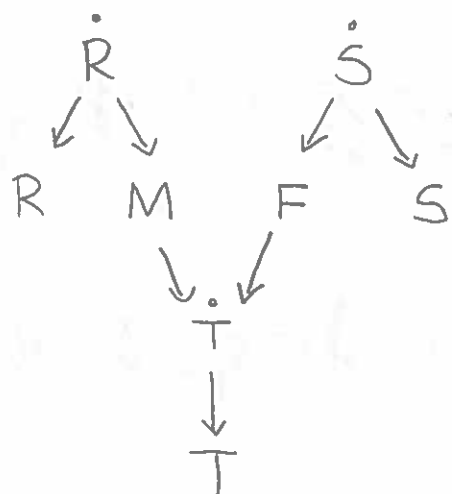


EXACT INFERENCE IN BAYESIAN NETWORKS

① Consider the following Bayesian network for blood types:



where:

- $\dot{R}, \dot{S}, \dot{T}$ are the blood genotypes for Rhonda, Sam, and Tim ($\in \{AA, AB, AO, BB, BO, OO\}$).
- R, S, T are the blood types for Rhonda, Sam, and Tim ($\in \{A, B, AB, O\}$).
- M, F are the genes passed to Tim by Rhonda (his Mother) and Sam (his Father), $M, F \in \{A, B, O\}$.

② What if we want to compute the probability of Rhonda having blood type AB, given that we know Tim has blood type A?

EXACT INFERENCE

③ We could try reasoning about it intuitively...

well, if Tim has type A,
then Rhonda has genotype AA, AB, AO, or BO...
and I guess AA is more likely than AB,
but AB is more likely than BO...



But it's a bit much. Plus what about harder questions like: what is the probability of Rhonda having blood type AB, given we know Tim has blood type A, AND Tim's cousin on his father's side has blood type O?

④ We need a method to automate complicated probability questions.

This problem is called probabilistic inference.

Note that $P(R=AB|T=A) = \frac{P(R=AB, T=A)}{P(T=A)}$, so if we

can answer joint queries like $P(R=AB, T=A)$ and $P(T=A)$, then we can immediately also answer conditional queries

EXACT INFERENCE

⑤ On the surface, it doesn't look that hard to solve:

$$P(R=AB, T=A) = \sum_{\dot{r}} \sum_{\dot{s}} \sum_{\dot{t}} \sum_m \sum_f \sum_s P(R=AB, T=A, \dot{r}, \dot{s}, \dot{t}, m, f, s)$$

But this expression has $6 \cdot 6 \cdot 6 \cdot 3 \cdot 3 \cdot 4 = 7776$ terms. While a computer could handle that, I certainly can't. Plus, as the number of variables in the Bayesian network increases, the number of terms grows exponentially.

⑥ We can begin by replacing the joint distribution by the factored representation assumed by the Bayesian network.

$$P(r, t)$$

$$= \sum_{\dot{r}} \sum_{\dot{s}} \sum_{\dot{t}} \sum_m \sum_f \sum_s P(\dot{r}) P(\dot{s}) \underbrace{P(r|\dot{r})}_{\text{this is just a function } g_3(r, \dot{r}) = P(r|\dot{r})} P(m|\dot{r}) P(f|\dot{s}) \underbrace{P(s|\dot{s}) P(\dot{t}|m, f) P(t|\dot{t})}_{\text{this is just a function } g_6(\dot{t}, m, f) = P(\dot{t}|m, f)}$$

EXACT INFERENCE

⑦ So $P(r, t)$ is just a sum of products:

$$P(r, t) = \sum_i \sum_{\dot{s}} \sum_{\dot{t}} \sum_m \sum_f \sum_s g_1(i) g_2(\dot{s}) g_3(r, i) g_4(m, i) g_5(f, \dot{s}) g_6(s, \dot{s}) g_7(\dot{t}, m, f) g_8(t, \dot{t})$$

How can we compute this efficiently?

⑧ The first thing to notice is we can arrange the summations in whatever order we want, e.g.

$$P(r, t) = \sum_m \sum_f \sum_{\dot{t}} \sum_i \sum_{\dot{s}} \sum_s g_1(i) g_2(\dot{s}) g_3(r, i) g_4(m, i) g_5(f, \dot{s}) g_6(s, \dot{s}) g_7(\dot{t}, m, f) g_8(t, \dot{t})$$

⑨ Now, we can move the innermost summation inward:

$$P(r, t) = \sum_m \sum_f \sum_{\dot{t}} \sum_i \sum_{\dot{s}} g_1(i) g_2(\dot{s}) g_3(r, i) g_4(m, i) g_5(f, \dot{s}) g_7(\dot{t}, m, f) g_8(t, \dot{t}) \sum_s g_6(s, \dot{s})$$

Note that $\sum_s g_6(s, \dot{s})$ is a function of \dot{s} , so we can define a new function $h_1(\dot{s}) = \sum_s g_6(s, \dot{s})$ and rewrite:

$$P(r, t) = \sum_m \sum_f \sum_{\dot{t}} \sum_i \sum_{\dot{s}} g_1(i) g_2(\dot{s}) g_3(r, i) g_4(m, i) g_5(f, \dot{s}) g_7(\dot{t}, m, f) g_8(t, \dot{t}) h_1(\dot{s})$$

We have "eliminated" the summation over variable S (at least, we're hiding the fact that it exists.)

