CMSC 471
Fall 2012

Class #21

Thursday, November 8, 2012
Bayesian Networks

Kevin Winner, winnerk1@umbc.edu

Today's class

- Design documents
- Late HW4
- Uncollected midterms
- Bayesian networks
  - Network structure
  - Conditional probability tables
  - Conditional independence
- Inference in Bayesian networks
  - Exact inference
  - **Approximate inference (time permitting)**
- Review HW4 (time permitting)

Bayesian Networks

Chapter 14.1-14.4

Some material borrowed
from Lise Getoor

Bayesian Belief Networks (BNs)

- Definition: **BN = (DAG, CPD)**
  - **DAG**: directed acyclic graph (BN's **structure**)
    - **Nodes**: random variables (typically binary or discrete, but methods also exist to handle continuous variables)
    - **Arcs**: indicate probabilistic dependencies between nodes (*lack* of link signifies conditional independence)
  - **CPD**: conditional probability distribution (BN's **parameters**)
    - Conditional probabilities at each node, usually stored as a table (conditional probability table, or **CPT**)

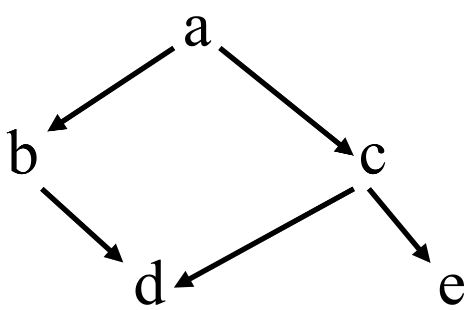
$$P(x_i \mid \pi_i) \text{ where } \pi_i \text{ is the set of all parent nodes of } x_i$$

  - Root nodes are a special case – no parents, so just use priors in CPD:

$$\pi_i = \varnothing, \text{ so } P(x_i \mid \pi_i) = P(x_i)$$

Example BN

P(A) = 0.001



P(B|A) = 0.3
P(B|¬A) = 0.001

P(C|A) = 0.2
P(C|¬A) = 0.005

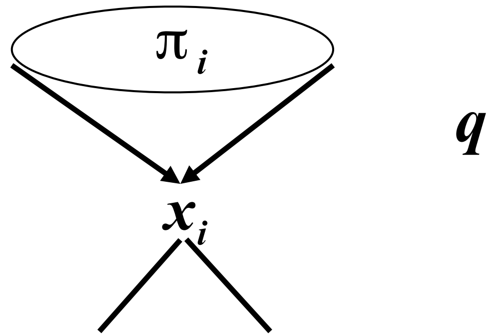
P(D|B,C) = 0.1
P(D|B,¬C) = 0.01
P(D|¬B,C) = 0.01
P(D|¬B,¬C) = 0.00001

P(E|C) = 0.4
P(E|¬C) = 0.002

Note that we only specify P(A) etc., not P(¬A), since they have to add to one

Conditional Independence and Chaining

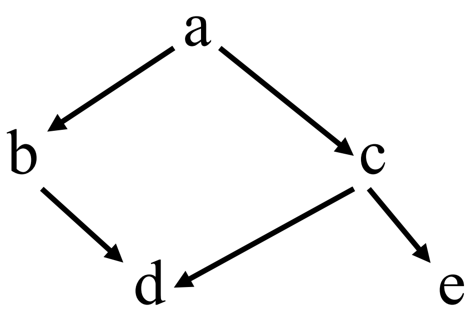
- Conditional independence assumption
  - $P(x_i \mid \pi_i, q) = P(x_i \mid \pi_i)$
    where q is any set of variables (nodes) other than x_i and its predecessors
  - π_i **blocks influence** of other nodes on x_i and its successors (q influences x_i only through variables in π_i)
  - With this assumption, the complete joint probability distribution of all variables in the network can be represented by (recovered from) local CPDs by chaining these CPDs:



$$P(x_1, \ldots, x_n) = \Pi_{i=1}^{n} P(x_i \mid \pi_i)$$

Chaining: Example

$P(A) = 0.001$



$P(B|A) = 0.3$
$P(B|\neg A) = 0.001$

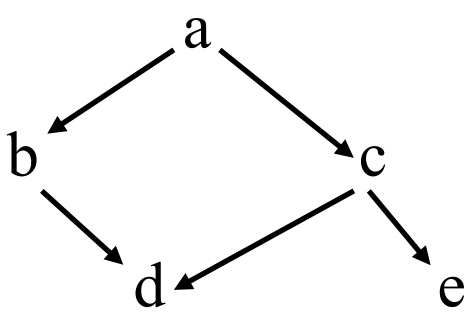
$P(C|A) = 0.2$
$P(C|\neg A) = 0.005$

$P(D|B,C) = 0.1$
$P(D|B,\neg C) = 0.01$
$P(D|\neg B,C) = 0.01$
$P(D|\neg B,\neg C) = 0.00001$

