CS 482/682 – AI: Reasoning Under Uncertainty

Fall 2021

Failures of Logic

- First-order logic represents a certainty

  $\forall p\ Symptom(p, Toothache) \Rightarrow Disease(p, Cavity)$

- To make the rule true, we must add an almost unlimited set of causes

  $\forall p\ Symptom(p, Toothache) \Rightarrow Disease(p, Cavity) \vee Disease(p, GumDisease) \vee Disease(p, ImpactedTooth) \vee \ldots$

- Conversion to a causal rule does not help

  $\forall p\ Disease(p, Cavity) \Rightarrow Symptom(p, Toothache)$

Why Does Logic Fail?

- Laziness
  - Too much work to list entire sets of consequents or antecedents
  - List all possible causes for a toothache
- Theoretical Ignorance
  - No complete theory for the domain exists
  - Describe the conditions that cause AIDS
- Practical Ignorance
  - Even if we know all the rules, we may be uncertain about a particular event
  - What was the white blood cell count of the patient two years ago?

Probability provides a way of summarizing the uncertainty that comes from our laziness and ignorance

Basic Probability

- I'm assuming a basic understanding of probability theory, but as a quick review:
- Unconditional (prior) probability:
  - P (Cavity) = 0.1
- Random Variables
  - P (Weather = snow) = 0.05
- Probability distribution

- Conditional (posterior) probability:
  - P (Cavity|Toothache) = 0.8

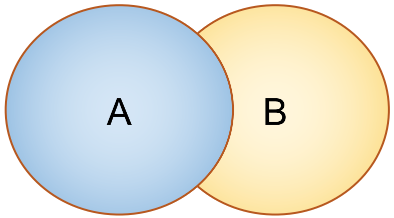
Where do Probabilities come from?

- Frequentist
  - Probabilities come from experiments
  - “9 out of 10 dentists agree”
- Objectivist
  - Probabilities are real aspects of the universe
  - Propensities of objects to act in certain ways
- Subjectivist
  - Probabilities characterize an agent’s beliefs
  - “In my opinion, there is a 30% chance of success”

Basic Probability II

Axioms:

- $0 \leq P(A) \leq 1$
- $P(True) = 1$
- $P(False) = 0$
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$
- $P(\neg A) = 1 - P(A)$
- $P(A \wedge B) = P(A|B)\,P(B)$ (the product rule)



Bayes' Rule

From the product rule:

$$P(A \wedge B) = P(A|B)\,P(B)$$
$$P(B \wedge A) = P(B|A)\,P(A)$$
$$P(A \wedge B) = P(B \wedge A)$$

Bayes' Rule:

$$P(A|B)\,P(B) = P(B|A)\,P(A)$$
$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$$

Application of Bayes' Rule

Medical example:

- 1 in 20 patients reports a stiff neck
- 1 in 50,000 patients has meningitis
- Meningitis causes a stiff neck 50% of the time
- If I have a stiff neck, what is the chance that I have meningitis?

Apply Bayes' Rule:

$$P(S) = .05$$

$$P(M) = .00002$$

$$P(S \mid M) = 0.5$$

$$P(M \mid S) = \frac{P(S \mid M)\,P(M)}{P(S)}$$

$$P(M \mid S) = \frac{0.5 * .00002}{.05}$$

$$P(M \mid S) = 0.0002$$

Bayes theorem and practice

$$P(B|A) = \frac{P(B) \cdot P(A|B)}{P(A)}$$

Bayes theorem

- In the medical field, it is more intuitive if we use the notation H for hypothesis, and O for observed event.
- Upon the **observance** of O, one goes back and assesses the probability of the causal **hypothesis** H.

$$P(H|O) = \frac{P(H) \cdot P(O|H)}{P(O)}$$

Bayes theorem

- Furthermore, the probability of the observed event P(O) can be written as:

$$P(O) = P\big((O \cap H) \cup (O \cap nonH)\big)$$
$$\mathrm{P}(O) = P(O \cap H) + P(O \cap nonH)$$
$$\mathrm{P(O)} = P(H) \cdot P(O|H) + P(nonH) \cdot P(O|nonH)$$

