Tutorial 8 - Factor Elimination and Jointrees

COMP9418 – Advanced Topics in Statistical Machine Learning

Lecturer: Gustavo Batista

Lecture: Factor Elimination and Jointrees

Topic: Questions from lecture topics

Last revision: Tuesday 2nd November, 2021 at 15:16

Question 1

Answer the following queries with respect to the Bayesian network below:

$P(A)$

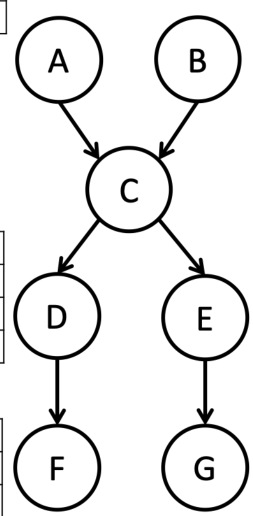
+a	.6
-a	.4

$P(B)$

+b	.5
-b	.5

$P(C|A,B)$

+a	+b	+c	.9
+a	+b	-c	.1
+a	-b	+c	.1
+a	-b	-c	.9
-a	+b	+c	.5
-a	+b	-c	.5
-a	-b	+c	.3
-a	-b	-c	.7



$P(D|C)$

+c	+d	.2
+c	-d	.8
-c	+d	.7
-c	-d	.3

$P(E|C)$

+c	+e	.1
+c	-e	.9
-c	+e	.2
-c	-e	.8

$P(F|D)$

+d	+f	.7
+d	-f	.3
-d	+f	.6
-d	-f	.4

$P(G|E)$

+e	+g	.2
+e	-g	.8
-e	+g	.8
-e	-g	.2

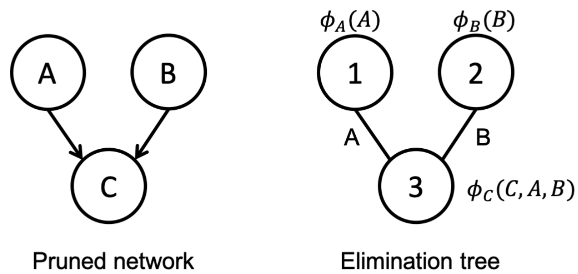
1. $P(B,C)$
2. $P(C, D = true)$
3. $P(C|D = true, E = true)$

You may prune the network before attempting each computation and use any inference method you find most appropriate.

Answer

1. $P(B,C)$

Since the query has no evidence, we can prune only nodes of the Bayesian network. We are free to choose our inference method. For this marginal, we decided to use an elimination tree. Our elimination tree has three nodes, one for each factor of the pruned Bayesian network, as shown below.



If we select node three as root, we can eliminate the remaining factors since they have only one neighbour.

We can eliminate the factor 1 first:

$\tau_1(C, A, B) = \phi_A(A)\phi_C(C, A, B)$

This results in the following factor:

A	B	C	$\tau_1(C, A, B)$
+a	+b	+c	.54
+a	+b	-c	.06
+a	-b	+c	.06
+a	-b	-c	.54
-a	+b	+c	.20
-a	+b	-c	.20
-a	-b	+c	.12
-a	-b	-c	.28

And eliminate factor 2:

$\tau_2(C, A, B) = \phi_B(B)\phi_C(C, A, B)$

This results in the following factor:

A	B	C	$\tau_2(C, A, B)$
+a	+b	+c	.27
+a	+b	-c	.03
+a	-b	+c	.03
+a	-b	-c	.27
-a	+b	+c	.10
-a	+b	-c	.10
-a	-b	+c	.06
-a	-b	-c	.14

Eliminating variable A gives the desired result:

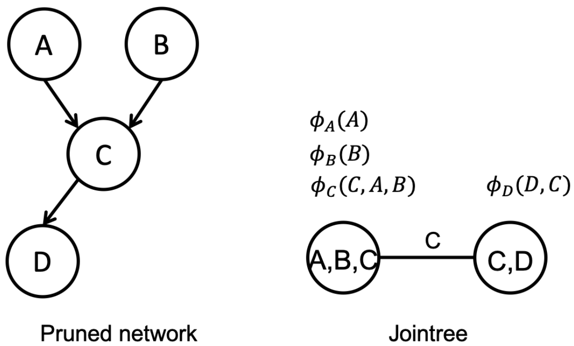
B	C	$P(B, C)$
+b	+c	.37
+b	-c	.13
-b	+c	.09

B	C	$P(B,C)$
-b	-c	.41

2. $P(C, D = true)$

This time, we can prune edges and nodes. Again, we are free to choose our inference method. Let's try a jointree this time. We will handcraft the jointree. It is not difficult, because the pruned network is small, as the figure below illustrates. We will need a single cluster for nodes A, B and C because of the factor $P(C|A, B)$ mentions these three variables. We will also need a cluster for variables C and D. In this case, the separator will be just variable C. We can check that this simple jointree has the required properties.

