Computational Intelligence

A Logical Approach

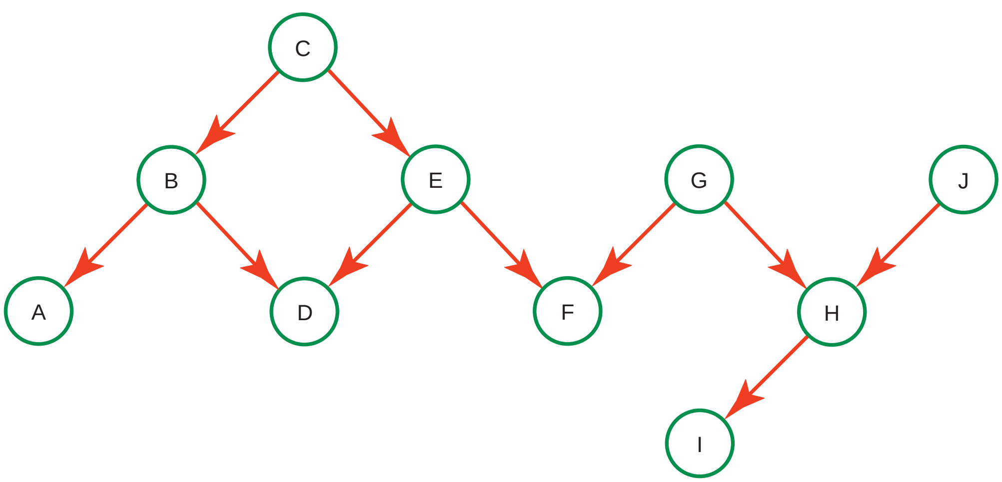
Problems for Chapter 10

Here are some problems to help you understand the material in Computational Intelligence: *A Logical Approach*. They are designed to help students understand the material and practice for exams.

This file is available in html, or in pdf format, either without solutions or with solutions. (The pdf can be read using the free acrobat reader or with recent versions of Ghostscript).

1 Qualitative Effect of Observations in Belief Networks

Consider the following belief network:



We say a variable is independent of another variable in a belief network if it is independent for all probability distributions consistent with the network. In this question you are to consider what variables could have their belief changed as a result of observing a value for variable X, in other words what variables are not independent of X.

(a) Suppose you had observed a value for variable G. What other variables could have their belief changed as a result of this observation?

(b) Suppose you had observed a value for variable I. What other variables could have their belief changed as a result of this observation?

(c) Suppose you had observed a value for variable A. What other variables could have their belief changed as a result of this observation?

(d) Suppose you had observed a value for variable F. What other variables could have their belief changed as a result of this observation?

Solution to part (a)

Suppose you had observed a value for variable G. What other variables could have their belief changed as a result of this observation?

Answer: F, H, and I. As G has no ancestors, only those variables that are descendants of G can be affected by observing G.

Solution to part (b)

Suppose you had observed a value for variable I. What other variables could have their belief changed as a result of this observation?

Answer: H, G, J, and F.

All of the ancestors of I, (namely H, G, and J) and all of their descendants (in this case only F).

Solution to part (c)

Suppose you had observed a value for variable A. What other variables could have their belief changed as a result of this observation?

Answer: B, C, D, E, and F.

All of the ancestors of A, (namely B and C) and all of their descendants (D, E, and F).

Solution to part (d)

Suppose you had observed a value for variable F. What other variables could have their belief changed as a result of this observation?

Answer: A, B, C, D, E, G, H, and I.

