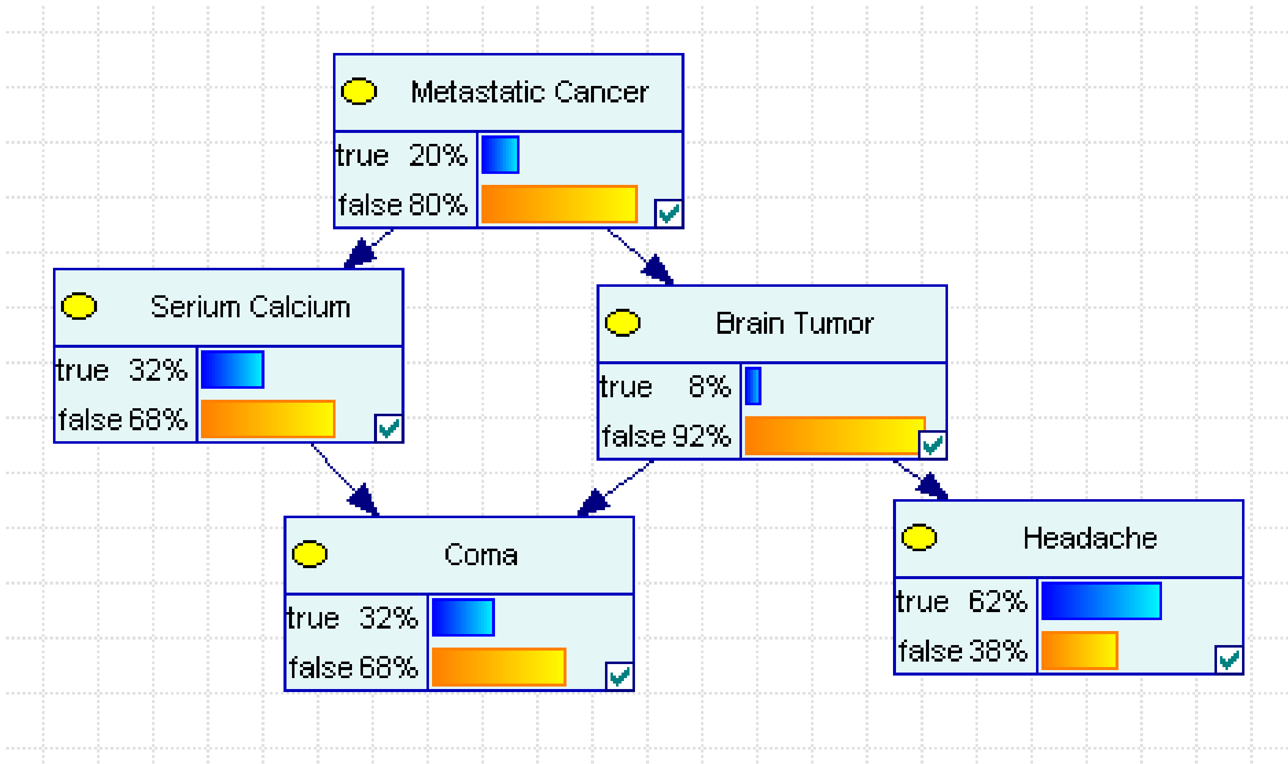
Inference Problem

- The canonical inference problem: find the posterior probability distribution for some variable(s) given direct or virtual evidence about other variable(s)

Cancer Example (Pearl)

- Metastatic cancer is a possible cause of a brain tumor and is also an explanation for increased total serum calcium. In turn, either of these could explain a patient falling into a coma. Severe headache is also possibly associated with a brain tumor.

Cancer Example (Pearl)

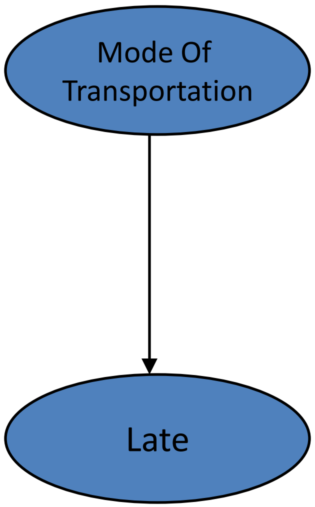


Bayes Theorem Example

- Suppose that Bob can decide to go to work by one of three modes of transportation, car, bus, or commuter train. Because of high traffic, if he decides to go by car, there is a 50% chance he will be late. If he goes by bus, which has special reserved lanes but is sometimes overcrowded, the probability of being late is only 20%. The commuter train is almost never late, with a probability of only 1%, but is more expensive than the bus.
- Suppose that Bob is late one day, and his boss wishes to estimate the probability that he drove to work that day by car. Since he does not know which mode of transportation Bob usually uses, he gives a prior probability of 1/3 to each of the three possibilities. What is the boss' estimate of the probability that Bob drove to work?