$P(E|C) = 0.4$
$P(E|\neg C) = 0.002$

Compute P(a, b, c, d, e) [the probability that all 5 are true]

Chaining: Example



Computing the joint probability for all variables is easy:

P(a, b, c, d, e)

=	P(e | a, b, ***c,*** d) P(a, b, c, d)	by the product rule
=	P(e | c) P(a, b, c, d)	by cond. indep. assumption
=	P(e | c) P(d | a, ***b, c***) P(a, b, c)	
=	P(e | c) P(d | b, c) P(c | ***a,*** b) P(a, b)	
=	P(e | c) P(d | b, c) P(c | a) P(b | a) P(a)	
=	0.0000024	

Inference Tasks

- **Simple queries:** Compute posterior marginal $P(X_i \mid E=e)$
  - E.g., P(NoGas | Gauge=empty, Lights=on, Starts=false)
- **Conjunctive queries:**
  - $P(X_i, X_j \mid E=e) = P(X_i \mid e=e)\, P(X_j \mid X_i, E=e)$
- **Optimal decisions:** *Decision networks* include utility information; probabilistic inference is required to find P(outcome | action, evidence)
- **Value of information:** Which evidence should we seek next?
- **Sensitivity analysis:** Which probability values are most critical?
- **Explanation:** Why do I need a new starter motor?

Approaches to Inference

- Exact inference
  - **Enumeration**
  - Belief propagation in polytrees
  - **Variable elimination**
  - Clustering / join tree algorithms
- Approximate inference
  - Stochastic simulation / sampling methods
  - Markov chain Monte Carlo methods
  - Genetic algorithms
  - Neural networks
  - Simulated annealing
  - Mean field theory

Direct Inference with BNs

- Instead of computing the joint, suppose we just want the probability for *one* variable
- Exact methods of computation:
  - **Enumeration**
  - **Variable elimination**

Inference by Enumeration

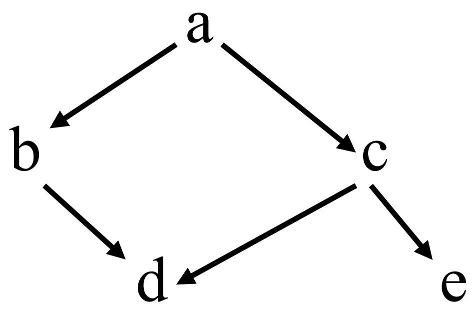
- Add all of the terms (atomic event probabilities) from the full joint distribution
- If **E** are the evidence (observed) variables and **Y** are the other (unobserved) variables, then:

  $$P(X|\mathbf{e}) = \alpha\, P(X, \mathbf{E}) = \alpha \sum P(X, \mathbf{E}, \mathbf{Y})$$

  $$\alpha = 1 / P(\mathbf{e})$$

- Each P(X, **E**, **Y**) term can be computed using the chain rule
- Computationally expensive!