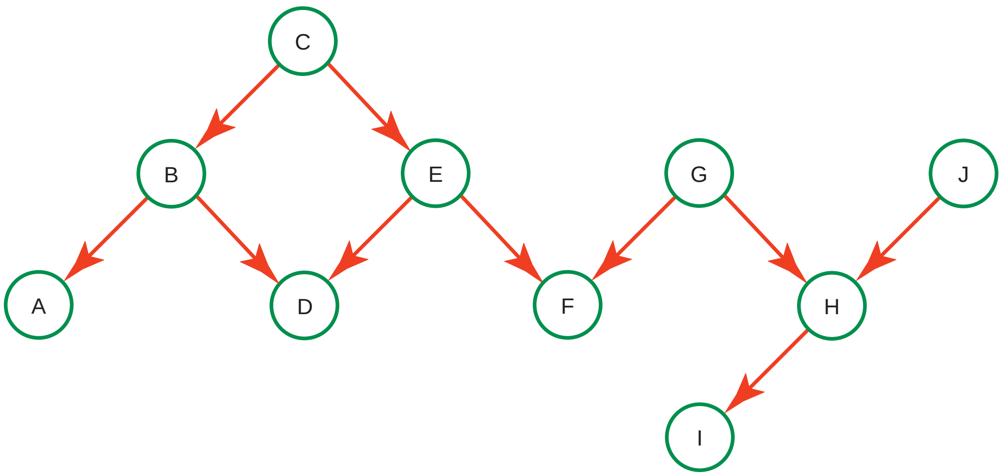
All of the ancestors of F, (namely C, E, and G) and all of their descendants. This means all variables except J can have their belief changes as a result of observing F.

2 Independence Entailed by a Belief Networks

In this question you should try to answer the following questions intuitively without recourse to a formal definition. Think about what information one set of variables could provide us about another set of variables, given that you know about a third set of variables. The purpose of this question is to get you to understand what independencies are entailed by the semantics of belief networks.

This intuition about what variables are independent of other variables is formalized by what is called *d-separation*. You are not expected to know about d-separation to answer the question.

Consider the following belief network:



Suppose X and Y are variables and Z is a set of variables. $I(X, Y|Z)$ means that X is independent of Y given Z for all probability distributions consistent with the above network. For example:

- $I(C, G|\{\})$ is true, as $P(C|G) = P(C)$ by the definition of a belief network.
- $I(C, G|\{F\})$ is false, as knowing something about C could explain why F had its observed value, which in turn would explain away G as a cause for F's observed value. [Remember, you just need to imagine one probability distribution to make the independence assertion false.]
- $I(F, I|\{G\})$ is true because the only way that knowledge of F can affect I is by changing our belief in G, but we are given the value for G.

Answer the following questions about what independencies can be inferred from the above network.

(a) Is $I(A, F|\{\})$ true or false? Explain.

(b) Is $I(A, F|\{C\})$ true or false? Explain.

(c) Is $I(A, F|\{D, C\})$ true or false? Explain.

(d) Is $I(C, F|\{D, E\})$ true or false? Explain.

(e) Is $I(G, J|\{F\})$ true or false? Explain.

(f) Is $I(G, J|\{I\})$ true or false? Explain.

(g) Is $I(F, J|\{I\})$ true or false? Explain.

(h) Is $I(A, J|\{I\})$ true or false? Explain.

(i) Is $I(A, J|\{I, F\})$ true or false? Explain.

Solution to part (a)

Is $I(A, F|\{\})$ true or false? Explain.

It is **false**. Knowing a value for A tells us something about C, which in turn tells us something about F.

Solution to part (b)

Is $I(A, F|\{C\})$ true or false? Explain.

It is **true**. The only way that knowledge of A can affect belief in F is because it provides evidence for C, but C is observed, so A is independent of F given $\{C\}$.

Solution to part (c)

Is $I(A, F|\{D, C\})$ true or false? Explain.

It is **false**. Knowing a value for A could explain away a reason for D being observed, which could change the belief in E, and so in F.

Solution to part (d)

Is $I(C, F|\{D, E\})$ true or false? Explain.

It is **true**. The only way that knowledge of C can affect belief in F is by changing belief in E, but that is given.

Solution to part (e)

Is $I(G, J|\{F\})$ true or false? Explain.

It is **true**. Neither G nor F provide any information about J.

Solution to part (f)

Is $I(G, J|\{I\})$ true or false? Explain.