Revised “Bob late to work” Example

- Mode of Transportation: {Car, Not Car (Public Transport)}
- Late: {True, False}
- $P(C) = 1/3$, $P(\neg C) = 2/3$
- $P(L \mid C) = 0.4$, $P(L \mid \neg C) = 0.2$



- **$P(L) = ?$**
- **$P(C \mid L)$, $P(\neg C \mid L)$**

Revised "Bob late to work" Example

- One approach is to compute the joint distribution through
  - $P(X_1, X_2, \ldots, X_n) = \prod P(X_i \mid parents(X_i))$
  - $P(L, C) = P(L \mid C)\, P(C) = 0.4 \times 0.33$
  - $P(L, \neg C) = P(L \mid \neg C)\, P(\neg C) = 0.2 \times 0.67$
  - $P(\neg L, C) = P(\neg L \mid C)\, P(C) = 0.6 \times 0.33$
  - $P(\neg L, \neg C) = P(\neg L \mid \neg C)\, P(\neg C) = 0.8 \times 0.67$
- Now compute P(L), P(C | L)
  - $P(L) = P(L, C) + P(L, \neg C)$
  - $P(C \mid L) = \dfrac{P(C, L)}{P(L)}$

Revised "Bob late to work" Example

Or we could use Bayes Theorm based propagation.

$$P(C \mid L) = \frac{P(L \mid C)\, P(C)}{P(L)} = \frac{0.4 \times 0.33}{P(L)}$$

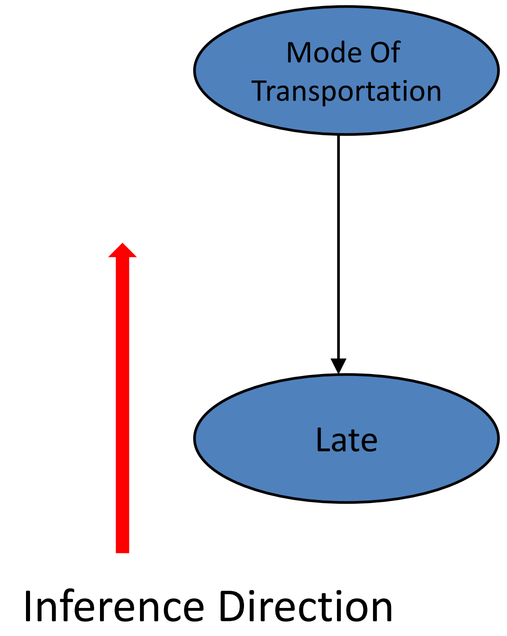
where

$$P(L) = P(L \wedge C) + P(L \wedge \neg C)$$

$$= P(L \mid C)P(C) + P(L \mid \neg C)\, P(\neg C)$$

$$= 0.4 \times 0.33 + 0.2 \times 0.67 = ?$$

Both approaches (on the previous and the current slides) should give the same result

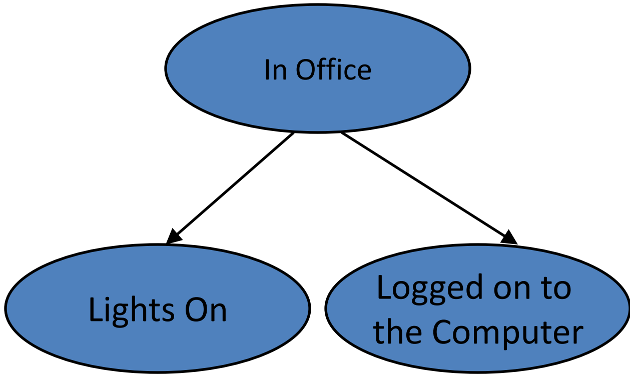


Lecturer’s Life Example (Korb & Nicholson)

- Dr. Ann Nicholson spends 60% of her work time in her office. The rest of her work time is spent elsewhere. When Ann is in her office, half the time her light is off (when she is trying to hide from students and get some real work done).
- When she is not in her office, she leaves her light on only 5% of the time.
- 80% of the time she is in her office, Ann is logged onto the computer.
- Because she sometimes logs onto the computer from home, 10% fo the time she is not in her office, she is still logged onto the computer.
- Suppose a student checks Dr. Nicholson’s login status and sees that she is logged on. What effect does this have on the student’s belief that Dr. Nicholson’s light is on?

