

Reasoning MPE to Multiply Connected Belief Networks Using Message Passing

Bon K. Sy

Queens College
of the City University of New York
Department of Computer Science
65-30 Kissena Boulevard
Flushing, NY 11367-0904

Email: bon@qcvax.bitnet bon@cunyvms1.gc.cuny.edu

Abstract

Finding the l Most Probable Explanations (MPE) of a given evidence, S_e , in a Bayesian belief network is a process to identify and order a set of composite hypotheses, H_i s, of which the posterior probabilities are the l largest; i.e., $Pr(H_1|S_e) \geq Pr(H_2|S_e) \geq \dots \geq Pr(H_l|S_e)$. A composite hypothesis is defined as an instantiation of all the non-evidence variables in the network. It could be shown that finding all the probable explanations is a NP-hard problem. Previously, only the first two best explanations (i.e., $l = 2$) in a singly connected Bayesian network could be efficiently derived without restrictions on network topologies and probability distributions. This paper presents an efficient algorithm for finding l (≥ 2) MPE in singly-connected networks and the extension of this algorithm for multiply-connected networks. This algorithm is based on a message passing scheme and has a time complexity $O(lkn)$ for singly-connected networks; where l is the number of MPE to be derived, k the length of the longest path in a network, and n the maximum number of node states — defined as the product of the size of the conditional probability table of a node and the number of the incoming/outgoing arcs of the node.

1. Introduction

A Bayesian belief network [Pearl, 1986] is a graph-theoretic approach for the representation of probabilistic knowledge about the inter-dependencies among a set of random variables. This kind of knowledge representation technique has recently been applied to various problem domains [Charniak & Goldman, 1991] [Dean, 1990][Andreassen *et al*, 1991], and particularly in the domain of diagnosis [Shwe, *et al* 1991]. A problem to be addressed in this paper is related to the generation of the *most probable explanations* to an observation from a Bayesian network. We believe that the solution of this problem will be very useful to the development of probabilistic inference algorithms for solving diagnostic problems which are modeled using Bayesian belief network representation.

Whenever a variable in a Bayesian network is observable, this variable is referred to as an evidence variable. The set of evidence variables is represented by S_e . Given a S_e , an instantiation of all the non-evidence variables — H — in a Bayesian belief network is referred to as a composite hypothesis. Each H is said to be a probable explanation of the given observation, S_e , in a Bayesian belief network if $Pr(H|S_e) > 0$. Finding the l Most Probable Explanations (MPE) of a given evidence, S_e , in a Bayesian belief network is to identify and order a set of composite hypotheses, H_i s, of which the posterior probabilities are the l largest; i.e., $Pr(H_1|S_e) \geq Pr(H_2|S_e) \geq \dots \geq Pr(H_l|S_e)$.

Since the number of composite hypotheses exponentially increases with respect to the number of non-evidence variables, finding all the probable explanations is generally NP-hard. [Santos, 1991] proposed a linear programming approach to solve this problem but the complexity is found to be subexponential. Even if we are interested in only the few most probable composite hypotheses, there is still exponential number of composite hypotheses to consider. It is possible to reduce the search space if certain probability distributions of the variables in a network are assumed. [Cooper, 1984] demonstrated that if $Pr(E_i|S_e) \geq Pr(E_j|S_e)$ holds for all E_i s and E_j s whenever the number of instantiated variables in E_j is larger than that in E_i , then finding the most probable composite hypotheses can be formulated as a search problem, and the best-first search strategy with branch and bound pruning can be applied. Although this approach is efficient and permits reasoning on hypotheses which do not include all the non-evidence variables, [Henrion, 1990] pointed out that the assumption is too strong and such an assumption is hardly valid in any real world problem. Another approach being explored by [Shimony & Charniak, 1990] is to transform a Bayesian network into a Weighted Boolean Function Directed Acyclic Graph (WBFDAG) which permits the application of the best-first search strategy. The major distinction of this approach in comparison to Cooper's approach is that no

assumption is made about the probability distributions. However, the spatial complexity of a WBF DAG is in an exponential order of the original DAG; where the spatial complexity is defined in terms of the size of a network. Although the time complexity is shown to be linear with respect to the size of a graph, the exponential complexity problem remains. Another approach being taken is to take advantage of the structure of a network. [Pearl, 1988] has shown a message passing algorithm which can efficiently derive two most probable composite hypotheses in a singly-connected network. In a singly-connected network, any pair of nodes is connected by at most one unique path. Unfortunately, Pearl's algorithm has two limitations. First, it cannot be applied to multiply-connected networks (i.e., non-singly connected networks). Second, Pearl's message passing scheme cannot be extended to finding more than the first two most probable composite hypotheses.