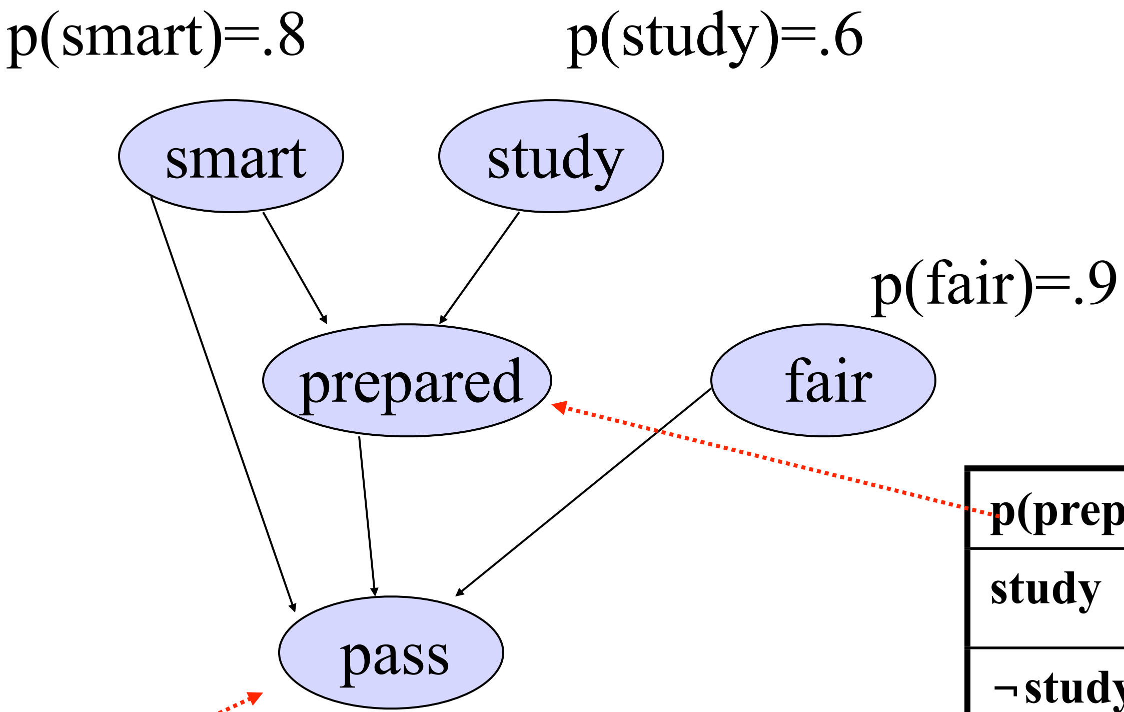
Example: Enumeration



- $P(x_i) = \Sigma_{\pi i} P(x_i \mid \pi_i) P(\pi_i)$
- Suppose we want P(D=true), and only the value of E is given as true
- $P(d|e) = \alpha \Sigma_{ABC} P(a, b, c, d, e)$ *(where $\alpha = 1/P(e)$)*

  $= \alpha \Sigma_{ABC} P(a) P(b|a) P(c|a) P(d|b,c) P(e|c)$
- With simple iteration to compute this expression, there's going to be a lot of repetition (e.g., P(e|c) has to be recomputed every time we iterate over C=true)

Exercise: Enumeration



p(prep|…)	smar	¬smart
study	.9	.7
¬study	.5	.1

p(pass|…)	smart		¬smart	
	prep	**¬prep**	**prep**	**¬prep**
fair	.9	.7	.7	.2
¬fair	.1	.1	.1	.1

Query: What is the probability that a student studied, given that they pass the exam?

Variable Elimination

- Basically just enumeration, but with caching of local calculations
- Linear for polytrees (singly connected BNs)
- Potentially exponential for multiply connected BNs

  ⇒**Exact inference in Bayesian networks is NP-hard!**
- Join tree algorithms are an extension of variable elimination methods that compute posterior probabilities for **all** nodes in a BN simultaneously

Variable Elimination Approach

General idea:

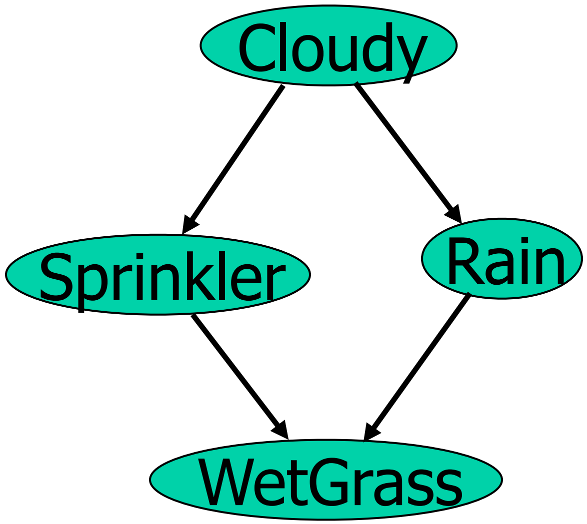
- Write query in the form

$$P(X_n, \mathbf{e}) = \sum_{x_k} \cdots \sum_{x_3} \sum_{x_2} \prod_i P(x_i \mid pa_i)$$

  – (Note that there is no α term here, because it's a conjunctive probability, not a conditional probability...)

- Iteratively
  – Move all irrelevant terms outside of innermost sum
  – Perform innermost sum, getting a new term
  – Insert the new term into the product