Lecturer's Life Example (Cont'd)

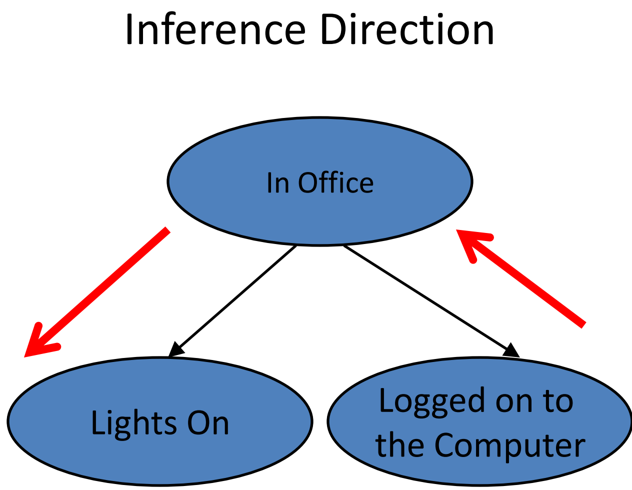
- In Office (O): {T, F}
- Lights On (L): {T, F}
- Logged on to the Computer (C) : {T, F}
- $P(O) = 0.6$
- $P(L \mid O) = 0.5$, $P(L \mid \neg O) = 0.05$
- $P(C \mid O) = 0.8$, $P(C \mid \neg O) = 0.1$



- **$P(L \mid C) = ?$**

Lecturer's Life Example (Cont'd)

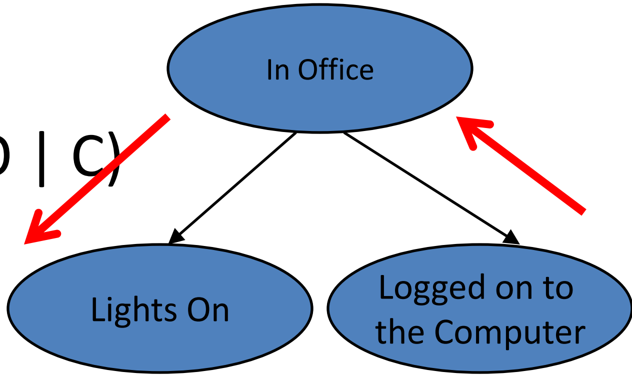
- One possibility is to compute the joint distribution and then compute the conditional probability
  - $P(X_1, X_2, \ldots, X_n = \prod P(X_i \mid parents(X_i))$
  - $P(O, L, C) = P(L \mid O)\, P(C \mid O)\, P(O)$
  - $P(O, L, \neg C) = P(L \mid O)\, P(\neg C \mid O)\, P(O)$
  - $P(O, \neg L, C) = P(\neg L \mid O)\, P(C \mid O)\, P(O)$
  - $P(O, \neg L, \neg C) = P(\neg L \mid O)\, P(\neg C \mid O)\, P(O)$
  - $P(\neg O, L, C) = P(L \mid \neg O)\, P(C \mid \neg O)\, P(\neg O)$
  - $P(\neg O, L, \neg C) = P(L \mid \neg O)\, P(\neg C \mid \neg O)\, P(\neg O)$
  - $P(\neg O, \neg L, C) = P(\neg L \mid O)\, P(C \mid \neg O)\, P(\neg O)$
  - $P(\neg O, \neg L, \neg C) = P(\neg L \mid \neg O)\, P(\neg C \mid \neg O)\, P(O)$
- After computing the joint distribution, we can compute $P(L \mid C)$