The objective of this paper is to introduce a message passing scheme for the derivation of the most probable composite hypotheses. The mechanism of our message passing process, in essence, is similar to Pearl's algorithm [Pearl, 1988]. However, our method differs from Pearl's and others' algorithms in four ways. First, the message passing in our method is unidirectional as opposed to bidirectional in Pearl's algorithm. Second, each "message stream" in our method is a vector but not a value as in [Pearl, 1988]. Third, we retain all processed information to permit their reuse in a systematically ordered fashion for the successive derivation of the most probable composite hypotheses. Finally, our message passing scheme can be applied to a tree-type hypergraph which is critical in dealing with multiply-connected networks.

In section 2 we will discuss the formalism of Bayesian belief networks and the realization of our message passing scheme as an unidirectional path traversal in a hierarchically organized graph. In section 3 we will detail the mathematical formulation of our message passing scheme. The treatment of multiply-connected networks will be discussed in section 4. In section 5 we will conclude this paper with a discussion of the relationship of this research with others in the field and future research work.

2. Formalism of Bayesian belief networks

A Bayesian network [Pearl, 1986] is a Directed Acyclic Graph (DAG) within which a set of nodes are connected by a set of arcs. Each node in a graph represents a random variable, and an arc connecting two nodes indicates the dependency between them. In particular, *head-to-head*, *head-to-tail*, and *tail-to-tail* are three configurations to specify the marginal and condi-

tional independencies among 3 adjacent nodes. These three configurations, together with the definitions of "joint" and "separate" discussed in [Pearl, 1986], permit the joint distribution of a Bayesian belief network shown in Fig. 1 to be re-written as below:

$$\begin{aligned} &Pr(abcdefghij) \\ &= Pr(a)Pr(b|a)Pr(c|b)Pr(d)Pr(e|cd)Pr(f|b) \\ &\quad Pr(g|f)Pr(h|g)Pr(i|fj)Pr(j) \end{aligned} \quad (1)$$

Note that the Right Hand Side (RHS) expression of (1) is in a form of $\prod_{x_i \in N} Pr(x_i|J_{x_i})$; where N is the set of nodes in a Bayesian belief network, and J_{x_i} is the set of immediate parents of x_i . This expression is a simplified form of the Bayes expression obtained from the marginal and conditional independency characteristics of the distribution of the network. The realization of the RHS expression of equation (1) signifies two important characteristics. First, the spatial complexity of the RHS expression is in the linear order of the sum of the number of states in each probability term, as opposed to the exponential order (with respect to the number of variables) as the expression in the LHS. Second, each probability term in the RHS, $Pr(x_i|J_{x_i})$, can be conceptualized as a local probability term associated with the node x_i . Each of these local probability terms constitutes the basis of the information to be passed to its neighbors in our proposed message passing scheme.

Let's first consider the singly-connected network shown in Fig. 1. The idea of our message passing scheme is to propagate the minimum amount of infor-

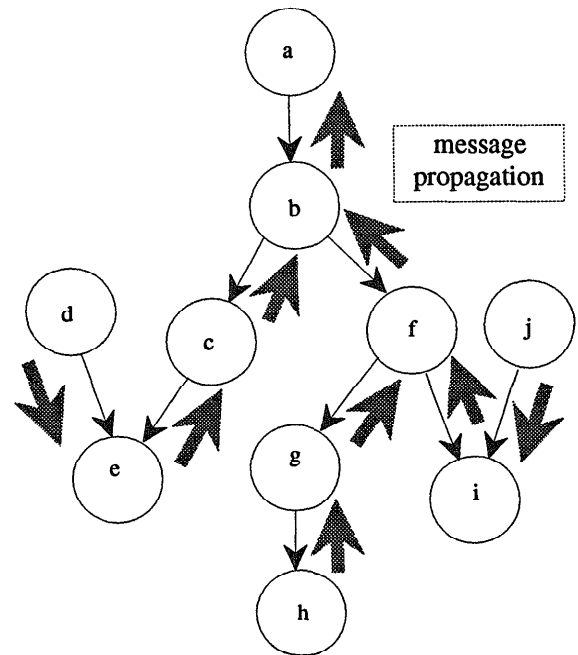


Fig. 1: A ten-node singly connected Bayesian network.