- The above formula can be extended to "marginalization" as used in the midterm practice (slides number 11 in A[4][1] – Midterm Practice.pdf on Canvas)
    - $P(w,r,h) = \sum_{b'} P(w,r,h,b') = \sum_{b} P(w,r,h|b') * P(b')$

- **O** may happen if the hypothesis **H** is **true** (first term) or if the hypothesis **H** is **not true** (second term)

Bayes theorem

- Using the value of P(O), we get

$$P(H|O) = \frac{P(H) \cdot P(O|H)}{P(H) \cdot P(O|H) + P(nonH) \cdot P(O|nonH)}$$

Bayes theorem

- A certain disease has an incidence of about 20%. This disease is to be screened for using a test procedure that provides positive results for 70% of the subjects which suffer from it and for about 1 in every 100 healthy subjects. What is the probability that someone with a positive test result actually has the disease?

$$P(H|O) = \frac{P(H) \cdot P(O|H)}{P(H) \cdot P(O|H) + P(nonH) \cdot P(O|nonH)}$$

Naïve Bayes Models

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\prod_i \mathbf{P}(Effect_i \mid Cause).$$

- It is called "naïve" because it assumes that the effects are independent given the cause.
- In practice, naïve Bayes systems often work very well, even when the conditional independence assumption is not strictly true

- See Naïve Bayes slides in AI[4][0] – NaiveBayes.pdf !!!

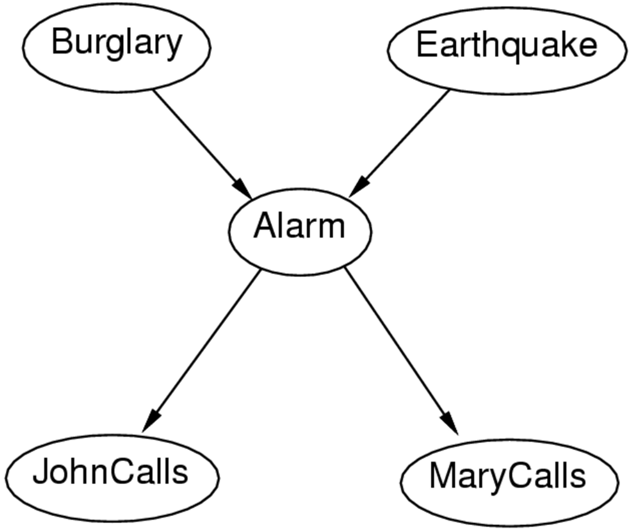
Belief Networks

Probabilistic Reasoning Systems

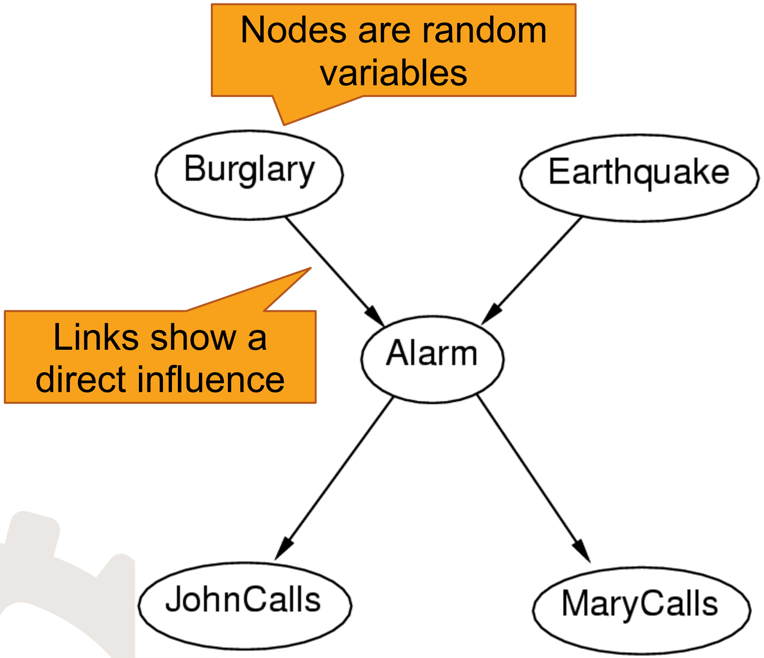
We've seen the syntax and semantics of probability

Now we look at an inference mechanism:

Belief networks



A Basic Belief Network



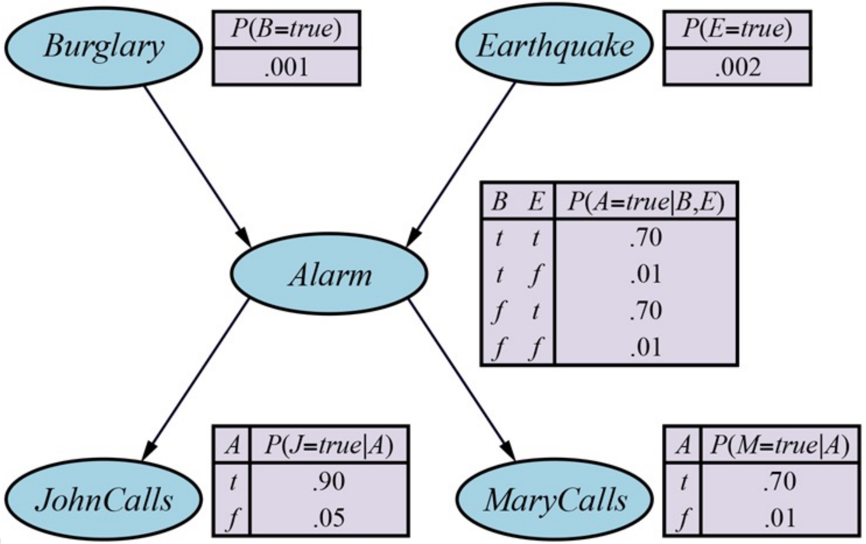
You have a new home alarm that responds

- Accurately to burglaries
- Occasionally responds to earthquakes

When the alarm rings, your neighbors call you at work

- John usually calls, but sometimes confuses the telephone for the alarm
- Mary listens to loud music and sometimes misses the alarm, but almost only calls when the alarm actually rings

A Basic Belief Network



A conditional probability table gives the likelihood of a particular combination of values

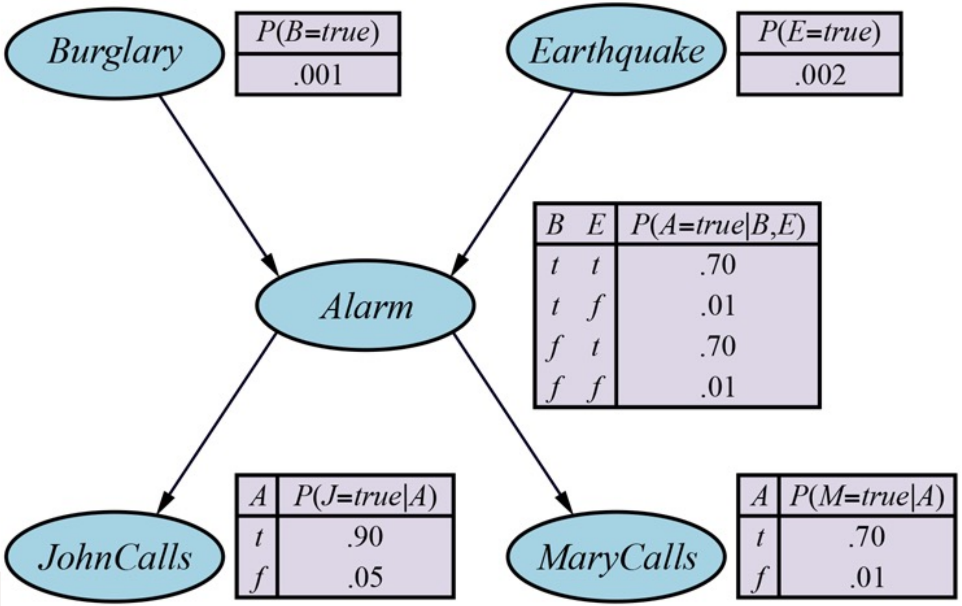
You have a new home alarm that responds

- Accurately to burglaries
- Occasionally responds to earthquakes

When the alarm rings, your neighbors call you at work

- John always calls, but sometimes confuses the telephone for the alarm
- Mary sometimes misses the alarm, but only calls when the alarm actually rings