It is **false**. Knowing G could explain away the observation for a value for I, thus changing the belief in J.

Solution to part (g)

Is $I(F, J|\{I\})$ true or false? Explain.

It is **false**. Knowing J could explain away the observation for a value for I, thus changing the belief in G, which would change the belief in F.

Solution to part (h)

Is $I(A, J|\{I\})$ true or false? Explain.

It is **true**. A only depends on its ancestors, and neither of these are influenced by I or J.

Solution to part (i)

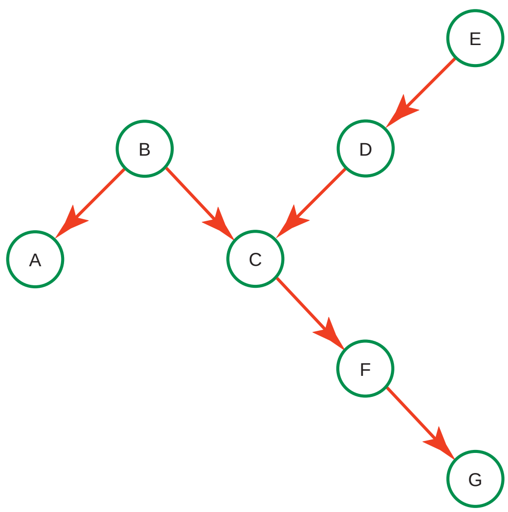
Is $I(A, J|\{I, F\})$ true or false? Explain.

It is **false**. Knowing A would change your belief in C, which would influence G, as it explains away the observation of F. Changing G would change your belief that J was an explanation for I.

3 Variable Elimination Algorithm (Singly Connected)

In this question we trace through one instance of the variable elimination algorithm for a singly connected belief net (i.e., if we ignore the arc directions, there is at most one path between any two nodes).

Consider the following belief network:



Assume that all of the variables are Boolean (i.e., have domain $\{true, false\}$.

We will write variables in upper case, and use lower-case letters for the corresponding propositions. In particular, we will write $A = true$ as a and $A = false$ as $\sim a$, and similarly for the other variables.

Suppose we have the following conditional probability tables:

$P(a|b) = 0.88$

$P(a|\sim b) = 0.38$

$P(b) = 0.7$

$P(c|b \wedge d) = 0.93$

$P(c|b \wedge \sim d) = 0.33$
$P(c|\sim b \wedge d) = 0.53$
$P(c|\sim b \wedge \sim d) = 0.83$
$P(d|e) = 0.04$
$P(d|\sim e) = 0.84$
$P(e) = 0.91$
$P(f|c) = 0.45$
$P(f|\sim c) = 0.85$
$P(g|f) = 0.26$
$P(g|\sim f) = 0.96$

We will draw the factors as tables. The above conditional probability tables are all we need to build the factors. For example, the factor representing $P(E)$ can be written as:

E	Value
true	0.91
false	0.09

The factor for $P(D|E)$ can be written as

E	D	Value
true	true	0.04
true	false	0.96
false	true	0.84
false	false	0.16

and similarly for the other factors.

In this question you are to consider the following elimination steps in order (i.e., assume that the previous eliminations and observations have been carried out). We want to compute $P(A|g)$. (Call your created factors f_1, f_2, etc.)

(a) Suppose we first eliminate the variable E. Which factor(s) are removed, and show the complete table for the factor that is created. Show explicitly what numbers were multiplied and added to get your answer.

(b) Suppose we were to eliminate D. What factor(s) are removed and which factor is created. Give the table for the created factor.

(c) Suppose we were to observe g (i.e., observe $G = true$). What factor(s) are removed and what factor(s) are created?

(d) Suppose we now eliminate F. What factor(s) are removed, and what factor is created?

(e) Suppose we now eliminate C. What factor(s) are removed, and what factor is created?

(f) Suppose we now eliminate B. What factor(s) are removed, and what factor is created?

(g) What is the posterior probability distribution of E? What is the prior probability of the observations?