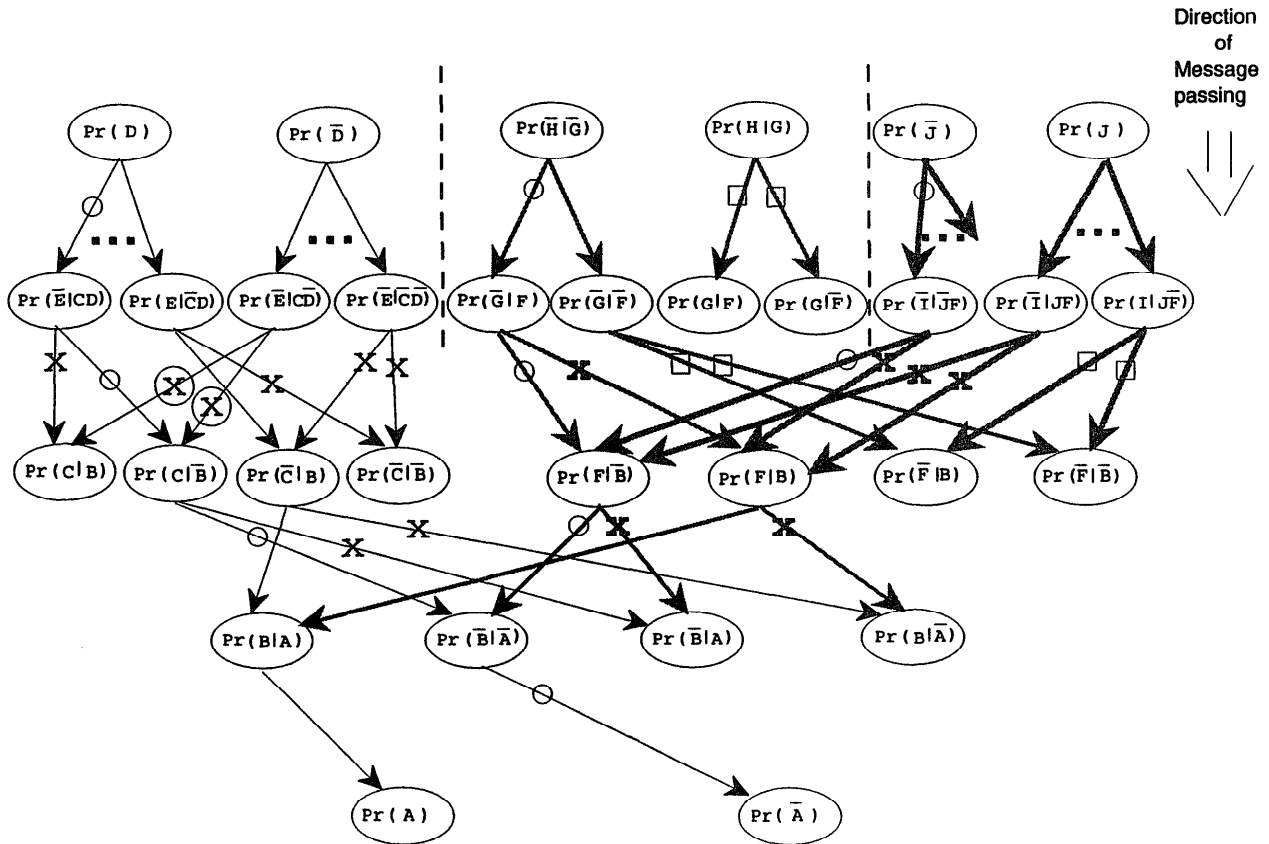


Fig. 2 Message Flow Graph

mation needed towards a designated node via a set of unique paths which cover all the nodes in a network. For example, suppose node a is selected as a designated node (i.e., a sink) to absorb the incoming information from every other node in the network. Then three unique paths can be identified; namely, $d \rightarrow e \rightarrow c \rightarrow b \rightarrow a$, $h \rightarrow g \rightarrow f \rightarrow b \rightarrow a$, and $j \rightarrow i \rightarrow f \rightarrow b \rightarrow a$.

For the sake of discussion, each variable in Fig. 1 is represented by a lower case letter and is assumed to be binary-valued — true or false. An upper case letter represents the value of a variable. For example, X and \bar{X} represent $x = X$ (i.e., true) and $x = \bar{X}$ (i.e., false) respectively. In addition, each node in the three unique paths discussed previously can be replaced by the probability terms of all its possible instantiations, and the three unique paths can now be represented in terms of the probability terms¹ of the nodes as shown in Fig. 2. The purpose of conceptualizing these three unique paths as a graph is for the realization of our proposed message passing scheme to be discussed in

the next section.