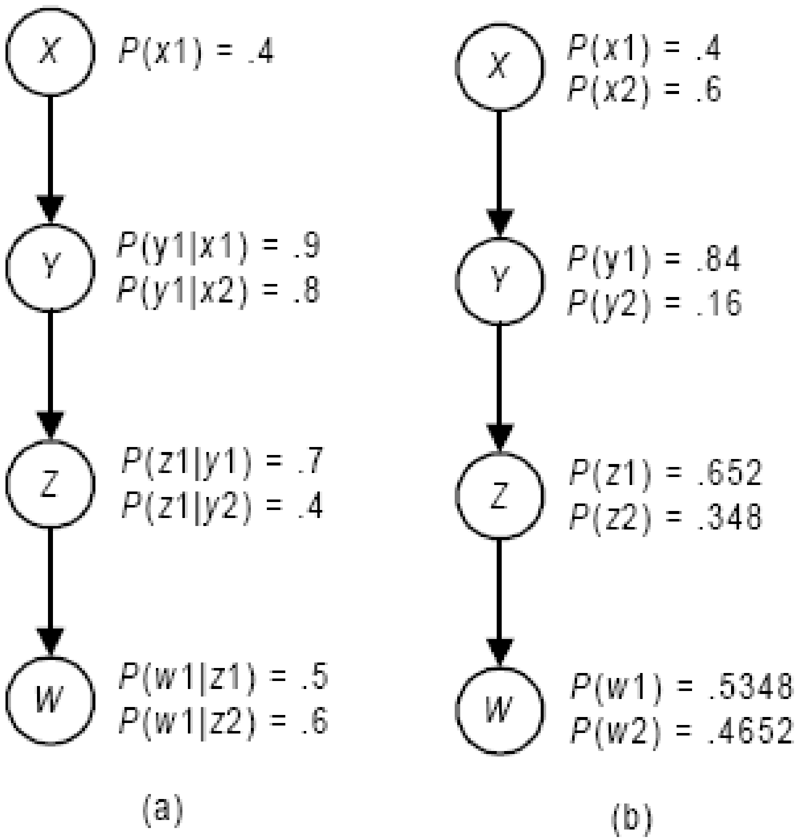
Lecturer’s Life Example (Cont’d)

- The other approach is to do inference using probability propagation based on Bayes Theorem
- First compute the posterior probability P(O | C) {referred to as P(O*)}
  - $P(O^*) = \dfrac{P(C \mid O)\, P(O)}{P(C)}$
  - And then P(L | C) {referred to as P(L*)}
  - $P(L^*) = P(L \mid O)P(O^*) + P(L \mid \neg O)P(\neg O^*)$

Inference Direction



Inference in Singly Connected Networks



$P(y1) = P(y1 \mid x1)\, P(x1) + P(y1 \mid x2)\, P(x2)$

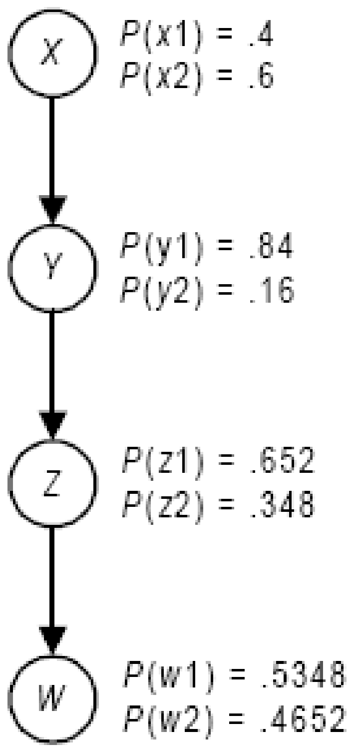
$P(y2) = 1 - P(y1)$

$P(z1) = P(z1 \mid y1)\, P(y1) + P(z1 \mid y2)\, P(y2)$

$P(z2) = 1 - P(z1)$

$P(w1) = ???$

Inference in Singly Connected Networks



Suppose we get evidence that w1 is true, i.e., P(w1) = 1.