The resulting Bayesian network and a corresponding jointree are shown below:



Since this is a small example with a single query, we will set the evidence by eliminating rows. We set the evidence $D = true$ to the only remaning factor that includes D, $\phi_D(D, C)$. Its final format is:

C	D	$\phi_D(+d, C)$
+c	+d	.2
-c	+d	.7

We set CD as the root node. But we notice that choosing ABC as root will lead to the same results. Now, we can send a message from cluster ABC (1) to cluster CD (2). The message $M_{1,2}$ has domain C.

$M_{1,2}(C) = \sum_{A,B} \phi_A(A)\phi_B(B)\phi_C(C, A, B)$

C	$M_{1,2}(C)$
+c	.46
-c	.54

The message $M_{1,2}$ is incorporated to cluster CD (2).

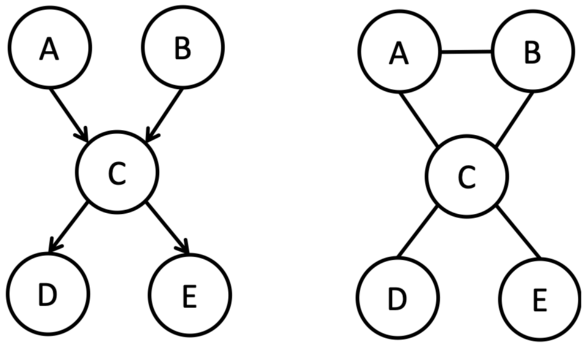
$\beta_2(C, D) = M_{1,2}\phi_D(+d, C)$

C	$M_{1,2}(C)$
+c	.092
-c	.378

Therefore, $P(C, D = true) = 0.092$. Normalizing, we get $P(C|D = true) = 0.1957$.

3. $P(C, D = true, E = true)$

For this one, let's use variable elimination. The next figure shows the pruned graph as well as its corresponding interaction graph.



Pruned network Interaction Graph

We have the following factors:

$\phi_A(A), \phi_B(B), \phi_C(C, A, B), \phi_D(D, C), \phi_E(E, C)$

We start setting the evidence $D = true$ and $E = true$ to the factors.

C	D	$\phi_D(+d, C)$
+c	+d	.2
-c	+d	.7

C	E	$\phi_D(+e, C)$
+c	+e	.1
-c	+e	.2

We use min-degree to heuristic to drive the variable elimination procedure. We start with the variable B since it not query nor evidence.

$\sigma_1(C, A, B) = \phi_B(B), \phi_C(C, A, B)$

A	B	C	$\sigma_1(A, B, C)$
+a	+b	+c	.45
+a	+b	-c	.05
+a	-b	+c	.05
+a	-b	-c	.45
-a	+b	+c	.25
-a	+b	-c	.25
-a	-b	+c	.15
-a	-b	-c	.35

and we eliminate variable B:

$\tau_1(A, C) = \sum_B \sigma_1(C, A, B)$

A	C	$\tau_1(A,C)$
+a	+c	.50
+a	-c	.50
-a	+c	.40
-a	-c	.60

We have the following factors now:

$\phi_A(A), \phi_D(+d, C), \phi_E(+e, C), \tau_1(A, C)$

And we will eliminate A now:

$\sigma_2(C, A) = \phi_A(A)\tau_1(A, C)$

A	C	$\sigma_2(A,C)$
+a	+c	.30
+a	-c	.30
-a	+c	.16
-a	-c	.24

$\tau_2(C) = \sum_A \sigma_2(C, A)$:

C	$\tau_2(C)$
+c	.46
-c	.54

Now joining together all remaining factors:

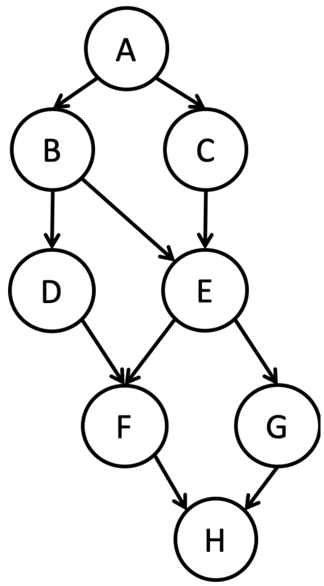
$\tau_3(C) = \tau_2(C)\phi_D(+d, C)\phi_E(+e, C)$

C	$\tau_3(C)$
+c	.0092
-c	.0756

Therefore, $P(C, D = true, E = true) = \tau_3(C)$. Normalizing, we get $P(C|D = true, E = true) = 0.108$.