(h) For each factor created, can you give an interpretation of what the function means?

Solution to part (a)

Suppose we were to eliminate the variable E. Which factors are removed, and show the complete table for the factor that is created. Show explicitly what numbers were multiplied and added to get your answer.

We will eliminate the factors $P(E)$ and $P(D|E)$ and create a factor $f_1(D)$ which can be defined by the table:

D	Value
true	0.91 * 0.04 + 0.09 * 0.84 = 0.112
false	0.91 * 0.96 + 0.09 * 0.16 = 0.888

Solution to part (b)

Suppose we were to eliminate D. What factor(s) are removed and which factor is created. Give the table for the created factor.

We remove the factors containing D, these are $f_1(D)$ and $P(C|B,D)$. We create a new factor $f_2(B,C)$ on the remaining factors.

B	C	Value
true	true	0.93 * 0.112 + 0.33 * 0.888 = 0.3972
true	false	0.07 * 0.112 + 0.67 * 0.888 = 0.6028
false	true	0.53 * 0.112 + 0.83 * 0.888 = 0.7964
false	false	0.47 * 0.112 + 0.17 * 0.888 = 0.2036

Solution to part (c)

Suppose we were to observe g (i.e., observe $G = true$). What factor(s) are removed and what factor(s) are created?

We remove the factor $P(G|F)$, as this is the only factor in which G appears and create the factor $f_3(F)$ defined by:

F	Value
true	0.26
false	0.96

Solution to part (d)

Suppose we now eliminate F. What factor(s) are removed, and what factor is created?

The factors that contain F are $P(F|C)$ and $f_3(F)$. We create the factor $f_4(C)$, defined by:

C	Value
true	0.45 * 0.26 + 0.55 * 0.96 = 0.645
false	0.85 * 0.26 + 0.15 * 0.96 = 0.365

Solution to part (e)

Suppose we now eliminate C. What factor(s) are removed, and what factor is created?

The factors that contain C are $f_2(B, C)$ and $f_4(C)$. We thus create a factor $f_5(B)$, defined by:

B	Value
true	0.3972 * 0.645 + 0.6028 * 0.365 = 0.476216
false	0.7964 * 0.645 + 0.2036 * 0.365 = 0.587992

Solution to part (f)

Suppose we now eliminate B. What factor(s) are removed, and what factor is created?

Variable B appears in three factors: $P(A|B)$, $P(B)$, and $f_5(B)$. We thus create a factor $f_6(A)$, defined by:

A	Value
true	0.88 * 0.7 * 0.476216 + 0.38 * 0.3 * 0.587992 = 0.360380144
false	0.12 * 0.7 * 0.476216 + 0.62 * 0.3 * 0.587992 = 0.149368656

Solution to part (g)

What is the posterior probability distribution of A? What is the prior probability of the observations?

There is only one factor that contains A, this represents $P(A \wedge g)$. We can get the probability of g is $0.360380144 + 0.149368656 = 0.50975$.

We can then calculate the posterior distribution on A (to 3 significant digits):

$$P(a|g) = \frac{0.360380144}{0.50975} = 0.707$$

$$P(\sim a|g) = \frac{0.149368656}{0.50975} = 0.293$$

Solution to part (h)

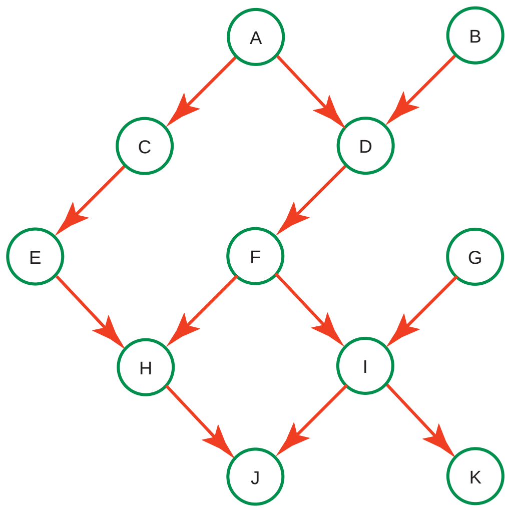
For this example, we can interpret all of the factors as conditional probabilities:

(a) $f_1(D)$ represents $P(D)$.
(b) $f_2(B, C)$ represents $P(C|B)$.
(c) $f_3(F)$ represents $P(g|F)$.
(d) $f_4(C)$ represents $P(g|C)$.
(e) $f_5(B)$ represents $p(g|B)$.
(f) $f_6(A)$ represents $p(g \wedge A)$.