3. Message passing using local propagation

To illustrate the local computation involved in our proposed message passing scheme, let's suppose we are interested in finding the most probable composite hypotheses without evidence ($S_e = \emptyset$) in Fig. 1; i.e., $\text{ArgMax}[Pr(abcdefghij)]$. Finding the most probable composite hypothesis is equivalent to finding the optimal setting for each of the local probability terms in the RHS of equation (1). Since Fig. 1 is a singly-connected network, any two non-adjacent nodes are conditionally independent given that at least one node in the unique path connecting the two nodes is (un)instantiated to separate them. This conditional independence property implies that the identification of the optimal setting of a particular local probability term depends on only the *ascendant* terms. For example, in order to determine the optimal setting for the term $Pr(c|b)$, it depends only on the terms $Pr(e|cd)$ and $Pr(d)$, but not $Pr(b|a)$, $Pr(a)$ (the descendant terms in Fig. 2), or $Pr(h|g)$, $Pr(g|f)$, $Pr(f|b)$, $Pr(j)$, $Pr(i|jf)$ (the out-of-branch terms in Fig. 2). This permits the

¹ Only consistent terms are connected.

optimal setting of each local probability term to be searched locally in an unidirectional downward propagation (in Fig. 2). An important question to ask now is the kind of information which should be carried in the message passing process. One of the considerations is to anticipate the possible settings of the immediate descendant terms in Fig. 2. For example, the information that should go from $Pr(h|g)$ to $Pr(g|f)$ would be all the possible settings of g . That is, the information to be passed from node h to g would be $M_{h \rightarrow g} = [Max_h[Pr(h|G)], Max_h[Pr(h|\bar{G})]]$. The next message propagation from node g to f , however, should consider both $[Max_g[Pr(g|F)], Max_g[Pr(g|\bar{F})]]$ as well as $M_{h \rightarrow g}$. In order to integrate the information properly, a **convolution** operation, $*$, and a **Belief** matrix are defined for this purpose.

Definition1: Given $M_{d \rightarrow x} = [m_{d \rightarrow X_1} \ m_{d \rightarrow X_2} \ \dots \ m_{d \rightarrow X_n}]$, and $Pr(x|J_x) = [Pr(X_1|v1_x) \ Pr(X_1|v2_x) \ \dots \ Pr(X_n|vk_x)]$ (where vi_x is an instantiation of the variables in J_x), the convolution of $M_{d \rightarrow x}$ with $Pr(x|J_x)$ is defined as the product of every single term in $Pr(x|J_x)$ with a consistent $m_{d \rightarrow x}$ in $M_{d \rightarrow x}$; where J_x is the set of immediate parent nodes of x in a Bayesian belief network. $Pr(x|J_x)$ and $m_{d \rightarrow x}$ are consistent with each other if the instantiation of x in $m_{d \rightarrow x}$ and that in $Pr(x|J_x)$ are identical.

Definition2: A Belief Matrix of a node x , $Bel(x)$, is defined as the convolution of all $M_{d \rightarrow x}$ with $Pr(x|J_x)$ — $M_{d_1 \rightarrow x} * M_{d_2 \rightarrow x} * \dots * M_{d_k \rightarrow x} * Pr(x|J_x)$; where d_i are the nodes which propagate $M_{d_i \rightarrow x}$ to x for $i = 1 \dots k$.

To illustrate the definitions of convolution and belief matrix, let's suppose

$$M_{h \rightarrow g} = [m_{h \rightarrow G} \ m_{h \rightarrow \bar{G}}] = [(HG \ 0.6) \ (H\bar{G} \ 0.8)] \text{ and}$$

$$Pr(g|f) = \begin{bmatrix} (GF \ 0.3) & (G\bar{F} \ 0.45) \\ (\bar{G}F \ 0.7) & (\bar{G}\bar{F} \ 0.55) \end{bmatrix}$$

then we have

$$Bel(g) =$$

$$M_{h \rightarrow g} * Pr(g|f) = \begin{bmatrix} (HGF \ 0.18) & (HGF\bar{F} \ 0.27) \\ (H\bar{G}F \ 0.56) & (H\bar{G}\bar{F} \ 0.44) \end{bmatrix}$$

Remark: $Pr(h|g)$ in $M_{h \rightarrow g}$ is represented by a 2-tuple. The first tuple is the settings of h and g , and the second tuple is $Pr(h|g)$. For example, $(HG \ 0.6)$ in $M_{h \rightarrow g}$ is equivalent to $Pr(H|G) = 0.6$.