EXACT INFERENCE

⑩ We can continue to do this for each summation:

$$P(r, t) = \sum_m \sum_f \sum_{\dot{t}} \sum_{\dot{r}} g_1(\dot{r}) g_3(r, \dot{r}) g_4(m, \dot{r}) g_7(\dot{t}, m, f) g_8(t, \dot{t}) \sum_{\dot{s}} g_2(\dot{s}) g_5(f, \dot{s}) h_1(\dot{s})$$

$$= \sum_m \sum_f \sum_{\dot{t}} \sum_{\dot{r}} g_1(\dot{r}) g_3(r, \dot{r}) g_4(m, \dot{r}) g_7(\dot{t}, m, f) g_8(t, \dot{t}) h_2(f)$$

$$\left[\text{letting } h_2(f) = \sum_{\dot{s}} g_2(\dot{s}) g_5(f, \dot{s}) h_1(\dot{s}) \right]$$

$$= \sum_m \sum_f \sum_{\dot{t}} g_7(\dot{t}, m, f) g_8(t, \dot{t}) h_2(f) \sum_{\dot{r}} g_1(\dot{r}) g_3(r, \dot{r}) g_4(m, \dot{r})$$

$$= \sum_m \sum_f \sum_{\dot{t}} g_7(\dot{t}, m, f) g_8(t, \dot{t}) h_2(f) h_3(r, m)$$

$$\left[\text{letting } h_3(r, m) = \sum_{\dot{r}} g_1(\dot{r}) g_3(r, \dot{r}) g_4(m, \dot{r}) \right]$$

$$= \sum_m \sum_f h_2(f) h_3(r, m) \sum_{\dot{t}} g_7(\dot{t}, m, f) g_8(t, \dot{t})$$

$$= \sum_m \sum_f h_2(f) h_3(r, m) h_4(t, m, f)$$

$$\left[\text{letting } h_4(t, m, f) = \sum_{\dot{t}} g_7(\dot{t}, m, f) g_8(t, \dot{t}) \right]$$

$$= \sum_m h_3(r, m) \sum_f h_2(f) h_4(t, m, f)$$

$$= \sum_m h_3(r, m) h_5(t, m)$$

$$\left[\text{letting } h_5(t, m) = \sum_f h_2(f) h_4(t, m, f) \right]$$

$$= h_6(r, t)$$

$$\left[\text{letting } h_6(r, t) = \sum_m h_3(r, m) h_5(t, m) \right]$$

EXACT INFERENCE

11) What have we done? Well, we have laid out a strategy for computing $P(r, t)$.

- | | <u>num operations</u> |
|--|---|
| - $\forall \dot{s}$, compute $h_1(\dot{s}) = \sum_s g_6(s, \dot{s})$ | $ \dot{s} \cdot S $ |
| - $\forall F$, compute $h_2(F) = \sum_{\dot{s}} g_2(\dot{s}) g_5(F, \dot{s}) h_1(\dot{s})$ | $ \dot{s} \cdot F $ |
| - $\forall r, m$, compute $h_3(r, m) = \sum_{\dot{r}} g_1(\dot{r}) g_3(r, \dot{r}) g_4(m, \dot{r})$ | $ \dot{r} \cdot R \cdot M $ |
| - $\forall t, m, F$, compute $h_4(t, m, F) = \sum_{\dot{t}} g_7(\dot{t}, m, F) g_8(t, \dot{t})$ | $ \dot{t} \cdot T \cdot M \cdot F $ |
| - $\forall t, m$, compute $h_5(t, m) = \sum_F h_2(F) h_4(t, m, F)$ | $ F \cdot T \cdot M $ |
| - $\forall r, t$, compute $h_6(r, t) = \sum_m h_3(r, m) h_5(t, m)$ | $ M \cdot R \cdot T $ |

Then return $P(r, t) = h_6(r, t)$.

