Pearl's Belief Propagation: the Proofs

Linda van der Gaag

RUU-CS-92-47
December 1992

Utrecht University
Department of Computer Science
Padualaan 14, P.O. Box 80.089,
3508 TB Utrecht, The Netherlands,
Tel. : ... + 31 - 30 - 531454

Pearl's Belief Propagation: the Proofs

Linda van der Gaag

Technical Report RUU-CS-92-47
December 1992

Department of Computer Science
Utrecht University
P.O.Box 80.089
3508 TB Utrecht
The Netherlands

ISSN: 0924-3275

Pearl's Belief Propagation: the Proofs

Linda van der Gaag
Utrecht University, Department of Computer Science
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands

Abstract

The belief network framework for plausible reasoning provides both a formalism for representing knowledge concerning a joint probability distribution on a problem domain, and a set of algorithms for efficiently propagating evidence and computing probabilities of interest. This paper presents a rigorous review of the set of algorithms for belief propagation proposed by J. Pearl, including full proofs.

1 Introduction

Halfway through the 1980s, the theory of *belief networks* was introduced for reasoning with uncertainty in knowledge-based systems. The belief network framework provides a formalism for representing knowledge about a problem domain, or to be more precise, for representing knowledge concerning a joint probability distribution on a set of variables discerned in the domain. Beside this knowledge-representation formalism the framework also provides a set of algorithms for reasoning with knowledge represented in the formalism. Several such sets of algorithms have been proposed. The earliest and most well-known is the set of algorithms proposed by J. Pearl from the University of California, Los Angeles, [Pearl, 1988]. Since its introduction the belief network framework is becoming increasingly popular, and at present more and more knowledge-based systems are being developed using this framework, most notably in the area of medical diagnosis and therapy selection, see for example [Andreassen et al., 1987], [Andreassen et al., 1991], [Bellazzi et al., 1991] and [Shwe et al., 1991]; for applications in other domains, see for example [Bruza & van der Gaag, 1992] and [Jensen et al., 1990].

The aim of this paper is to present a formal review of the set of algorithms proposed by Pearl; for an informal introduction to these algorithms, the reader is referred to [Morawski, 1989]. The paper is organized as follows. In Section 2 the basic notions involved in the belief network framework are provided. Section 3 defines the belief network formalism and its semantics. Section 4 reviews Pearl's algorithms and presents full proofs of their correctness. The paper is rounded off with some concluding observations in Section 5.

2 Preliminaries

In this section, we briefly review some notions from graph theory that will play a central role in this paper; for further information the reader is referred to [Harary, 1969]. Also some preliminaries concerning probability theory are provided.

Generally two types of graphs are discerned: undirected and directed ones.

Definition 2.1 *An* undirected graph *G is an ordered pair $G = (V(G), E(G))$ where $V(G)$ is a finite set of* vertices *and $E(G)$ is a set of unordered pairs (V_i, V_j), $V_i, V_j \in V(G)$, called* edges. *A* directed graph, *or* digraph *for short, is an ordered pair $G = (V(G), A(G))$ where $V(G)$ is a finite set of vertices and $A(G)$ is a set of ordered pairs (V_i, V_j), $V_i, V_j \in V(G)$, called* arcs.

Definition 2.2 *Let $G = (V(G), A(G))$ be a digraph. Vertex $V_i \in V(G)$ is called a* predecessor *of vertex $V_j \in V(G)$ in G if $(V_i, V_j) \in A(G)$; the set of all predecessors of vertex V_j in G is denoted by $\rho_G(V_j)$. Vertex V_j is called a* successor *of vertex V_i in G if $(V_i, V_j) \in A(G)$; the set of all successors of vertex V_i in G is denoted by $\sigma_G(V_i)$. The set of all* neighbours *of a vertex V_i in G is defined as $\nu_G(V_i) = \sigma_G(V_i) \cup \rho_G(V_i)$. The transitive closure of the set of predecessors of vertex V_i is denoted by $\rho^*_G(V_i)$; an element from $\rho^*_G(V_i)$ is called an* ancestor *of V_i. The transitive closure of the set of successors of vertex V_i is denoted by $\sigma^*_G(V_i)$; an element from $\sigma^*_G(V_i)$ is called a* descendant *of V_i. The* in-degree *of a vertex $V_i \in V(G)$ is defined as the number of vertices in $\rho_G(V_i)$; the* out-degree *of V_i is defined as the number of vertices in $\sigma_G(V_i)$.*

We will often drop the subscript G from ρ_G etc. as long as ambiguity cannot occur.