Conditional Probabilities



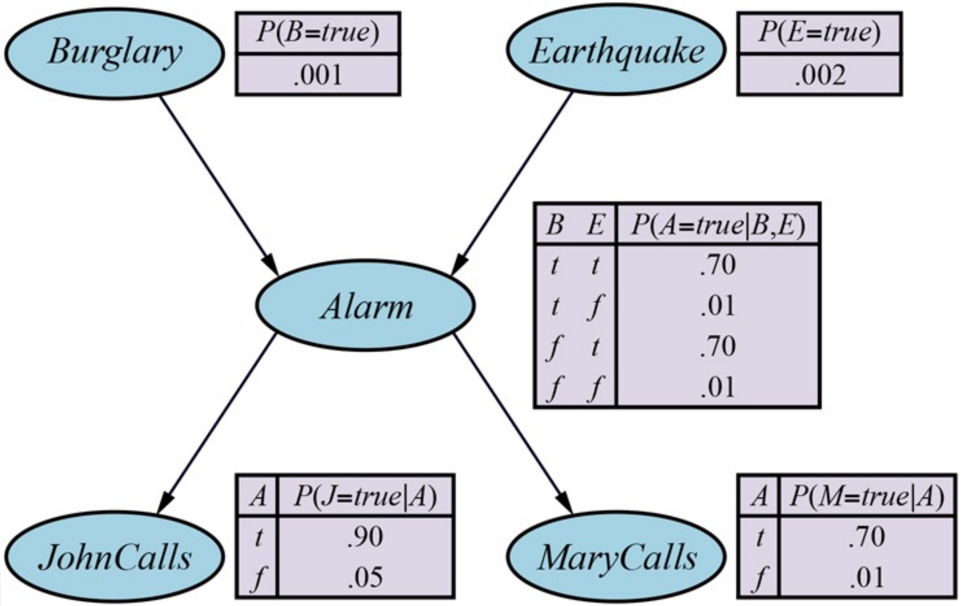
$$P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n)$$

or $P(x_1, \ldots, x_n)$ for short

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \,|\, parents(X_i))$$

$$\begin{aligned} P(x_1, \ldots, x_n) &= P(x_n \,|\, x_{n-1}, \ldots, x_1) P(x_{n-1} \,|\, x_{n-2}, \ldots, x_1) \cdots P(x_2 \,|\, x_1) P(x_1) \\ &= \prod_{i=1}^{n} P(x_i \,|\, x_{i-1}, \ldots, x_1)\,. \end{aligned}$$

Conditional Probabilities



$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \,|\, parents(X_i))$$

What is $P(j, m, a, \neg b, \neg e)$?

The alarm has sounded, but neither a burglary nor an earthquake has occurred, and both John and Mary call.

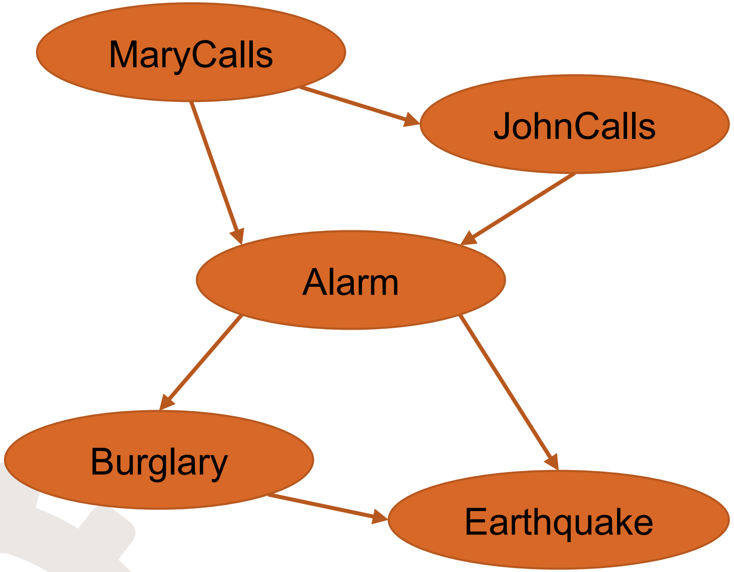
$P(j, m, a, \neg b, \neg e) =$
$P(j|m, a, \neg b, \neg e))P(m, a, \neg b, \neg e) =$
$P(j|a)P(m, a, \neg b, \neg e) =$
$P(j|a)P(m|a, \neg b, \neg e)P(a, \neg b, \neg e) =$
$P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b, \neg e) =$
$P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e) =$
$0.9 \times 0.7 \times 0.01 \times 0.999 \times 0.998 = 0.00628$