Question 2

Consider the Bayesian network in the figure below. Construct an elimination tree for the Bayesian network CPTs that has the smallest width possible and assigns at most one CPT to each tree node. Compute the separators, clusters and width of the elimination tree. It may be useful to know that this network has the following jointree $ABC - BCE - BDE - DEF - EFG - FGH$.

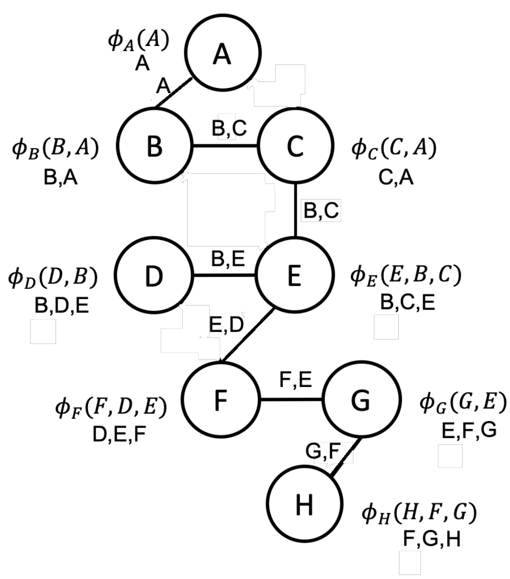


Answer

The Bayesian network has factors with three variables associated with the variables E, F and H. Therefore, we know the treewidth cannot be smaller than 2. At the same time, the jointree has width equal 2, and consequently it is optimal. Therefore, we need to find an elimination tree that is also optimal, i.e., the largest cluster must have size of 3.

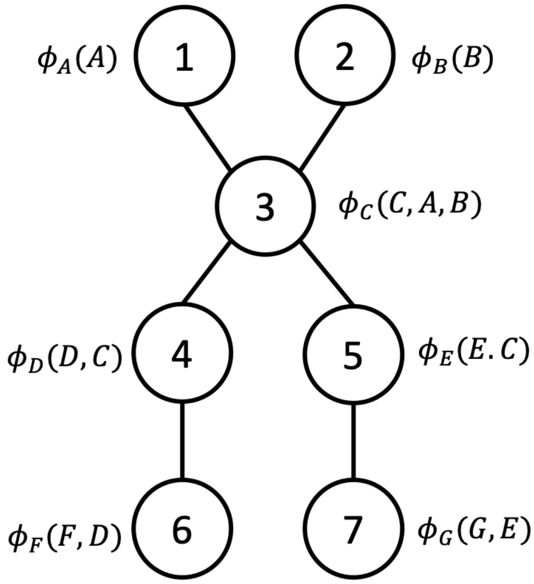
We start building this tree by copying the nodes of the original Bayesian network. It is useful to note the factors associated with each node and their domains. We need to link the nodes avoiding creating large clusters carefully.

In particular, the jointree has the following separators: $(ABC) - BC - (BCE) - BE - (BDE) - DE - (DEF) - EF - (EFG) - FG - (FGH)$. Therefore, indicating an elimination order of $A, C, B, D, E, \{F, G, H\}$. We can use this information to connect the nodes. In particular, we can try to connect nodes that are in the same jointree cluster together, being careful to not create elimination tree clusters larger than 3 variables. The next figure illustrate one of those elimination trees. We note that other elimination trees with the same width are possible.