With these two definitions, $M_{g \rightarrow f}$ can be formulated as $Max_g[Bel(g|F) \ Bel(g|\bar{F})]$; i.e., $M_{g \rightarrow f} = [(H\bar{G}F \ 0.56) \ (H\bar{G}\bar{F} \ 0.44)]$. Lemma1 summarizes the formulation of a message stream $M_{b \rightarrow a}$:

Lemma1: A message stream that a node b propagates to a node a in a Bayesian belief network (in Fig. 1 but

not Fig. 2) is defined as

$$M_{b \rightarrow a} = \begin{cases} Max_b[Bel(b|A) \ Bel(b|\bar{A})] & \text{if } a \text{ is an immediate parent of } b; \\ Max[Bel(B) \ Bel(\bar{B})] & \text{if } b \text{ is an immediate parent of } a; \end{cases}$$

where

$$Bel(b) = \begin{cases} M_{d_1 \rightarrow b} * \dots * M_{d_k \rightarrow b} * Pr(b|ap_1 \dots p_k) & \text{if } a \text{ is an immediate parent of } b; \\ M_{p_1 \rightarrow b} * \dots * M_{p_k \rightarrow b} * Pr(b|p_1 \dots p_k) & \text{if } b \text{ is an immediate parent of } a; \end{cases}$$

$d_1 \dots d_k$ = are the immediate descendent nodes of b , and $p_1 \dots p_l$ are the immediate ascendent nodes of b .

Remark: If node b is a root node, $M_{b \rightarrow a}$ is simply $[(B \ Pr(B)) \ (\bar{B} \ Pr(\bar{B}))]$.

It is possible that the instantiation of a variable for the maximal value of $Bel(\bullet)$ is not unique. In this case, all such instantiations must be included in a message vector in order to find all the MPEs. However, if we are interested in only one of the MPEs, then we can break tie arbitrary. Due to the unidirectional message propagation, the arbitrarily selected instantiation is guaranteed to be one of the solutions. This eliminates the need of carrying all instantiations via explicit pointers, which is required in the Pearl's bidirectional message propagation [Pearl, 1988].

Note that $M_{b \rightarrow a}$ described in Lemma1 is a pruning process illustrated in Fig. 2 that, at each level, all the links, except one, in each group of the variable instantiations are pruned (marked by "x"). For example, Fig. 2 illustrates part of the pruning process². If we look at the message passing from node e to c , only one link from e to c remains for each possible instantiation of c ; i.e., from $Pr(e = \bar{E}|C, d = D) \rightarrow Pr(C|b = \bar{B})$, and from $Pr(e = E|\bar{C}, d = D) \rightarrow Pr(\bar{C}|b = B)$. Similarly, only the links from $Pr(c = C|\bar{B}) \rightarrow Pr(\bar{B}|a = \bar{A})$ and $Pr(c = \bar{C}|B) \rightarrow Pr(B|a = A)$ remain unpruned. When the information from each node reaches the designated sink (node a), the most probable composite hypotheses can be realized from $ArgMax_a[Bel(A) \ Bel(\bar{A})]$.

Once the most probable composite hypothesis is found, let's say, $\bar{A}\bar{B}CDE\bar{F}\bar{G}\bar{H}\bar{I}\bar{J}$, the path corresponding to this instantiation in Fig. 2 is marked (those marked with "o"). To find the second most probable composite hypothesis, the candidate must be from one

² The actual spatial complexity of Fig. 2 is only proportional to the total number of unpruned links because only unpruned links are stored.

of the unpruned links, or *undeleting* some of the previously pruned links in which variable instantiation is exhausted. For example, once the link from $Pr(\bar{E}|CD) \rightarrow Pr(C|\bar{B})$ (in Fig. 2) is marked after the most probable composite hypothesis is identified, the information to be passed from node e to c with node c being instantiated to C is exhausted. In this case, the links, $Pr(e|Cd) \rightarrow Pr(C|b)$ (i.e., $M_{e \rightarrow c}$), are *undeleted* (indicated by “ \otimes ”); e.g., $Pr(\bar{E}|C\bar{D}) \rightarrow Pr(C|B)$ and $Pr(\bar{E}|C\bar{D}) \rightarrow Pr(C|\bar{B})$ are undeleted. These two pieces of new information are used in $Pr(e|Cd) * Pr(C|b)$ to update the belief matrix $Bel(c)$, and the second largest of $Bel(C)$ is added to the previous message stream to be propagated for the next local computation. Such an updating is then repeated at each level, and at each iteration for identifying the next most probable composite hypothesis.