In the sequel, we will use several types of vertex sequences.

Definition 2.3 *Let $G = (V(G), E(G))$ be an undirected graph. A* path *from V_0 to V_k, $V_0, V_k \in V(G)$, in G is a sequence of vertices $V_0, V_1, \ldots, V_k$ such that $(V_{i-1}, V_i) \in E(G)$, $i = 1, \ldots, k$, $k \geq 0$; k is called the* length *of the path. A* cycle *is a path of length at least one from V_0 to V_0 for some $V_0 \in V(G)$. A graph G is called* cyclic *if it contains at least one cycle; otherwise it is called an* acyclic *graph.*

The previous notions have been introduced for undirected graphs; they can easily be extended, however, to apply to directed graphs by taking the directions of the arcs into account. The following definitions apply to directed graphs only.

Definition 2.4 *Let $G = (V(G), A(G))$ be a digraph. The* underlying graph *H of G is the undirected graph $H = (V(H), E(H))$ where $V(H) = V(G)$ and $E(H)$ is obtained from $A(G)$ by replacing each arc $(V_i, V_j) \in A(G)$ by an edge (V_i, V_j). A* chain *from V_0 to V_k, $V_0, V_k \in V(G)$, in G is a sequence of vertices that is a path in the underlying graph H of G; k is called the* length *of the chain. A* loop *in G is a chain of length at least one from V_0 to V_0 for some $V_0 \in V(G)$.*

To conclude the preliminaries on graph theory, several types of digraph are introduced.

Definition 2.5 *A digraph G is called* singly connected *if it does not contain any loops; otherwise it is called* multiply connected. *A singly connected digraph G is called a* directed tree *if each vertex in G has at most one predecessor.*

Singly connected digraphs often are termed *polytrees*, [Pearl, 1988].

We now provide some preliminaries concerning probability theory; in doing so we take an algebraic point of view. In the following definition, the notion of a free Boolean algebra is defined.

Definition 2.6 *A* Boolean algebra *$\mathcal{B}$ is a set of elements with two binary operations $\wedge$ (conjunction) and $\vee$ (disjunction), a unary operation $\neg$ (negation) and two constants $false$ and $true$ which (by equality according to logical truth tables) adhere to the usual axioms. A subset of elements $\mathcal{G} = \{g_1, \ldots, g_n\}$, $n \geq 1$, of a Boolean algebra $\mathcal{B}$ is said to be a set of* generators *for $\mathcal{B}$ if each element of $\mathcal{B}$ can be represented in terms of the elements $g_i \in \mathcal{G}$, $i = 1, \ldots, n$, and the operations $\wedge$, $\vee$ and $\neg$. A set of generators $\mathcal{G}$ for $\mathcal{B}$ is said to be* free *if every mapping of elements of $\mathcal{G}$ into an arbitrary Boolean algebra $\mathcal{B}'$ can be extended to a homomorphism of $\mathcal{B}$ into $\mathcal{B}'$. A Boolean algebra $\mathcal{B}$ is* free *if it has a finite set $\mathcal{A} = \{a_1, \ldots, a_n\}$, $n \geq 1$, of free generators; we say that $\mathcal{B}$ is (finitely)* generated *by $\mathcal{A}$. In the sequel, we will use $\mathcal{B}(a_1, \ldots, a_n)$ to denote the free Boolean algebra $\mathcal{B}$ generated by $\mathcal{A}$.*