4 Variable Elimination Algorithm (Multiply Connected)

In this question we consider the qualitative aspects of the variable elimination algorithm for a multiply connected network.

Consider the following belief network:



Assume that all of the variables are Boolean (i.e., have domain $\{true, false\}$.

(a) Give all of the initial factors that represent the conditional probability tables.

(b) Suppose we observe a value for K, what factors are removed and what factor(s) are created (call these f_1...).

(c) Suppose (after observing a value for K) we were to eliminate the variables in order: B, D, A, C, E, G, F, I, H. For each step show which factors are removed and what factor(s) are created (call these $f_2, f_3, \ldots$, continuing to count from the previous part). What is the size of the maximum factor created (give both the number of variables and the table size).

(d) Suppose, instead that we were to observe a value for A and a value for I. What are the factors created by the observations? Given the variable ordering: K, B, D, C, E, G, F, H. For each step show which factors are removed and what factor is created.

(e) Suppose, without any observations, we eliminate F. What factors are removed and what factor is created. Give a general rule as to what variables are joined when a variable is eliminated from a Bayesian network.

(f) Suppose we change the graph, so that D is a parent of G, but F isn't a parent of I. Given the variable ordering in part (c) to compute the posterior distribution on J given an observation on K, what are the sizes of the factors created? Can you think of a better ordering?

(g) Draw an undirected graph, with the same nodes as the original belief network, and with an arc between two nodes X and Y if there is a factor that contains both X and Y. [This is called the moral graph of the Bayesian network; can you guess why?]

Draw another undirected graph, with the same nodes where there is an arc between two nodes if there is an original factor or a created factor (given the elimination ordering for part (c) that contains the two nodes. Notice how the second graph triangulates the moral graph.

Draw a third undirected graph where the nodes correspond to the maximal cliques of the second (triangulated) graph. [Note that a clique of a graph G is a set of nodes of G so that there is an arc in G between each pair of nodes in the the clique. We refer to the nodes of the Bayesian network/moral graph as variables.] Draw arcs between the nodes to maintain the properties (i) there is exactly one path between any two nodes and (ii) if a variable is in two nodes, every node on the path between these two nodes contains that variable. The graph you just drew is called a junction tree or a clique tree. On the arc between two cliques, write the variables that are in the intersection of the cliques. What is the relationship between the clique tree and the VE derivation?

Solution to part (a)

Give all of the initial factors that represent the conditional probability tables.

Answer: $P(A)$, $P(B)$, $P(C|A)$, $P(D|A,B)$, $P(E|C)$, $P(F|D)$, $P(G)$, $P(H|E,F)$, $P(I|F,G)$, $P(J|H,I)$, $P(K|I)$.

Solution to part (b)

Suppose we observe a value for K, what factors are removed and what is created.

We remove a factor $P(K|I)$ and replace it with a factor $f_1(I)$.

Solution to part (c)

Suppose (after observing a value for K) we were to eliminate the variables in order: B, D, A, C, E, G, F, I, H. For each step show which factors are removed and what factor is created. What is the size of the maximum factor created (give both the number of variables and the table size).