In essence, we need to keep track at most n terms for each instantiation x_i of a given variable, X , in computing the n most probable composite hypotheses. When an instantiation x_i of a variable X and some instantiation y_j of a variable Y in $M_{X \rightarrow Y}$ are identified as the settings for the most probable interpretation in, let's say, k -th iteration, then the k -largest term with the instantiation of the variable X to be x_i (and some instantiation of Y) is the only additional information to be included in the k -th iteration. Therefore, the size of the message vector will grow incrementally as the number of iteration increases. Such a linear increment is the worst case in which the size of a message vector will grow per iteration. In the best case, the size of a message vector will remain the same as the one in the previous iteration. Due to the page limit, the readers are referred to [Sy, 1992a] for the detailed discussion. We have also proved in [Sy, 1992a] that the information propagated at each level and at each iteration according to the process just described is complete and is the minimal amount which is needed in identifying the next most probable composite hypothesis. The following is the algorithm of this message passing scheme:

Step 1: Define $l \leftarrow$ length of partial ordering (i.e., number of most probable composite hypotheses to be sought), and $S_e \leftarrow$ the evidence.

Step 2: Designate a node as a sink, identify the paths for the propagation of message streams, and construct the corresponding flow graph using the (instantiated) local probability terms associated with each variable. Initialize the iteration count, $i = 1$.

Step 3: Compose $M_{b \rightarrow a|S_e}$ using Lemma1, and propagate messages along the proper paths identified in step 2.

Step 4: Perform convolution operation to integrate incoming messages and update Belief matrix, $Bel(x)$,

at each node traversal.

Step 5: Identify the setting of the composite hypothesis with the i -largest $Pr(H_i|S_e)$ in the designated sink node at the i th iteration.

Step 6: Mark the path which corresponds to the most probable composite hypothesis just found.

Step 7: Undelete the previously pruned paths in which variable instantiations are exhausted.

Step 8: Repeat steps 3 to 8 until i reaches l .

Noted that the algorithm shown above is based on the propagation of quantitative vector streams towards a designated sink node in a network. In a complete iteration of propagating the vector streams to the sink node, one composite hypothesis of the ordering can be identified. To obtain the l most probable composite hypotheses, l iterations will be needed.

When parallel processing is permitted, the amount of time required for each iteration will be at most the amount of time required for the convolution operations in the longest path (i.e., length k stated in the theorem). Note that the node states, n — defined as the product of the size of the conditional probability table of a node and the number of incoming/outgoing (depends on the direction of message flow) arcs of the node in a Bayesian belief network (Fig. 1) — is the worst case of the time complexity of one convolution operation. The time complexity for one iteration is $O(kn)$, and for l iterations, the time complexity is $O(lkn)$. The details of a formal proof of this complexity order and the completeness of the message passing scheme are referred to [Sy, 1992a].

5. Coping multiply-connected network

The message passing scheme discussed and illustrated in the previous sections fails to produce correct inference when the Bayesian belief network is multiply connected. One of the main reasons is that a common parent node or a common daughter node (such as nodes a and d in Fig. 3 respectively) may receive conflicting messages along the paths of propagation. For example, suppose the propagation paths relevant to nodes a , b , c , and d are $d \rightarrow b \rightarrow a$ and $d \rightarrow c \rightarrow a$ in Fig. 3. Node a may receive different values of b and c when the maximum of $Pr(b|a)Pr(d|bc)$ and $Pr(c|a)Pr(d|bc)$ are considered via different paths.

In order to apply the message passing algorithm to a multiply connected network, the knots in a multiply connected network must be broken up; for example, $a-b-d-c-a$ in Fig. 3. Clustering and cutset conditioning are two techniques [Pearl, 1988][Neapolitan, 1990][Peot & Shachter, 1991] for resolving these knots. The basic idea of clustering is to lump variables together to form compound variables as a method of eliminating knots.

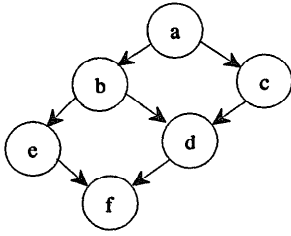


Fig. 3. Multiply Connected Network

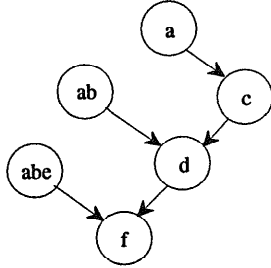


Fig. 4. Singly Connected Compound-Variable Network

For example, nodes b and c in Fig. 3 can be lumped together as a compound variable bc . However, a disadvantage of this approach is that the belief relevant to a compound variable explains less in terms of the belief accrued by each individual variable in a compound variable. The other technique — cutset conditioning — is based on the idea that a parent node can be *absorbed* into one or more of its immediate descendant nodes for the purpose of breaking up knots. For example, node a in Fig. 3 can be absorbed into node b to obtain a singly connected network once node a is instantiated to have a fixed value. This approach, however, renders an expensive computation because all possible values of the absorbed nodes (node a in this case) have to be considered in the process of local computation.