Now compute the posterior probabilities:
P*(z1), P*(y1), P*(x1)

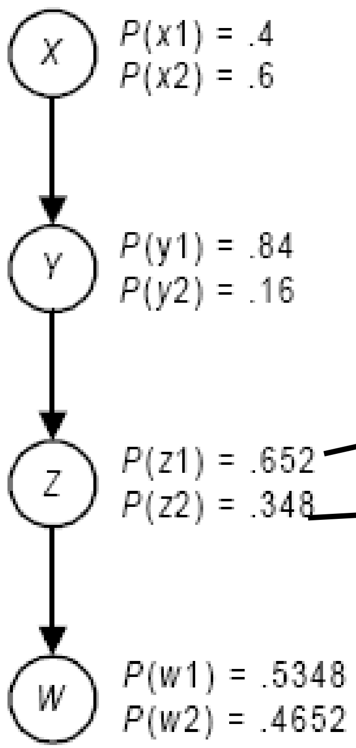
P*(z1) = P(z1 | w1) P(w1) ← P(w1)=1
+ P(z1 | w2) P(w2) ← P(w2)=0

Computing P(z1 | w1) using Bayes theorem:
P(z1 | w1) = P(w1 | z1) P(z1) / P(w1)
P(z1 | w1) = 0.5 x 0.652 / 0.5348 = 0.61

=>
P*(z1) = 0.61 * 1 + 0 = 0.61

Computation of P*(y1) on the next slide

Inference in Singly Connected Networks



Now update the probability of Y.
P*(y1) = P(y1 | z1) P(z1)
+ P(y1 | z2) P(z2)

P(z1)=0.61
P(z2)=0.39
} Computed on the previous slide

Computing P(y1 | z1) and P(y1 | z2) using Bayes theorem:
P(y1 | z1) = P(z1 | y1) P(y1) / P(z1)
P(y1 | z1) = 0.7 x 0.84 / 0.652 = 0.90

P(y1 | z2) = P(z2 | y1) P(y1) / P(z2)
P(y1 | z2) = 0.3 x 0.84 / 0.348 = 0.92

Finally plug the values in the equation at the top
P*(y1) = P(y1 | z1) P(z1) + P(y1 | z2) P(z2)
= 0.90 x 0.61 + 0.92 X 0.39
= 0.91
Confirm the results using Genie.

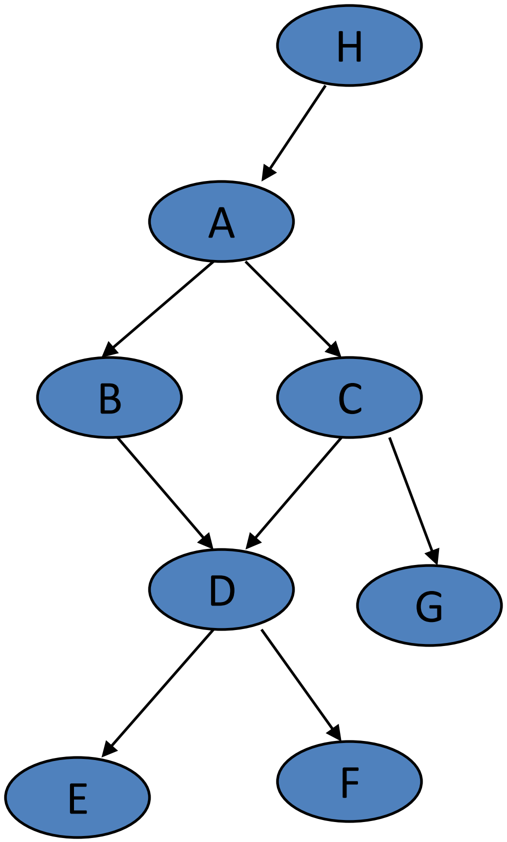
Finally compute P*(x1)

d-Separation

- d-separation is short for direction-dependent separation
- d-separation of vertices in a graph corresponds to conditional independence of the associated random variables
- d-separation is the mathematical basis for efficient inference algorithms in Bayesian networks
- Two nodes X and Y are d-separated by a set Z if all paths between X and Y are blocked by Z.
- When X and Y are d-separated by Z, no information can be transmitted between X and Y given Z. Hence X and Y are conditional independent given Z.