Question 3

Consider the Bayesian network of Question 1 and the corresponding elimination tree:



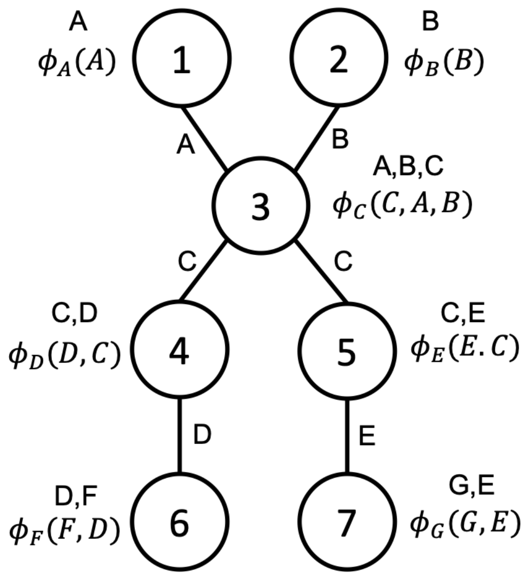
Suppose that the evidence indicator for each variable is assigned to the node corresponding to that variable (e.g. λ_C is assigned to node 3). Suppose we are answering the following queries according to the given order: $P(G = true)$, $P(G, F = true)$, $P(F, A = true, G = true)$ using the following algorithm:

Data: N: Bayesian network
Data: (T, Φ): elimination tree for the CPTs of network N
Data: $\mathbf{e}$: evidence
Result: the joint marginal $P(\mathbf{C}, \mathbf{e})$ for each node i in elimination tree
begin
    for *each variable E in evidence* $\mathbf{e}$ **do**
        $i \leftarrow$ node in the tree T such that $E \in \mathbf{C}_i$;
        $\lambda_E \leftarrow$ evidence indicator for variable E $\{\lambda_E(e) = 1$ if $e \sim \mathbf{e}$ and $\lambda_E(e) = 0$ otherwise $\}$;
        $\phi_i \leftarrow \phi_i \lambda_E$ {entering evidence at node i };
    end
    Choose a root node r in the tree T;
    Pull/collect messages toward root r;
    Push/distribute messages away from root r;
end
return $\phi_i \prod_k M_{ki}$ for each node i in the tree T {joint marginal $P(\mathbf{C}_i, \mathbf{e})$}

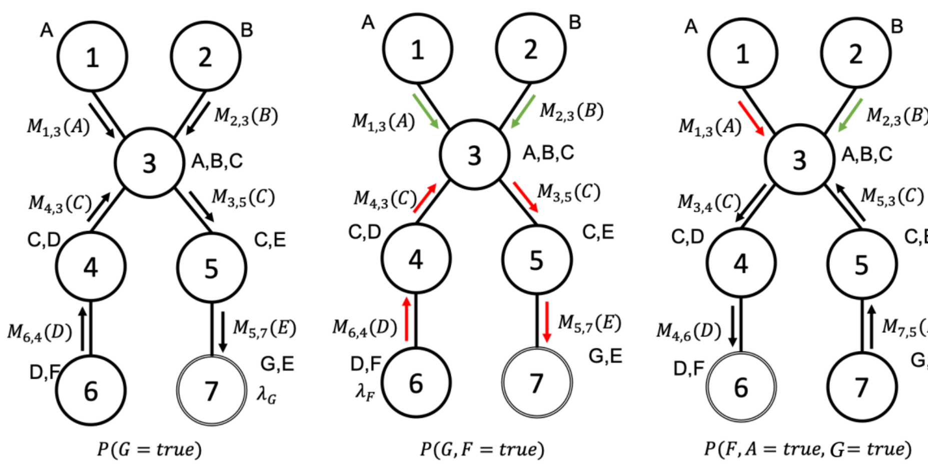
1. Compute the separators, clusters, and width of the given elimination tree.
2. What messages are computed while answering each query according to the previous sequence? State the origin, destination, and value of each message. Use node 7 as the root to answer the first two queries, and node 6 as the root to answer the last query. For each query, compute only the messages directed toward the corresponding root.
3. What messages are invalidated due to new evidence as we attempt each new query?
4. What is the answer to each of the previous queries?

Answer

1. The following figure shows the elimination tree annotated with separators and clusters. The width of this tree is the size of the largest cluster minus one. The largest cluster has three variables. Therefore the width is two.



2. The next figure illustrates the messages for each query. The messages arrows are marked in black when they are computed for the first time. Green when they are reused from a previous query and red when they need to be recomputed.



3. For $P(G, F = true)$ we will need to recompute $M_{6,4}$, $M_{4,3}$, $M_{3,5}$, and $M_{5,7}$. For $P(F, A = true, G = true)$, we will need to recompute $M_{1,3}$.