The technique that we employ to resolve knots is a combination of clustering and cutset conditioning. There are two major steps in our proposed technique in breaking up knots. The first step is the formulation of a singly connected version of the multiply connected network. The second step is the construction of the density function. Whenever a parent node, x , is *absorbed* by its immediate descendant nodes, this parent node and the compound variable are marked as special nodes where all possible values of x must be propagated in the message passing process. For example, the multiply connected network shown in Fig. 3 can be formulated to have a singly connected version (Fig. 4) after node a is absorbed by node b to form a compound-variable node ab , and node ab is absorbed by node e to form abe . These compound-variable nodes are marked as special nodes and all the possible values of a , ab , and abe must be included as the messages during the message prop-

agation. In order to illustrate the necessity of this, let's suppose the message propagation paths are from $a \rightarrow c \rightarrow d \rightarrow ab$ and $abe \rightarrow e \rightarrow d \rightarrow ab$. Since nodes a , ab , and abe will all influence the setting of a in finding the most probable composite hypotheses, all possible values of a and b must be carried along the message propagation. This step has the same effect as creating multiple copies of the absorbed nodes and keeping track of their values to ensure consistency. Before we can apply our proposed message passing scheme, the density function of the network in Fig. 4 can be realized as $P'(a)P'(c|a)P'(d|ab, c)P'(ab)P'(abe)P'(f|abe, d)$ which has a form $\prod_x P'(x|J_x)$. Since the product of all $P'(\bullet)$ s must equal to the product of all $Pr(\bullet)$ s, each $P'(\bullet)$ can be realized as the product of the relevant $Pr(\bullet)$ terms with proper scalings. Each term, $Pr(\bullet)$, is scaled to be $Pr(\bullet)^{\frac{1}{n}}$; where n is the number of occurrences of $Pr(\bullet)$ in the singly connected version of the network⁴. For example,

$$\begin{aligned} P'(a) &= Pr(a)^{\frac{1}{3}} & P'(abe) &= Pr(e|b)Pr(b|a)^{\frac{1}{3}}Pr(a)^{\frac{1}{3}} \\ P'(ab) &= Pr(b|a)^{\frac{1}{3}}Pr(a)^{\frac{1}{3}} & P'(c|a) &= Pr(c|a) \\ P'(d|ab, c) &= Pr(d|bc)Pr(a)^{\frac{1}{3}} \\ P'(f|abe, d) &= Pr(f|de)Pr(b|a)^{\frac{1}{3}}Pr(a)^{\frac{1}{3}} \end{aligned}$$

There are two important points about the treatment of multiply connected networks. First, the terms $P'(\bullet)$ s are not necessarily probability functions. Nevertheless, the product of all $P'(\bullet)$ s yields the same values of the probability density function of the network. This property is similar to the concept of potential function introduced in [Lauritzen & Spiegelhalter, 1988] for dealing with multiply connected networks through triangulation techniques [Kjaerulff, 1990]. Second, if a $P'(x|y)$ is related to only one $Pr(x|y)$ term, the ordering of $P'(x|y)$, with respect to different instantiation of x and y , is the same as that of $Pr(x|y)$ even though $Pr(x|y)$ is scaled. If a $P'(\bullet)$ is related to several $Pr(\bullet)$ terms, the influence of non-immediate parent or daughter nodes are brought to the scope in the case of compound-variable node such as $P'(d|ab, c)$ shown in Fig. 4. These two properties ensure the sufficiency of the information required for the message passing and the correctness of the belief computation.

6. Conclusion

In this paper we have presented a message passing algorithm for the derivation of the first l most probable explanations. If a network is singly connected, the first l most probable explanations can be found in order of $O(kln)$; where l is the number of most probable

⁴ The purpose of scaling $Pr(\bullet)$ is to reduce the problem arising from the non-unique instantiations of variables which share the same maximal Bel value.

explanations to be derived, k the length of the longest path in a network, and n the maximum number of node states. We believe that this result is important in the domains such as diagnosis because previously we were able to only efficiently derive the first two ($l = 2$) most probable explanations.