Variable Elimination: Example



$$P(w) = \sum_{r,s,c} P(w \mid r,s)P(r \mid c)P(s \mid c)P(c)$$

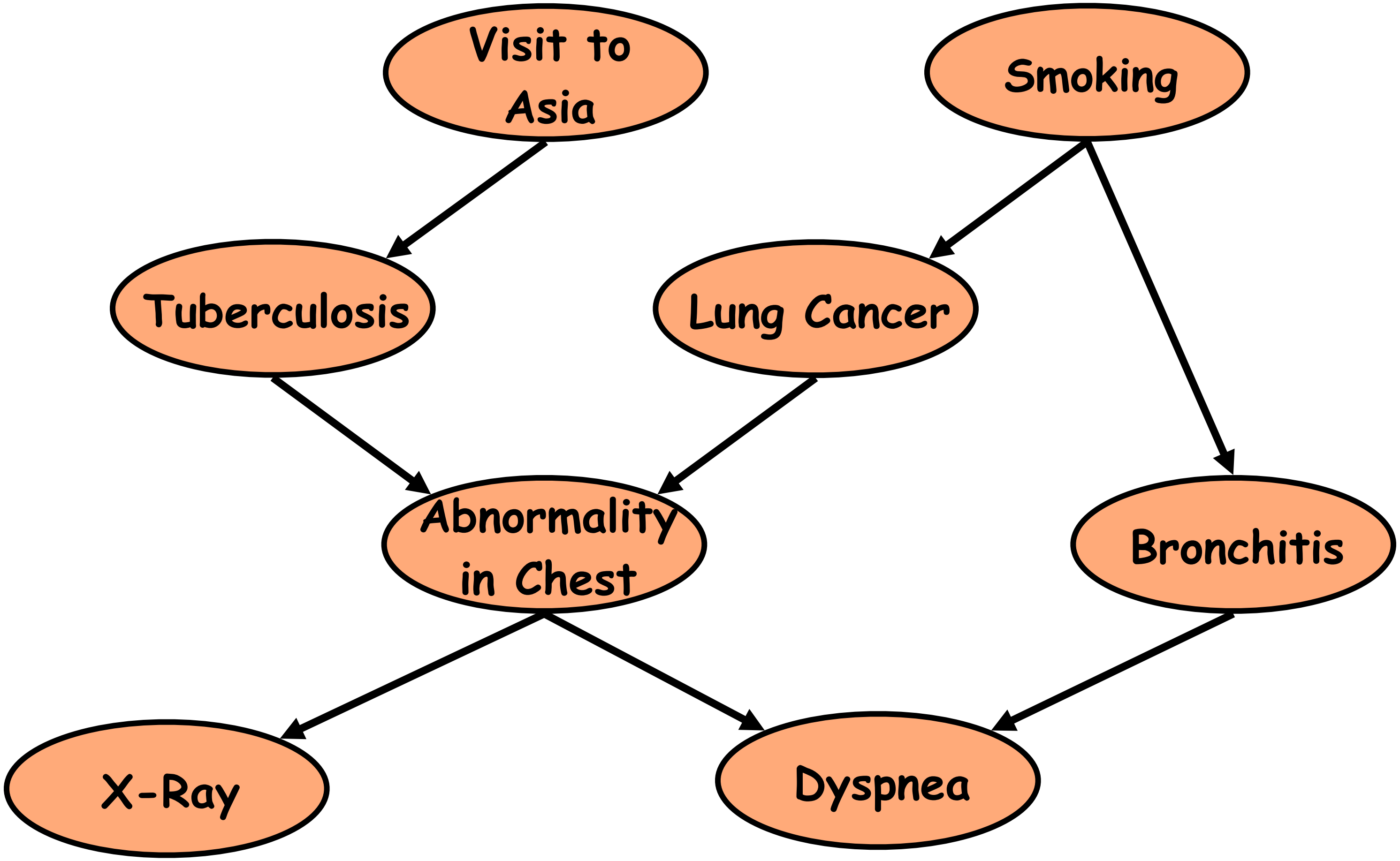
$$= \sum_{r,s} P(w \mid r,s) \sum_{c} P(r \mid c)P(s \mid c)P(c)$$

$f_1(r,s)$

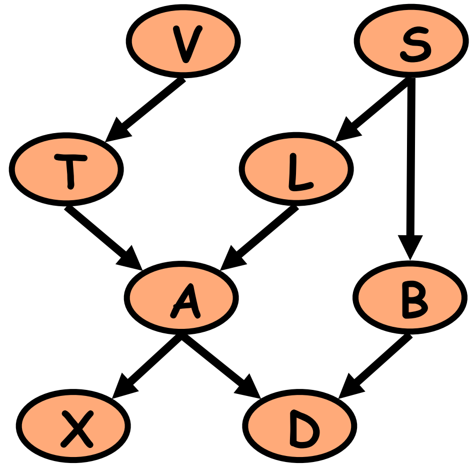
$$= \sum_{r,s} P(w \mid r,s) f_1(r,s)$$

A More Complex Example

- “Asia” network:



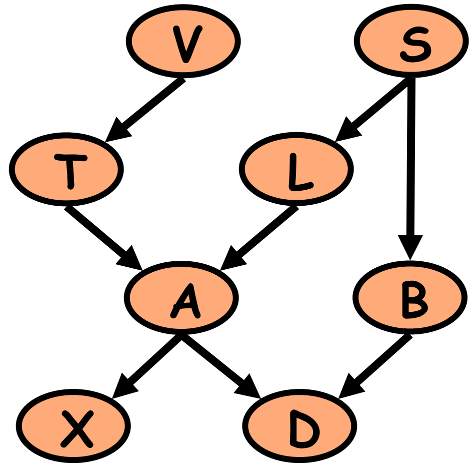
- We want to compute $P(d)$
- Need to eliminate: v,s,x,t,l,a,b



Initial factors

$$P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$

- We want to compute $P(d)$
- Need to eliminate: v,s,x,t,l,a,b



Initial factors

$$P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$

Eliminate: v

Compute: $$f_v(t) = \sum_v P(v)P(t|v)$$

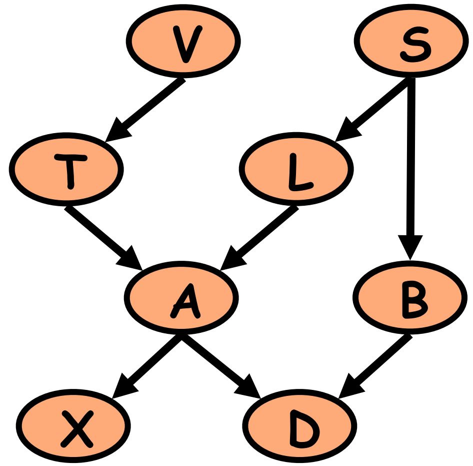
$$\Rightarrow f_v(t)P(s)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$

Note: $f_v(t) = P(t)$

In general, result of elimination is not necessarily a probability term

- We want to compute $P(d)$
- Need to eliminate: s,x,t,l,a,b

- Initial factors



$$P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$

$$\Rightarrow f_v(t)P(s)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$

Eliminate: s

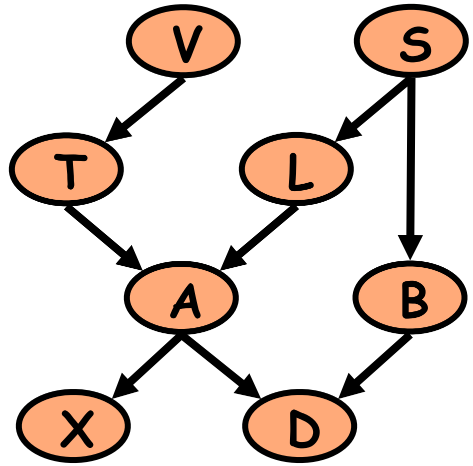
Compute: $$f_s(b,l) = \sum_s P(s)P(b|s)P(l|s)$$

$$\Rightarrow f_v(t)f_s(b,l)P(a|t,l)P(x|a)P(d|a,b)$$

Summing on s results in a factor with two arguments $f_s(b,l)$

In general, result of elimination may be a function of several variables

- We want to compute $P(d)$
- Need to eliminate: x,t,l,a,b



- Initial factors

$$P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$

$$\Rightarrow f_v(t)P(s)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$

$$\Rightarrow f_v(t)f_s(b,l)P(a|t,l)P(x|a)P(d|a,b)$$

Eliminate: x

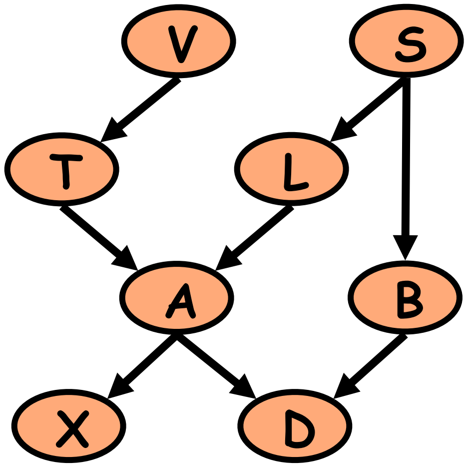
Compute: $$f_x(a) = \sum_x P(x|a)$$

$$\Rightarrow f_v(t)f_s(b,l)f_x(a)P(a|t,l)P(d|a,b)$$

Note: $f_x(a) = 1$ for all values of a !!