Definition 2.7 *Let $\mathcal{B}(a_1, \ldots, a_n)$ be the free Boolean algebra generated by the set of free generators $\mathcal{A} = \{a_1, \ldots, a_n\}$, $n \geq 1$. Let $A = \{A_{i_1}, \ldots, A_{i_k}\}$, $0 \leq k \leq n$, be a set of variables over $\mathcal{B}(a_1, \ldots, a_n)$. Now, let $F_A : \mathcal{B}(a_1, \ldots, a_n)^k \rightarrow \mathcal{B}(a_1, \ldots, a_n)$ be the Boolean polynomial function defined by $F_A = true$ if $k = 0$ and $F_A(A_{i_1}, \ldots, A_{i_k}) = A_{i_1} \wedge \cdots \wedge A_{i_k}$ otherwise. Let $B_i = \{a_i, \neg a_i\}$, $i = 1, \ldots, n$. We define the* configuration function *C_A as the restriction of F_A to $B_{i_1} \times \cdots \times B_{i_k}$, that is, $C_A = F_A|_{B_{i_1} \times \cdots \times B_{i_k}}$. A function value c_A of C_A is called a* configuration *of A.*

In the sequel, we will often use the notation $\{c_A\}$ to denote the set of all configurations of the set of variables A; for a single variable A_i we will often write c_{A_i} instead of $c_{\{A_i\}}$. Furthermore, we will take the point of view of a free Boolean algebra $\mathcal{B}(a_1, \ldots, a_n)$ as a sample space being 'spanned' by a set of variables A_i taking values from $\{a_i, \neg a_i\}$, $i = 1, \ldots, n$. A variabele A_i over $\{a_i, \neg a_i\}$ will be termed a *probabilistic variable*. Note that the generalization to variables with more than two discrete multiple values is straightforward.

We now define the notion of a joint probability distribution on a Boolean algebra.

Definition 2.8 *Let $\mathcal{B}$ be a free Boolean algebra as defined above. Let Pr be a function $Pr\colon \mathcal{B} \rightarrow [0, 1]$ such that*

1. *Pr is* positive, *that is, for all $x \in \mathcal{B}$, we have $Pr(x) \geq 0$, and furthermore we have $Pr(false) = 0$,*

2. *Pr is* normed, *that is, we have $Pr(true) = 1$, and*

3. Pr is additive, *that is, for all* $x_1, x_2 \in \mathcal{B}$, *we have that if* $x_1 \wedge x_2 = false$ *then* $Pr(x_1 \vee x_2) = Pr(x_1) + Pr(x_2)$.

Then, Pr is called a joint probability distribution *on* $\mathcal{B}$.

It can easily be shown that the probability of an event is equivalent to the probability of the truth of the proposition asserting the occurrence of the event: we have that a joint probability distribution on a Boolean algebra of propositions has the usual properties. We now take conditional probabilities being defined as customary.

To conclude, we introduce the notion of an independency relation between sets of probabilistic variables.

Definition 2.9 *Let* $\mathcal{B}(a_1, \ldots, a_n)$ *be the free Boolean algebra generated by the set of free generators* $\mathcal{A} = \{a_1, \ldots, a_n\}$, $n \geq 1$. *Let Pr be a joint probability distribution on the algebra* $\mathcal{B}(a_1, \ldots, a_n)$. *Let* $A = \{A_1, \ldots, A_n\}$ *be the set of probabilistic variables* A_i *over* $B_i = \{a_i, \neg a_i\}$, $i = 1, \ldots, n$. *Now, let* $X, Y, Z \subseteq A$ *be sets of variables and let* C_X, C_Y *and* C_Z *be the configuration functions for the sets* X, Y *and* Z, *respectively. The set of variables X is said to be* conditionally independent *of Z given Y, denoted as* $I_{Pr}(X, Y, Z)$, *if* $Pr(C_X \mid C_Y \wedge C_Z) = Pr(C_X \mid C_Y)$; $I_{Pr}(X, Y, Z)$ *is called an* independency statement *for Pr. The set of all independency statements for Pr defines the relation* I_{Pr}, *called the* independency relation *of Pr*.

For an in-depth discussion of the properties of independency relations, the reader is referred to [Pearl, 1988].

3 Belief Networks

As we have mentioned before in our introduction, the belief network framework provides a formalism for representing knowledge concerning a joint probability distribution on a problem domain. In this section, the belief network formalism is defined and is assigned a meaning based on probability theory. Also, it is indicated which types of algorithm are required for reasoning with knowledge represented in the formalism.