4. Let's compute the messages.

$M_{6,4}(D) = \sum_F \phi_F(F, D)$

D	$M_{6,4}(D)$
+d	1
-d	1

$M_{4,3}(C) = \sum_D \phi_D(D, C) M_{(6,4)}(D)$

C	$M_{6,4}(C)$
+c	1
-c	1

$M_{(1,3)}(A) = \phi_A(A)$

A	$M_{1,3}(A)$
+a	.6
-a	.4

$M_{2,3}(B) = \phi_B(B)$

B	$M_{2,3}(B)$
+b	.5
-b	.5

$M_{3,5}(C) = \sum_{A,B} \phi_C(C, A, B) M_{1,3}(A) M_{2,3}(B) M_{4,3}(C)$

A	B	C	$\phi_C(C, A, B) M_{1,3}(A) M_{2,3}(B) M_{4,3}(C)$
+a	+b	+c	.27
+a	+b	-c	.03
+a	-b	+c	.03
+a	-b	-c	.27
-a	+b	+c	.1
-a	+b	-c	.1
-a	-b	+c	.06
-a	-b	-c	.14

C	$M_{3,5}(C)$
+c	.46
-c	.54

$M_{5,7}(E) = \sum_C \phi_E(E, C) M_{3,5}(C)$

E	$M_{5,7}(E)$
+e	.154
-e	.846

According to the algorithm, the answer to the query is

$P(G = true) = \phi_G(G, E) M_{5,7}(E) \lambda_G = 0.7076.$

Let's now compute $P(G, F = true)$. We will need to recompute some messages. We will start inserting the evidence in message $M_{6,4}$.

$M_{6,4}(D) = \sum_F \phi_F(F,D)\lambda_F$

D	$M_{6,4}(D)$
+d	.7
-d	.6

$M_{4,3}(C) = \sum_D \phi_D(D,C)M_{6,4}(D)$

C	$M_{4,3}(C)$
+c	.62
-c	.67

$M_{3,5}(C) = \sum_{A,B} \phi_C(C,A,B)M_{1,3}(A)M_{2,3}(B)M_{4,3}(C)$

A	B	C	$\phi_C(C,A,B)M_{1,3}(A)M_{2,3}(B)M_{4,3}(C)$
+a	+b	+c	.1674
+a	+b	-c	.0201
+a	-b	+c	.0186
+a	-b	-c	.1809
-a	+b	+c	.062
-a	+b	-c	.067
-a	-b	+c	.0372
-a	-b	-c	.0938

C	$M_{3,5}(C)$
+c	.2852
-c	.3618

$M_{5,7}(E) = \sum_C \phi_E(E,C)M_{3,5}(C)$

E	$M_{5,7}(E)$
+e	.10088
-e	.54612

According to the algorithm, the answer to the query is
$P(G, F = true) = \sum_E \phi_G(G,E)M_{5,7}(E)$.

G	$P(G, F = true)$
+g	.457072
-g	.189928

Let's now compute $P(F, G = true, A = true)$. We will need to recompute one message and compute some new messages. We will start inserting the evidence in message $M_{1,3}$.

$M_{1,3}(A) = \phi_A(A)\lambda_A$

A	$M_{1,3}(A)$
+a	.6
-a	0

$M_{7,5}(E) = \sum_G \phi_G(G, E)$

E	$M_{7,5}(E)$
+e	1
-e	1

$M_{5,3}(C) = \sum_E \phi_E(E, C)M_{7,5}(E)$

C	$M_{5,3}(E)$
+c	1
-c	1

$M_{3,4}(C) = \sum_{A,B} \phi_C(C, A, B)M_{1,3}(A)M_{2,3}(B)M_{5,3}(C)$

A	B	C	$\phi_C(C, A, B)M_{1,3}(A)M_{2,3}(B)M_{5,3}(C)$
+a	+b	+c	.27
+a	+b	-c	.03
+a	-b	+c	.03
+a	-b	-c	.27
-a	+b	+c	0
-a	+b	-c	0
-a	-b	+c	0
-a	-b	-c	0

C	$M_{3,4}(C)$
+c	.3
-c	.3

$M_{4,6}(D) = \sum_C \phi_D(D, C)M_{3,4}(C)$

D	$M_{4,6}(D)$
+d	.27
-d	.33

According to the algorithm, the answer to the query is $P(F, A = true, G = true) = \sum_D \phi_F(F, D)M_{4,6}(D)$.

F	$P(F, A = true, G = true)$
+f	.387
-f	.213