Incremental Construction of Belief Nets

Incremental belief net construction:

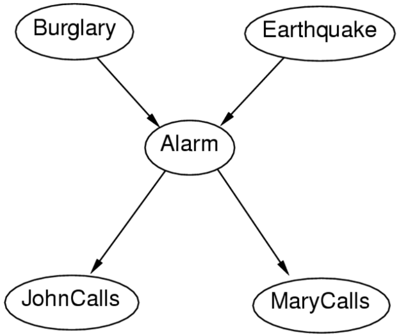
1. Choose the set of relevant variables
2. Choose an ordering for the variables
3. While there are variables left:
    a. Pick a variable *a1* and add a node to the network for it.
    b. Set Parents(a1) to the minimal set of nodes that satisfies the conditional independence property
    c. Define the conditional probability table for node *a1*

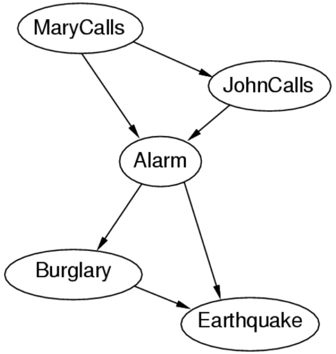
Incremental Construction of Belief Networks

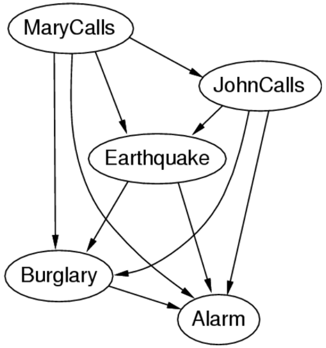


- Add MaryCalls
- Add JohnCalls
    - Dependence with MaryCalls since $P(John|Mary) \neq P(John)$
- Add Alarm
    - More likely if both calls are made
- Add Burglary
    - Phone calls don't tell us anything about the chance of a burglar, but the alarm does
- Add Earthquake
    - Alarm acts as earthquake predictor
    - Presence of a burglar helps determine whether or not an earthquake occurred

Incremental Construction of Belief Networks

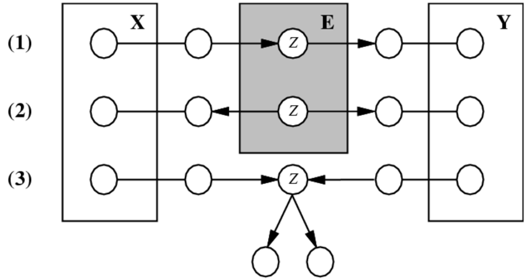


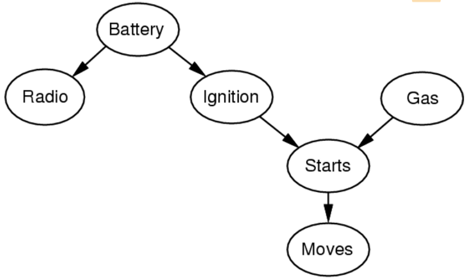




- Order in which you add nodes can make a difference on the number of links
- “Correct order” to add nodes is to add the “root causes” first and then the variables they influence, and so on...
- If we stick to a causal model, we need fewer probabilities and these probabilities will be easier to create

Conditional Independence Relations in Belief Nets





A set of nodes ***E*** d-separates two sets of nodes ***X*** and ***Y*** if every directed path from X to Y is blocked given ***E***

Presence of Gas and a working Radio are:

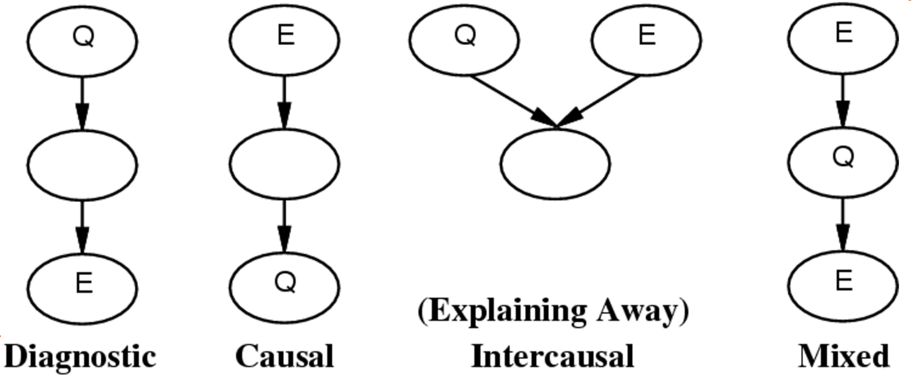
- Independent given evidence about the ignition
- Independent given evidence about the battery
- Independent given no evidence
- Dependent given evidence that the car starts
    - If the car does not start, but the radio plays, then the chance of being out of gas is increased

Types of Inference

Given that a burglary occurred, compute the chance that John calls

Given that John calls and that there was an earthquake, compute the chance of the alarm going off

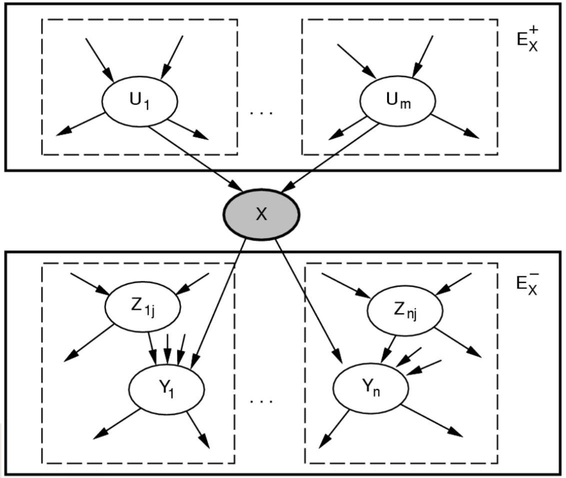
Given that John calls, infer the chance of a burglary



Given that the alarm rang and there was an earthquake, what is the chance that there was a burglary?

E = evidence
Q = query

Inference with Belief Nets

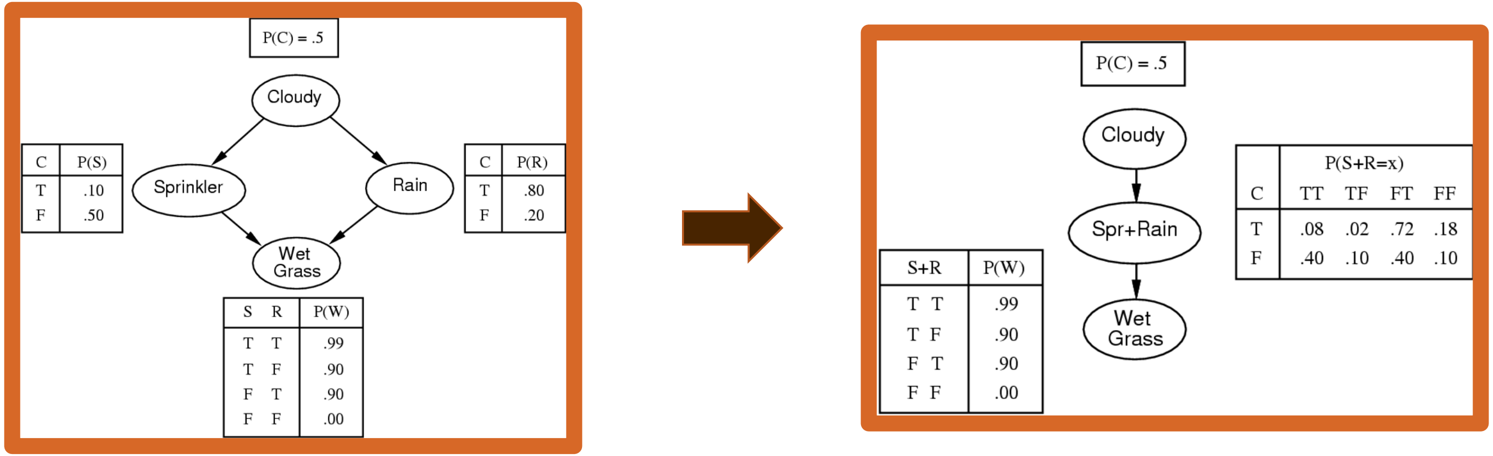


- Singly connected graph
  - only one path between any two nodes
- Multiply connected graph
  - multiple paths between nodes
- Inference in singly-connected graphs
  - Causal support for X:
    - Evidence "above" X (from the direction of X's parents)
  - Evidential support for X:
    - Evidence "below" X (from the direction of X's children)