3.1 The Belief Network Formalism

We introduce the notion of a belief network informally before giving a formal definition. A belief network comprises two parts: a *qualitative representation* and a *quantitative representation*. The qualitative part of a belief network takes the form of an acyclic digraph $G = (V(G), A(G))$ with vertices $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$, and arcs $A(G)$. Each vertex V_i in $V(G)$ represents a variable that can take one of the values *true* and *false*. We will adhere to the following notational convention: v_i denotes the proposition that the variable V_i takes the truth value *true*; $V_i = false$ will be denoted by $\neg v_i$. Informally speaking, we take an arc $(V_i, V_j) \in A(G)$ to represent a direct 'influential' or 'causal' relationship between the linked variables V_i and V_j; the direction of the arc designates V_j as the effect or consequence of the cause V_i. Absence of an arc between two vertices means that the corresponding variables do not influence each other directly. We take the digraph to be configured by an expert from human judgment; hence the phrase *belief* network. Associated with the graphical part of a belief network is a numerical assessment of the 'strengths' of the represented relationships: with each vertex is associated a set of (conditional) probabilities which describe the influence of the values of the predecessors of the vertex on the probabilities of the values of the vertex itself.

We now define the notion of a belief network more formally.

Definition 3.1 *A* belief network *is a tuple $B = (G, \Gamma)$ such that*

1. *$G = (V(G), A(G))$ is an acyclic digraph with vertices $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$, and*
2. *$\Gamma = \{\gamma_{V_i} \mid V_i \in V(G)\}$ is a set of real-valued nonnegative functions $\gamma_{V_i} : \{v_i, \neg v_i\} \times \{c_{\rho(V_i)}\} \rightarrow [0,1]$, called (conditional probability)* assessment functions, *such that for each configuration $c_{\rho(V_i)}$ of $\rho(V_i)$ we have that $\gamma_{V_i}(\neg v_i \mid c_{\rho(V_i)}) = 1 - \gamma_{V_i}(v_i \mid c_{\rho(V_i)})$, $i = 1, \ldots, n$.*

Note that in the previous definition V_i is viewed as a vertex from the graph and as a probabilistic variable, alternatively.

In order to link the qualitative and quantitative parts of a belief network, we assign a probabilistic meaning to the topology of the digraph of the network. We begin by introducing the notion of a blocked chain.

Definition 3.2 *Let $G = (V(G), A(G))$ be an acyclic directed graph with the vertex set $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$. Then, a chain s from vertex $V_i \in V(G)$ to vertex $V_j \in V(G)$ is* blocked *by a set $W \subseteq V(G)$ if one of the following conditions holds:*

1. *The chain s contains a vertex $X_2 \in W$ and two vertices $X_1, X_3 \in V(G)$ such that $(X_2, X_1) \in A(G)$ and $(X_2, X_3) \in A(G)$.*
2. *The chain s contains a vertex $X_2 \in W$ and two vertices $X_1, X_3 \in V(G)$ such that $(X_1, X_2) \in A(G)$ and $(X_2, X_3) \in A(G)$.*
3. *The chain s contains vertices $X_1, X_2, X_3 \in V(G)$ such that $(X_1, X_2) \in A(G)$ and $(X_3, X_2) \in A(G)$ and $\sigma^*(X_2) \cap W = \varnothing$.*

In defining the notion of a blocked chain, we have distinguished three cases; Figure 1 serves as a reference for these cases.

Building on the notion of a blocked chain, we define d-separation.

Definition 3.3 *Let $G = (V(G), A(G))$ be an acyclic digraph. Let $X, Y, Z \subseteq V(G)$ be sets of vertices. The set Y is said to* d-separate *the sets X and Z, denoted as $\langle X \mid Y \mid Z \rangle_G^d$, if for each $V_i \in X$ and $V_j \in Z$ every chain from V_i to V_j is blocked by Y.*

Case 1. Case 2. Case 3.

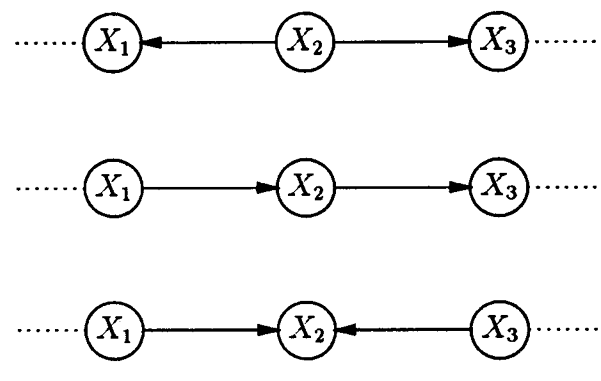

Figure 1: Chain Blocking.