Step	Eliminate	Removed	Added
1.	B	$P(B), P(D|A,B)$	$f_2(A,D)$
2.	D	$P(F|D), f_2(A,D)$	$f_3(A,F)$
3.	A	$P(C|A), P(A), f_3(A,F)$	$f_4(C,F)$
4.	C	$P(E|C), f_4(C,F)$	$f_5(E,F)$
5.	E	$P(H|E,F), f_5(E,F)$	$f_6(H,F)$
6.	G	$P(G), P(I|F,G)$	$f_7(I,F)$
7.	F	$f_6(H,F), f_7(I,F)$	$f_8(H,I)$
8.	I	$P(J|H,I), f_1(I), f_8(H,I)$	$f_9(H,J)$
9.	H	$f_9(H,J)$	$f_{10}(J)$

The largest factor has two variables, and has table size $2^2 = 4$.

Solution to part (d)

Suppose, instead that we were to observe a value for A and a value for I. What are the factors created by the observations? Given the variable ordering, K, B, D, C, E, G, F, H. For each step show which factors are removed and what factor is created.

Observing A results in replacing the factor $P(A)$ with $f_1()$ (i.e., just a number that isn't a function of any variable. It isn't needed to compute the posterior probability of any variable; it may be useful is we want the prior probability of the observations), replacing the factor $P(C|A)$ with $f_2(C)$, and replacing the factor $P(D|A, B)$ with $f_3(D, B)$.

Observing I results in replacing the factor $P(I|F, G)$ with $f_4(F, G)$, the factor $P(J|H, I)$ with $f_5(J, H)$ and $P(K|I)$ with $f_6(K)$.

Step	Eliminate	Removed	Added
1.	K	$f_6(K)$	f_7
2.	B	$P(B), f_3(D, B)$	$f_8(D)$
3.	D	$P(F|D), f_8(D)$	$f_9(F)$
4.	C	$P(E|C), f_2(C)$	$f_{10}(E)$
5.	E	$P(H|E, F), f_{10}(E)$	$f_{11}(H, F)$
6.	G	$P(G), f_4(F, G)$	$f_{12}(F)$
7.	F	$f_9(F), f_{11}(H, F), f_{12}(F)$	$f_{13}(H)$
8.	H	$f_5(H, J), f_{13}(H)$	$f_{14}(J)$

Note that f_7 is the constant 1 (as it is $P(k|i) + P(\sim k|i)$).

Solution to part (e)

Suppose, without any observations, we eliminate F. What factors are removed and what factor is created. Give a general rule as to what variables are joined when a variable is eliminated from a Bayesian network.

We remove the factors $P(F|D)$, $P(H|E, F)$, $P(I|F, G)$ and replace them with $f_1(D, E, H, I, G)$.

The general rule is that we join the parents of the node being removed, the children of the node being removed and the children's other parents.

Solution to part (f)

Suppose we change the graph, so that D is a parent of G, but F isn't a parent of I. Given the variable ordering in part (c) (i.e., B, D, A, C, E, G, F, I, H) to compute the posterior distribution on J given an observation on K, what are the sizes of the factors created? Is there an ordering that has a smaller factors?

The largest factor has three variables. The factors created have sizes: 2, 3, 3, 3, 3, 3, 2, 2, 1.

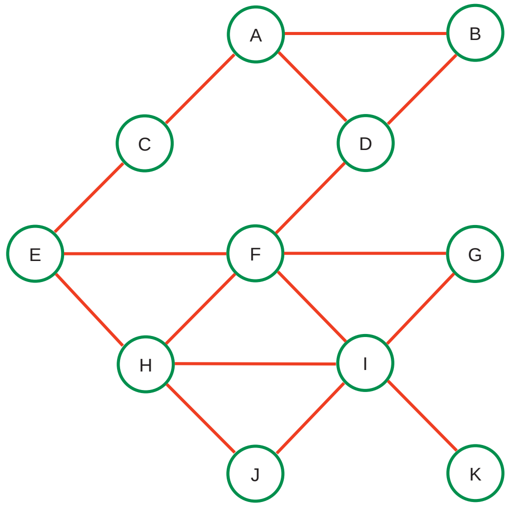
Step	Eliminating	Variables in factor	Number of Variables
1.	B	A, D	2
2.	D	A, F, G	3
3.	A	C, F, G	3
4.	C	E, F, G	3
5.	E	H, F, G	3
6.	G	F, H, I	3
7.	F	H, I	2
8.	I	H, J	2
9.	H	J	1

There is no ordering with a smaller largest factor. But the variable ordering B, A, C, G, F, E, D, I, H results in only one factor of size 3.