If a network is multiply-connected, the computational complexity is generally NP-hard and there is no algorithm which could efficiently derive the l most probable explanations. In this paper we have introduced a technique based on the idea of clustering and cutset conditioning to obtain a singly connected version of a multiply connected network in which our message passing scheme can be applied to identify the l MPE. Although the computational complexity is still NP-hard, we argue that our message passing algorithm only deals with exponential complexity proportional to the maximum number of node states of the compound variables.

Finally, we need to point out that an explanation is viewed as a composite hypothesis in this paper; where a composite hypothesis is defined as an instantiation of all non-evidence variables. Apparently an explanation can be viewed as an instantiation of any subset of non-evidence variables. This raises the issue of what it is meant by an explanation, which was discussed to some extent in [Peng & Reggia, 1987]. Unfortunately finding the l MPE of this kind would require the sum of those variables that are neither in the explanation nor in the evidence. This results in an exponential number of terms to be dealt with in such a summation. A challenging problem, as a follow-up to this research, is the development of an efficient algorithm for finding the l MPE of this kind and the preliminary results can be found in [Sy, 1992b].

Acknowledgements

This work is supported in part by the C.U.N.Y. PSC-CUNY Research Award Program, and a grant to Queens College from the General Research Branch, N.I.H. under grant No. RR-07064. I also thank the anonymous reviewers for their valuable comments.

References

- [Andreassen S., Woldbye M., Falck B., and Andersen S., 1991] MUNIN – A Causal Probabilistic Network for Interpretation of Electromyographic Findings, *Proc. of the 9th National Conf. on Artificial Intelligence*, pp. 121-123, Menlo Park, California: AAAI.
- [Charniak E., and Goldman R., 1991] A Probabilistic Model of Plan Recognition, *Proc. the 9th National Conf. on Artificial Intelligence*, pp. 160-165, Menlo Park, California: AAAI.
- [Cooper G., 1984] NESTOR: A Computer-Based Medical Diagnosis that Integrates Causal and Probabilistic Knowledge, *Technical Report HPP-84-48*, Stanford University, Stanford, CA.
- [Dean T., 1990] Coping with Uncertainty in a Control System for Navigation and Exploration, *Proc. of the 8th National Conf. on Artificial Intelligence*, pp. 1010-1015, Boston, MA: AAAI.
- [Henrion M., 1990] Towards Efficient Probabilistic Diagnosis in Multiply Connected Belief Networks, in R.M. Oliver and J.Q. Smith eds, *Influence Diagrams, Belief Nets and Decision Analysis*, pp. 385-409, John Wiley & Sons Ltd., New York.
- [Lauritzen S.L., Spiegelhalter D.J., 1988] "Local computations with Probabilities on Graphical Structures and their application to Expert System," *Journal of Royal Statistics Society — British*, 50(2), 157-224.
- [Kjaerulff U., 1990] "Triangulation of Graphs — Algorithms Giving Small Total State Space," *Internal Tech. Report, R-90-09*, University of Aalborg, Denmark.
- [Neapolitan R., 1990] *Probabilistic Reasoning in Expert Systems*, John Wiley & Sons, New York.
- [Pearl J., 1986] "Fusion, Propagation, and Structuring in Bayesian Networks," *Artificial Intelligence*, Vol. 28, pp. 241-288.
- [Pearl J., 1988] *Probabilistic reasoning in Intelligent Systems* Morgan Kaufmann Publishers.
- [Peng Y. and Reggia J.A., 1987] A Probabilistic Causal Model for Diagnostic Problem Solving (Part I & II), *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 17, (No. 2 & 3).
- [Peot M., Shachter R.D., 1991] Fusion and Propagation with Multiple Observations in Belief Networks, *Artificial Intelligence*, Vol. 48, pp. 299-318.
- [Santos E., 1991] On the Generation of Alternative Explanations with Implications for Belief Revision, *Proc. of the Conf. on Uncertainty in A.I.*, pp. 339-347.
- [Shimony S.E. and Charniak E., 1990] A New Algorithm for Finding MAP Assignments to Belief Networks, *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, Cambridge, MA, pp. 98-103.
- [Shwe M., Middleton B., Heckerman D., Henrion M., Horvitz E., and Lehmann H., 1991] Probabilistic Diagnosis Using a Reformulation of the INTERNIST-1/QMR Knowledge Base, *Methods of Information in Medicine*, 30:241-255.
- [Sy B.K., 1992a] A Recurrence Local Computation Approach Towards Ordering Composite Beliefs, to appear in the *International Journal of Approximate Reasoning, North-Holland*.
- [Sy B.K., 1992b] An Adaptive Reasoning Approach Towards Efficient Ordering of Composite Hypotheses, submitted to the *Annals of Mathematics and Artificial Intelligence*.