We now assign a probabilistic meaning to a digraph by defining several types of relationships between joint probability distributions and digraphs.

Definition 3.4 *Let* $G = (V(G), A(G))$ *be an acyclic directed graph with the vertex set* $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$. *Let* $\mathcal{B}(v_1, \ldots, v_n)$ *be the free Boolean algebra generated by the set of free generators* $\{v_i \mid V_i \in V(G)\}$. *Furthermore, let* Pr *be a joint probability distribution on* $\mathcal{B}(v_1, \ldots, v_n)$ *and let* I_{Pr} *be the independency relation of* Pr. *Then,*

1. *The digraph* G *is called a* dependency map, *or* D-map *for short, for* Pr *if for all sets* $X, Y, Z \subseteq V(G)$ *we have: if* $I_{Pr}(X, Y, Z)$ *then* $\langle X \mid Y \mid Z \rangle_G^d$.

2. *The digraph* G *is called an* independency map, *or* I-map *for short, for* Pr *if for all* $X, Y, Z \subseteq V(G)$ *we have: if* $\langle X \mid Y \mid Z \rangle_G^d$ *then* $I_{Pr}(X, Y, Z)$.

3. *The digraph* G *is called a* perfect map *for* Pr *if* G *is both a dependency map and an independency map for* Pr.

Note that vertices that are not d-separated in a D-map for a joint probability distribution Pr are guaranteed to be dependent in Pr (then viewed as probabilistic variables); the D-map, however, may display a pair of dependent variables as a pair of d-separated vertices. On the other hand, vertices found to be d-separated in an I-map for Pr correspond to independent variables; those not d-separated, however, need not necessarily be dependent. A perfect map faithfully displays all dependencies and independencies embodied in Pr. It can easily be verified that every probability distribution has at least one I-map and at least one D-map. However, not every probability distribution has a perfect map.

The following proposition now states that the initial assessment functions of a belief network provide all information necessary for uniquely defining a joint probability distribution on the variables discerned that respects the independency relationships portrayed by the graphical part of the network. Henceforth, we will call this the *joint probability distribution defined by the network.*

Proposition 3.5 *Let $B = (G, \Gamma)$ be a belief network as defined in Definition 3.1, where $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$. Let $\mathcal{B}(v_1, \ldots, v_n)$ be the free Boolean algebra generated by $\{v_i \mid V_i \in V(G)\}$. Then,*

$$Pr(C_{V(G)}) \quad = \quad \prod_{i=1,\ldots,n} \gamma_{V_i}(V_i \mid C_{\rho(V_i)})$$

defines a joint probability distribution Pr on $\mathcal{B}(v_1, \ldots, v_n)$ such that G is an I-map for Pr.

Proof. A digraph without any (directed) cycles allows at least one total ordering of its vertices such that any successor of a vertex in the graph follows it in the ordering. It follows that there is an ordering of the probabilistic variables such that in applying the chain rule each variable is conditioned only on the variables preceding it in the ordering. Choosing an appropriate ordering of $V(G)$, the independency relation portrayed by G can be exploited. By taking $Pr(v_i \mid c_{\rho(V_i)}) = \gamma_{V_i}(v_i \mid c_{\rho(V_i)})$ for each $V_i \in V(G)$ and all configurations $c_{\rho(V_i)}$ of $\rho(V_i)$, the property stated in the proposition follows immediately. □

In the sequel, it will be shown that a belief network may be exploited for reasoning purposes. When reasoning with the belief network, evidence may become available concerning some of the variables which is subsequently entered into the network. Pearl distinguishes between two types of incoming information: *specific evidence* and *virtual evidence.* Specific evidence represents direct observations that affect the probabilities of the values of some variables in the network; virtual evidence concerns judgments based on indirect observations that are out of the scope of the network but have bearing on variables within the network. In this paper, we restrict the discussion to specific evidence only and merely mention that virtual evidence is handled essentially the same way. For further details, the reader is referred to [Pearl, 1988].

To conclude this section, we will define some more notions.