Notice that:

$$|S| \cdot |\dot{S}| + |\dot{S}| \cdot |F| + |\dot{R}| \cdot |R| \cdot |M| + |\dot{T}| \cdot |T| \cdot |M| \cdot |F| + |F| \cdot |T| \cdot |M| + |M| \cdot |R| \cdot |T|$$

is much better than directly computing the sum of products; which is:

$$|R| \cdot |\dot{R}| \cdot |S| \cdot |\dot{S}| \cdot |T| \cdot |\dot{T}| \cdot |M| \cdot |F|$$

(414 operations, versus 124416 operations)

EXACT INFERENCE

⑫ If we execute this strategy, then we can get the probability that Rhonda has blood type AB and her son has type A.

$$h_1:$$

\dot{s}	$h_1(\dot{s})$
AA	1
AB	1
AO	1
BB	1
BO	1
OO	1

$$h_5:$$

t	m	$h_5(t, m)$
A	A	$\frac{2}{3}$
A	B	0
A	O	$\frac{1}{3}$

$$h_2:$$

f	$h_2(f)$
A	$\frac{1}{3}$
B	$\frac{1}{3}$
O	$\frac{1}{3}$

$$h_6:$$

r	t	$h_6(r, t)$
AB	A	$\frac{1}{18}$

$$P(R=AB, T=A) = \frac{1}{18} !$$



$$h_3:$$

r	m	$h_3(r, m)$
AB	A	$\frac{1}{12}$
AB	B	$\frac{1}{12}$
AB	O	0

$$h_4:$$

t	m	f	$h_4(t, m, f)$
A	A	A	1
A	A	B	0
A	A	O	1
A	B	A	0
A	B	B	0
A	B	O	0
A	O	A	1
A	O	B	0
A	O	O	0

We can use the same technique to compute $P(T=A) = \frac{1}{3}$. This gives us the probability that Rhonda's blood type is AB given Tim's blood type is A:

$$P(R=AB | T=A) = \frac{P(R=AB, T=A)}{P(T=A)} = \frac{1}{6}$$

EXACT INFERENCE

- ⑬ To analyze the running time of this "variable elimination" algorithm, let d be the maximum size of any variable domain:

compute all values of

num ops

$$h_1(s)$$

$$|S| \cdot |\dot{S}| \leq d^2 \leq d^4$$

$$h_2(f)$$

$$|\dot{S}| \cdot |F| \leq d^2 \leq d^4$$

$$h_3(r, m)$$

$$|\dot{R}| \cdot |R| \cdot |M| \leq d^3 \leq d^4$$

$$h_4(t, m, f)$$

$$|\dot{T}| \cdot |T| \cdot |M| \cdot |F| \leq d^4 \leq d^4$$

$$h_5(t, m)$$

$$|F| \cdot |T| \cdot |M| \leq d^3 \leq d^4$$

$$h_6(r, t)$$

$$|M| \cdot |R| \cdot |T| \leq d^3 \leq d^4$$

$$\leq n \cdot d^4$$

↑

number
of variables

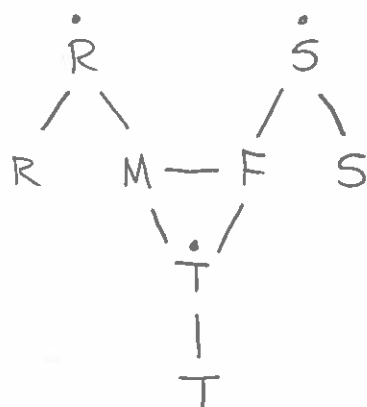
- ⑭ If we let w be the maximum number of variables in our h -functions (plus one), then the running time is

$$O(n \cdot d^w)$$

This quantity w is called the width of the elimination order.

EXACT INFERENCE

- ⑮ It can be convenient to view this graphically. Create an undirected graph whose vertices are the variables. There is an edge between two vertices iff they appear in the same g -function:



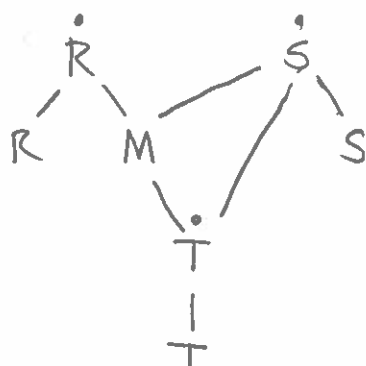
$$P(r|t) = \sum_m \sum_f \sum_t \sum_r \sum_s \sum_s g_1(r) g_2(s) g_3(r, r) g_4(m, r) g_5(f, s) g_6(s, s) g_7(t, m, f) g_8(t, t)$$

- ⑯ Suppose we choose to first eliminate F :

$$P(r|t) = \sum_m \sum_t \sum_r \sum_s \sum_s g_1(r) g_2(s) g_3(r, r) g_4(m, r) g_6(s, s) g_8(t, t) h_1(m, s, t)$$