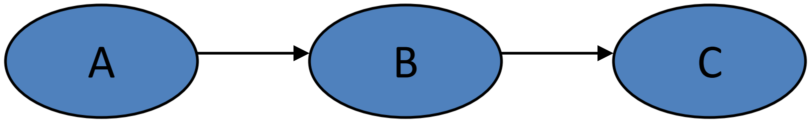
Markov Blanket (Laskey)

- **A node's Markov blanket consists of its parents, children, and other parents of its children (co-parents)**
  - **The Markov blanket of node B consists of all nodes whose local probability table mentions B and all nodes their local probability tables mention**
- **A node's Markov blanket d-separates it from all other nodes in the graph**
  - **A node is conditionally independent of all other nodes given its Markov blanket**

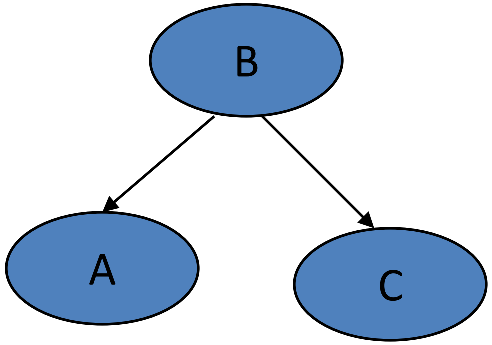


Conditional Dependence/Independence

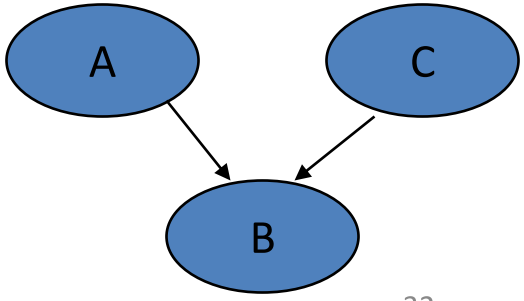
- **Causal Chains**
  - smoking causes cancer which causes dyspnoea
  - if we have evidence on B then A and C become independent



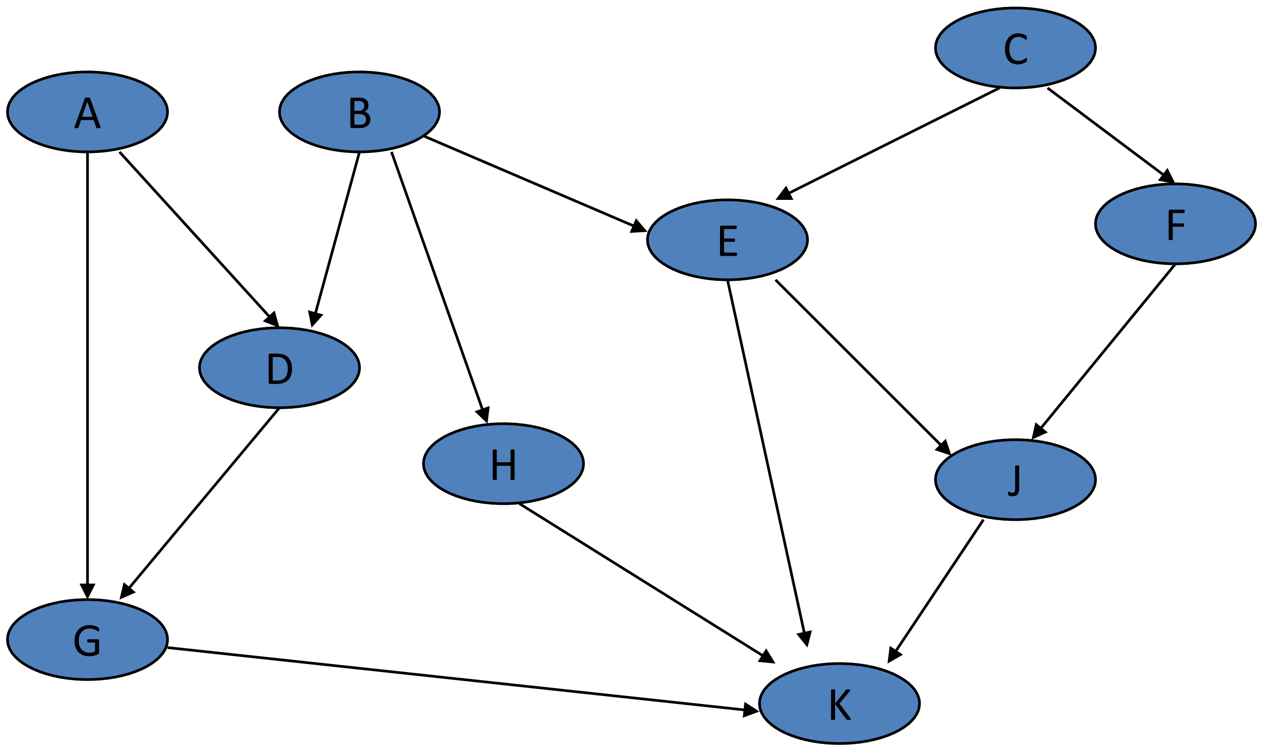
- **Common Causes**
  - cancer is a common cause of the two symptoms, a positive XRay result and dyspnoea
  - If we have evidence on B then A and C become independent