Three Techniques for Inference in Multiply Connected Belief Networks

- Clustering
    - Transform the network by merging nodes
        - Probabilistically equivalent
        - Topologically different
- Cutset Conditioning
    - Transform by instantiating variables to values and re-evaluating the network
- Stochastic Simulation
    - Generate many consistent concrete models and approximate an exact evaluation

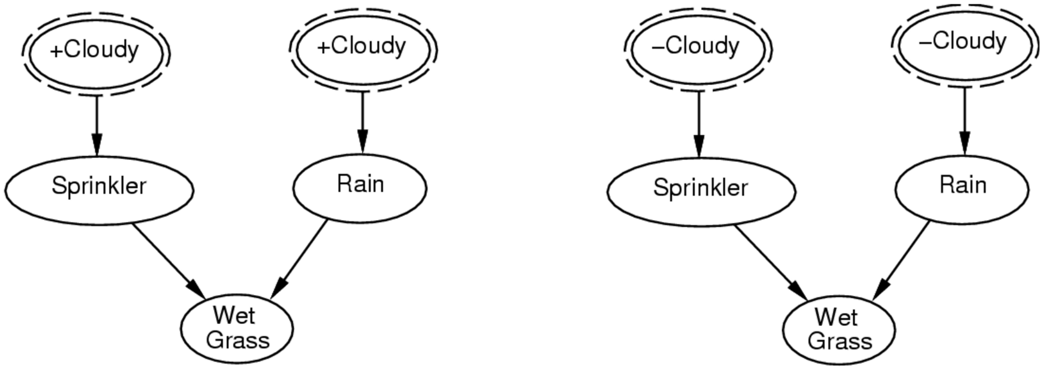
Clustering in Belief Networks



Try to change a multiply-connected graph into a singly-connected graph by merging nodes

Introduces complexity into each merged node, but the payoff is in the use of simpler inference mechanisms

Cutset Conditioning

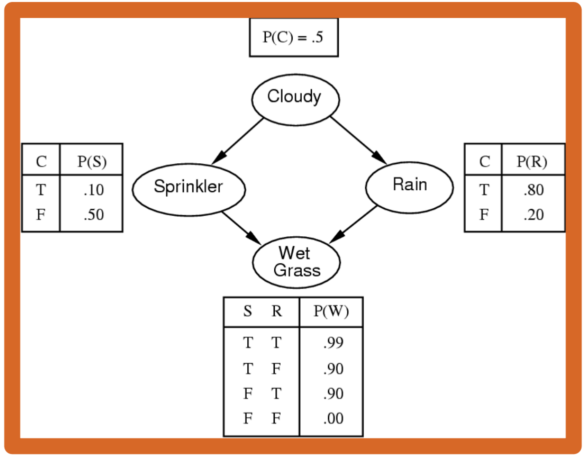


"Opposite" of clustering: transforms the network into several simpler graphs

- Number of graphs is exponential in the size of the cutset

Transform by instantiating variables to values and re-evaluating the network

Stochastic Simulation Methods



If we want to estimate P (WetGrass | Rain) then we just start at a root node and simulate a large number of trials

- Choose a value for Cloudy based on the probability
- Propagate this value through the network
- Count the number of instances of WetGrass and Rain to estimate the probability

Computing Probabilities and Reference Classes

Probability that the sun will exist tomorrow:

Undefined	There has never been an experiment that tested the existence of the sun tomorrow
1	In all previous experiments (previous days), the sun has continued to exist
$1-\varepsilon$	Where ε is the proportion of stars in the universe that go nova every day
$\frac{(d+1)}{(d+2)}$	Where d is the number of days that the sun has existed so far (Laplace formula)
???	Can be derived from the type, age, size, and temperature of the sun