- We want to compute $P(d)$
- Need to eliminate: t,l,a,b

- Initial factors



$$P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$

$$\Rightarrow f_v(t)P(s)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$

$$\Rightarrow f_v(t)f_s(b,l)P(a|t,l)P(x|a)P(d|a,b)$$

$$\Rightarrow f_v(t)f_s(b,l)f_x(a)P(a|t,l)P(d|a,b)$$

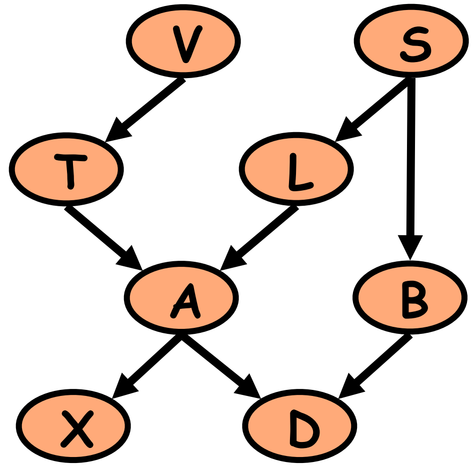
Eliminate: t

Compute: $$f_t(a,l) = \sum_t f_v(t)P(a|t,l)$$

$$\Rightarrow f_s(b,l)f_x(a)f_t(a,l)P(d|a,b)$$

- We want to compute P(d)
- Need to eliminate: l,a,b

- Initial factors



$$P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$

$$\Rightarrow f_v(t)P(s)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$

$$\Rightarrow f_v(t)f_s(b,l)P(a|t,l)P(x|a)P(d|a,b)$$

$$\Rightarrow f_v(t)f_s(b,l)f_x(a)P(a|t,l)P(d|a,b)$$

$$\Rightarrow f_s(b,l)f_x(a)f_t(a,l)P(d|a,b)$$

Eliminate: l

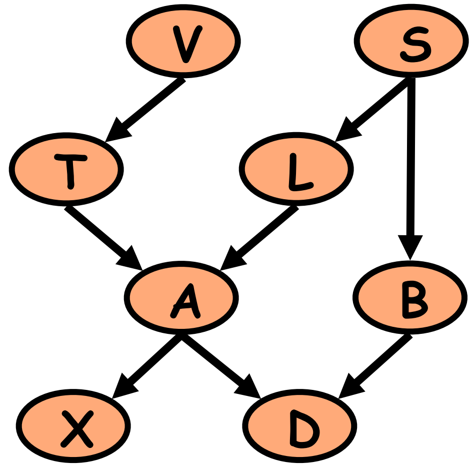
Compute: $f_l(a,b) = \sum_l f_s(b,l)f_t(a,l)$

$$\Rightarrow f_l(a,b)f_x(a)P(d|a,b)$$

Monday, November 12, 12

- We want to compute $P(d)$
- Need to eliminate: b

- Initial factors



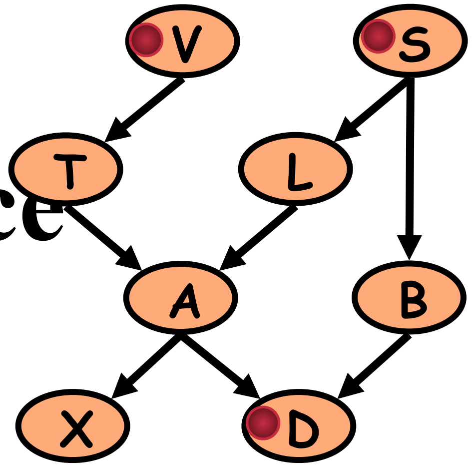
$$P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$
$$\Rightarrow f_v(t)P(s)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$
$$\Rightarrow f_v(t)f_s(b,l)P(a|t,l)P(x|a)P(d|a,b)$$
$$\Rightarrow f_v(t)f_s(b,l)f_x(a)P(a|t,l)P(d|a,b)$$
$$\Rightarrow f_s(b,l)f_x(a)f_t(a,l)P(d|a,b)$$
$$\Rightarrow f_l(a,b)f_x(a)P(d|a,b) \Rightarrow f_a(b,d) \Rightarrow f_b(d)$$

Eliminate: a,b

Compute:

$$f_a(b,d) = \sum_a f_l(a,b)f_x(a)p(d|a,b) \quad f_b(d) = \sum_b f_a(b,d)$$

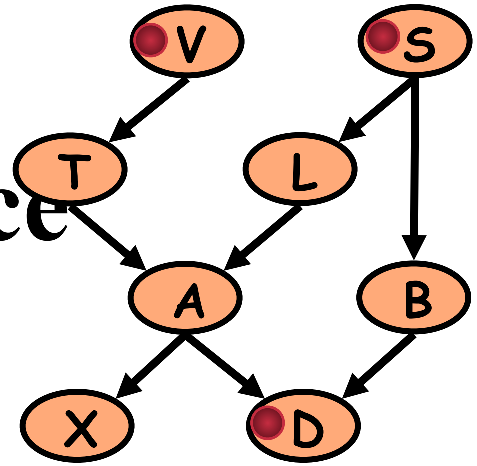
Dealing with Evidence



- How do we deal with evidence?