Step	Eliminating	Variables in factor	Number of Variables
1.	B	A, D	2
2.	A	C, D	2
3.	C	E, D	2
4.	G	D, I	2
5.	F	E, H, D	3
6.	E	H, D	2
7.	D	H, I	2
8.	I	H, J	2
9.	H	J	1

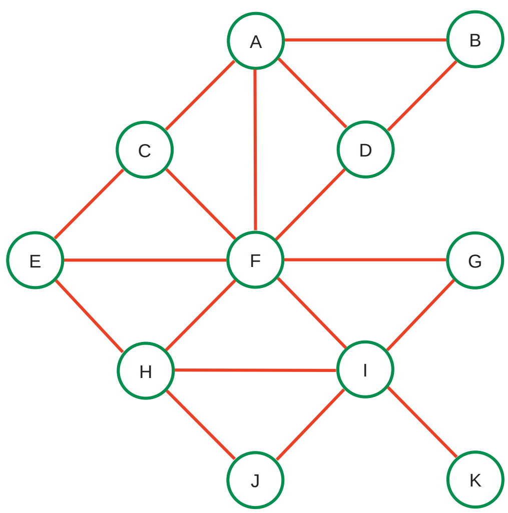
Solution to part (g)

- Draw an undirected graph, with the same nodes as the original belief network, and with an arc between two nodes X and Y if there is a factor that contains both X and Y. [This is called the moral graph of the Bayesian network; can you guess why?]

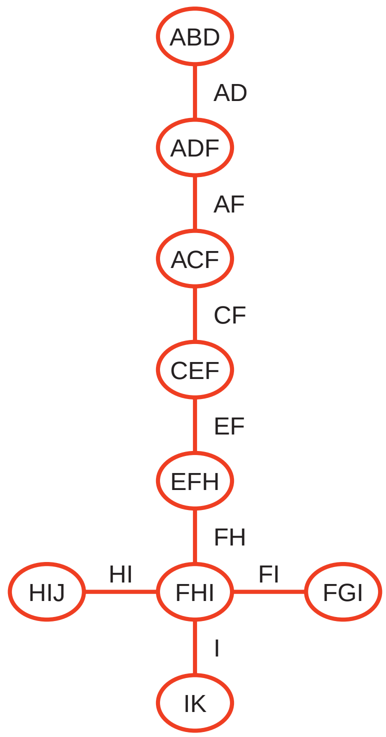


This is sometimes called the moral graph, as we married the parents of each node. Note this is all you ever do to get this graph; draw arcs between all of the parents of each node, and drop the arc directions.

- Draw another undirected graph, with the same nodes where there is an arc between two nodes if there is an original factor or a created factor (given the elimination ordering for part (a)) that contains the two nodes. Notice how the second graph triangulates the moral graph.



- Draw a third undirected graph where the nodes correspond to the maximal cliques of the second (triangulated) graph. [Note that a clique of a graph *G* is a set of nodes of *G* so that there is an arc in *G* between each pair of nodes in the the clique. We refer to the nodes of the Bayesian network/moral graph as variables.] Draw arcs between the nodes to maintain the properties (i) there is exactly one path between any two nodes and (ii) if a variable is in two nodes, every node on the path between these two nodes contains that variable. The graph you just drew is called a junction tree or a clique tree. On the arc between two cliques, write the variables that are in the intersection of the cliques. What is the relationship between the clique tree and the VE derivation?



The variables on the intersections of the cliques correspond to the factors added in the VE algorithm. The last two factors of the VE algorithm are needed to get a factor on J from the HIJ clique.

As an extra exercise, draw the junction tree corresponding to the two variable elimination orderings in the solution to part (f).