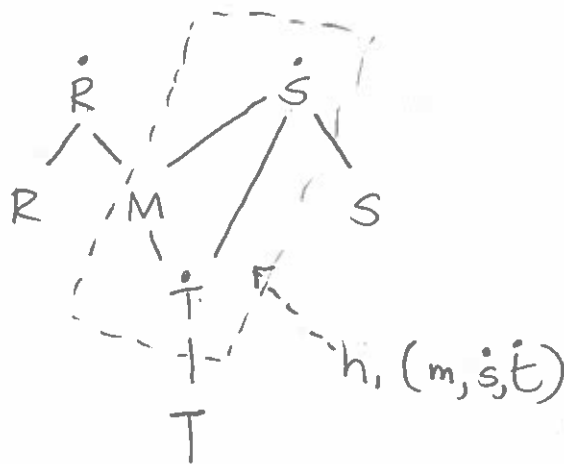
$$\text{where } h_1(m, s, t) = \sum_f g_5(f, s) g_7(t, m, f)$$

We create an h -function over m, s, t . Graphically, these are all the nodes that F was connected to. The new graph becomes:



EXACT INFERENCE

- ⑪ Note that "eliminating" F corresponds (graphically) to removing F and connecting all the nodes it was connected to. The size of this new clique corresponds to the number of variables in the new h -function:



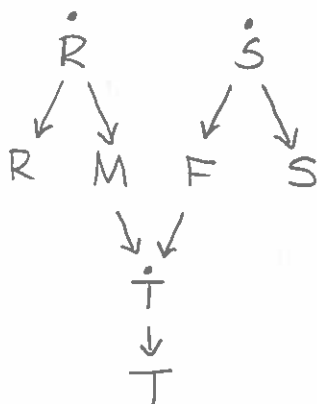
- ⑫ Therefore, we can compute the width of an elimination order as a graph algorithm:

- let $G = (V, E)$, where V corresponds to the variables and two vertices are connected if they co-occur in a g -function.
- $w = 0$
- iterate through V_1, \dots, V_n :
 - $w = \max(w, 1 + \text{num vertices } V_i \text{ is connected to})$
 - remove V_i from G and connect all its neighbors
- return w

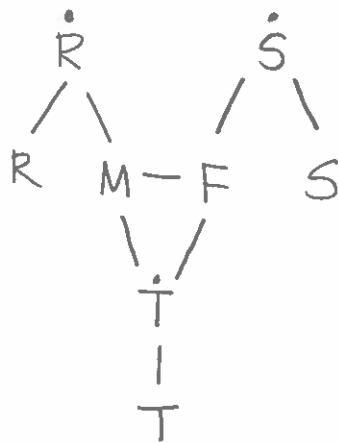
EXACT INFERENCE

⑱ So given a Bayesian network, we can predict the runtime of probabilistic inference based solely on its graph structure, and the domain size of its variables.

(i) Take the network:



(ii) Create an undirected graph where two vars are connected if they co-appear in a conditional probability function



this corresponds to connecting all parents of each node

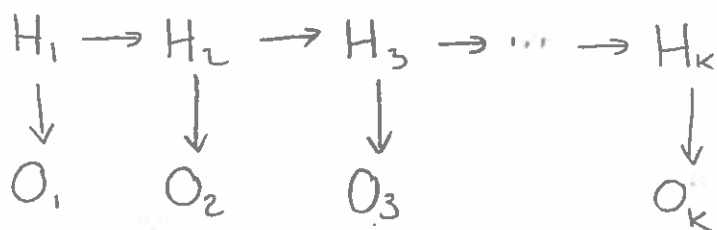
this is called a moral graph

(iii) Compute the width w of your selected elimination order (as described in ⑱)

(iv) The runtime of variable elimination is $O(n \cdot d^w)$, where n is the number of variables, and d is the maximum domain size.

EXACT INFERENCE

20) This means that certain Bayesian networks enjoy good performance, like the following:

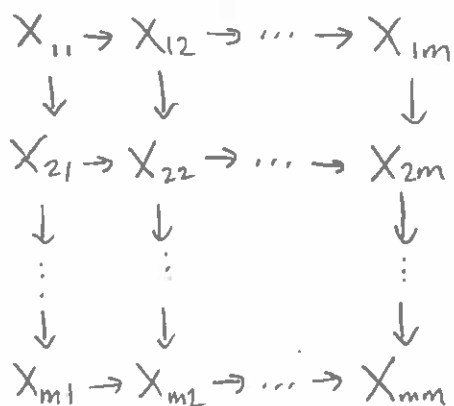


which is sometimes called a Hidden Markov Model and has an elimination order of width 2 (i.e. $O_1, H_1, O_2, H_2, \dots$)

Thus the runtime of variable elimination is $O(k \cdot d^2)$.

If the variables are all binary (or constant-size domain), then this is $O(k)$.

21) Unfortunately, others have guaranteed poor performance, like this grid network:



whose best width is $O(m^2)$, thus the runtime of VE is $O(m^2 \cdot d^{m^2})$