- Suppose we are give evidence V = t, S = f, D = t
- We want to compute P(L, V = t, S = f, D = t)

Dealing with Evidence



- We start by writing the factors:

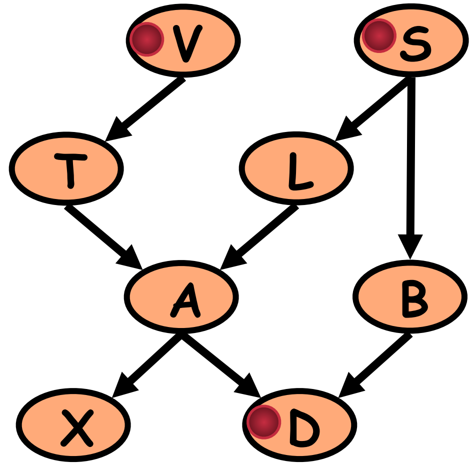
$$P(v)P(s)P(t|v)P(l|s)P(b|s)P(a|t,l)P(x|a)P(d|a,b)$$

- Since we know that V = t, we don't need to eliminate V
- Instead, we can replace the factors P(V) and P(T|V) with

$$f_{P(V)} = P(V = t) \quad f_{p(T|V)}(T) = P(T|V = t)$$

- These "select" the appropriate parts of the original factors given the evidence
- Note that $f_{p(V)}$ is a constant, and thus does not appear in elimination of other variables

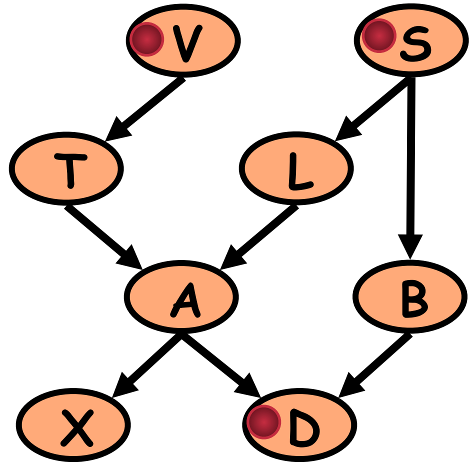
Dealing with Evidence



- Given evidence V = t, S = f, D = t
- Compute P(L, V = t, S = f, D = t)
- Initial factors, after setting evidence:

$$f_{P(v)} f_{P(s)} f_{P(t|v)}(t) f_{P(l|s)}(l) f_{P(b|s)}(b) P(a|t,l) P(x|a) f_{P(d|a,b)}(a,b)$$

Dealing with Evidence



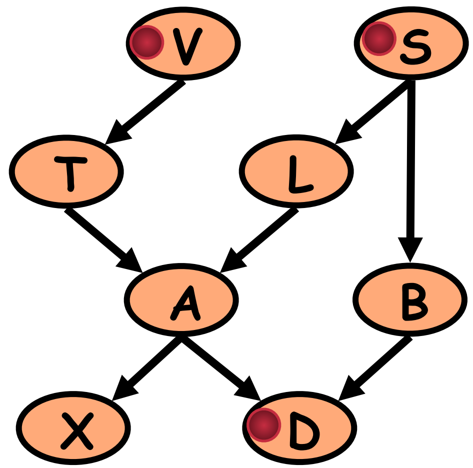
- Given evidence V = t, S = f, D = t
- Compute P(L, V = t, S = f, D = t)
- Initial factors, after setting evidence:

$$f_{P(v)} f_{P(s)} f_{P(t|v)}(t) f_{P(l|s)}(l) f_{P(b|s)}(b) P(a|t,l) P(x|a) f_{P(d|a,b)}(a,b)$$

- Eliminating x, we get

$$f_{P(v)} f_{P(s)} f_{P(t|v)}(t) f_{P(l|s)}(l) f_{P(b|s)}(b) P(a|t,l) f_x(a) f_{P(d|a,b)}(a,b)$$

Dealing with Evidence



- Given evidence $V = t, S = f, D = t$
- Compute $P(L, V = t, S = f, D = t)$
- Initial factors, after setting evidence:

$$f_{P(v)} f_{P(s)} f_{P(t|v)}(t) f_{P(l|s)}(l) f_{P(b|s)}(b) P(a|t,l) P(x|a) f_{P(d|a,b)}(a,b)$$