- **Common Effects**
  - Cancer is a common effect of pollution and smoking
  - If we have evidence on B then A and C become dependent

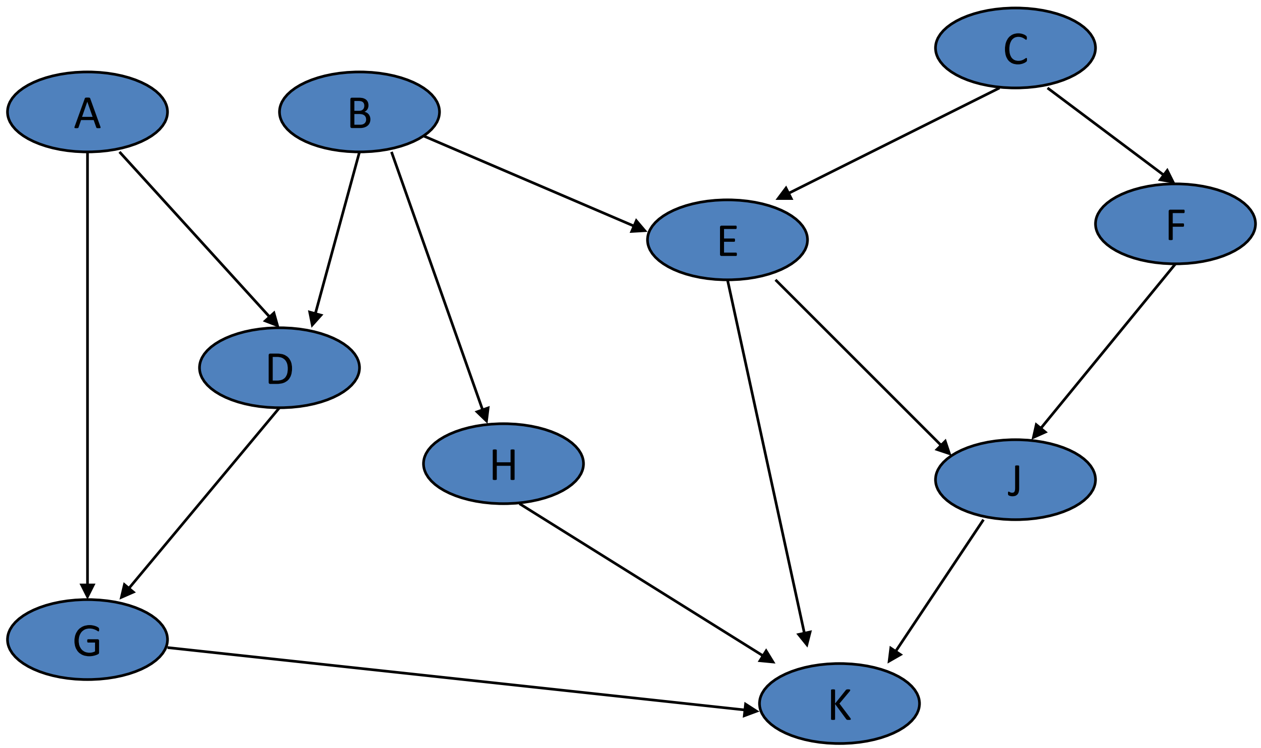


Inference Mechanism



- A and K are independent?
- What's the Markov blanket of H?
- A and B are independent given D?
- A and K are independent given G?
- E and J are independent given K?
- G and B are independent given D?
- E and H are independent given B?
- E and H are independent given B and K?

Inference Mechanism



- A and K are independent? (F)
- A and K are independent given G? (F)
- What's the Markov blanket of H?
- A and B are independent given D? (F)
- E and J are independent given K? (F)
- G and B are independent given D? (F)
- E and H are independent given B? (T)
- E and H are independent given B and K? (F)