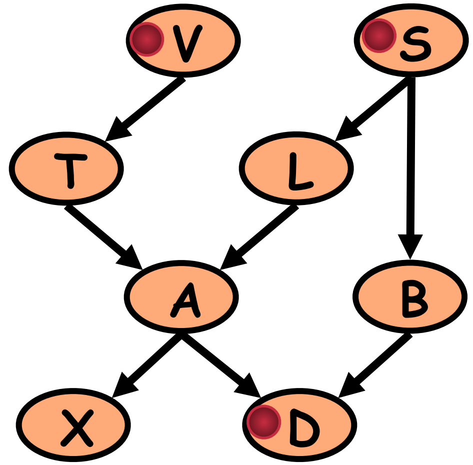
- Eliminating x, we get

$$f_{P(v)} f_{P(s)} f_{P(t|v)}(t) f_{P(l|s)}(l) f_{P(b|s)}(b) P(a|t,l) f_x(a) f_{P(d|a,b)}(a,b)$$

- Eliminating t, we get

$$f_{P(v)} f_{P(s)} f_{P(l|s)}(l) f_{P(b|s)}(b) f_t(a,l) f_x(a) f_{P(d|a,b)}(a,b)$$

Dealing with Evidence



- Given evidence $V = t, S = f, D = t$
- Compute $P(L, V = t, S = f, D = t)$
- Initial factors, after setting evidence:

$$f_{P(v)} f_{P(s)} f_{P(t|v)}(t) f_{P(l|s)}(l) f_{P(b|s)}(b) P(a|t,l) P(x|a) f_{P(d|a,b)}(a,b)$$

- Eliminating x, we get

$$f_{P(v)} f_{P(s)} f_{P(t|v)}(t) f_{P(l|s)}(l) f_{P(b|s)}(b) P(a|t,l) f_x(a) f_{P(d|a,b)}(a,b)$$

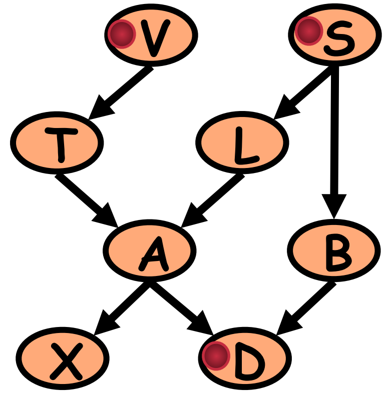
- Eliminating t, we get

$$f_{P(v)} f_{P(s)} f_{P(l|s)}(l) f_{P(b|s)}(b) f_t(a,l) f_x(a) f_{P(d|a,b)}(a,b)$$

- Eliminating a, we get

$$f_{P(v)} f_{P(s)} f_{P(l|s)}(l) f_{P(b|s)}(b) f_a(b,l)$$

Dealing with Evidence



- Given evidence V = t, S = f, D = t
- Compute P(L, V = t, S = f, D = t)
- Initial factors, after setting evidence:

$$f_{P(v)} f_{P(s)} f_{P(t|v)}(t) f_{P(l|s)}(l) f_{P(b|s)}(b) P(a|t,l) P(x|a) f_{P(d|a,b)}(a,b)$$

- Eliminating x, we get

$$f_{P(v)} f_{P(s)} f_{P(t|v)}(t) f_{P(l|s)}(l) f_{P(b|s)}(b) P(a|t,l) f_x(a) f_{P(d|a,b)}(a,b)$$

- Eliminating t, we get

$$f_{P(v)} f_{P(s)} f_{P(l|s)}(l) f_{P(b|s)}(b) f_t(a,l) f_x(a) f_{P(d|a,b)}(a,b)$$

- Eliminating a, we get

$$f_{P(v)} f_{P(s)} f_{P(l|s)}(l) f_{P(b|s)}(b) f_a(b,l)$$

- Eliminating b, we get

$$f_{P(v)} f_{P(s)} f_{P(l|s)}(l) f_b(l)$$

Variable Elimination Algorithm

- Let $X_1, \ldots, X_m$ be an ordering on the non-query variables
- For i = m, …, 1 $\sum_{X_1} \sum_{X_2} \cdots \sum_{X_m} \prod_j P(X_j \mid \text{Parents}(X_j))$
  - Leave in the summation for X_i only factors mentioning X_i
  - Multiply the factors, getting a factor that contains a number for each value of the variables mentioned, including X_i
  - Sum out X_i, getting a factor f that contains a number for each value of the variables mentioned, not including X_i
  - Replace the multiplied factor in the summation