Definition 3.6 *Let V be a set of probabilistic variables. A variable $V_i \in V$ for which either the specific evidence $V_i = true$ or $V_i = false$ has become available is called* instantiated*; if no evidence has been obtained as yet for a variable, it is called* uninstantiated. *Now, let $X \subseteq V$ be the set of instantiated variables from V. The configuration c_X of X is called a* partial configuration *of V, denoted by $\tilde{c}_V$.*

Note that if none of the variables in a set of probabilistic variables V are instantiated, we have $\tilde{c}_V = true$ by definition. Also note that the notation $\tilde{c}_V$ introduced above allows for referring to the subset of instantiated variables of the set V without specifying this subset explicitly.

3.2 Reasoning with a Belief Network

In the previous section, the notion of a belief network has been introduced as a means for representing a joint probability distribution. For making probabilistic statements concerning the variables discerned in the problem domain, two algorithms have to be associated with a belief network:

- an algorithm for (efficiently) computing probabilities of interest from the network, and
- an algorithm for processing evidence, that is, for entering evidence into the network and subsequently (efficiently) computing the revised probability distribution given the evidence.

Since a joint probability distribution on the variables is uniquely defined by the conditional probability assessment functions, any probability of interest can be computed from these functions. Equally, the impact of a value of a specific variable becoming known, on each of the other variables can be computed from the initial assessment functions. Now, observe that the conditional probability assessment functions describe the joint probability distribution locally for each vertex and its predecessors. Calculation of a (revised) probability from the joint probability distribution defined by the belief network in a straightforward manner, however, will generally not be restricted to performing computations which are local in terms of the graphical part of the network. In the literature therefore, several less naive algorithms for computing probabilities of interest from a belief network and for processing evidence in the network have been proposed, for example by J. Pearl, [Pearl, 1988], and by S.L. Lauritzen and D.J. Spiegelhalter, [Lauritzen & Spiegelhalter, 1988]. Although all schemes proposed for evidence propagation are based on probability theory, they differ considerably with respect to the algorithms employed and their complexity; it should be noted that in general probabilistic inference in belief networks without any restrictions is NP-hard, [Cooper, 1990].

All schemes for reasoning with a belief network have an important property in common: the graphical part of the network is exploited more or less directly as a computational architecture for both algorithms. In this paper which focuses on Pearl's work, we will often adhere to an object-oriented approach and view the graphical part of a belief network as a computational architecture by taking the vertices of the graph as autonomous objects having a local processor capable of performing certain probabilistic computations and a local memory in which the associated conditional probability assessment function is stored; the arcs of the graph are viewed as bi-directional communication channels through which the objects send messages.

4 Pearl's Algorithms

J. Pearl has presented a set of algorithms for computing probabilities and for processing evidence in a belief network comprising a singly connected digraph. The basic idea of these algorithms is that the vertices of the graphical part of the network, viewed as autonomous objects, send each other enough information about the joint probability distribution and the evidence obtained sofar to enable each vertex to compute the (revised) probabilities of its values from the information it receives from its neighbours and its own local conditional probability assessment function. The impact of a piece of evidence entered into the network is then viewed as a perturbation that spreads through the network by message-passing between neighbouring vertices. In Section 4.1, we present a simplified version of Pearl's algorithms that applies to directed trees only. In Section 4.2, these algorithms are extended to more general algorithms applying to singly connected digraphs.

4.1 Directed Trees

In this section, we confine ourselves to belief networks where the graphical part is a directed tree, that is, a vertex may have several successors and at most one predecessor. Now consider computing probabilities from such a network. It will be evident that at any time the probabilities of the values of a vertex in the directed tree are dependent upon the evidence entered for its ancestors and its descendants, that is, upon all data observed sofar.

Lemma 4.1 *Let $B = (G, \Gamma)$ be a belief network where $G = (V(G), A(G))$ is a directed tree, and let Pr be the joint probability distribution defined by B. Let $V_i \in V(G)$ be a vertex in G, and let $V_i^- = \{V_j \mid \text{there is a path from } V_i \text{ to } V_j\}$ and $V_i^+ = V(G) \setminus V_i^-$. Then,*

$$Pr(V_i \mid \tilde{c}_{V(G)}) = \alpha \cdot Pr(\tilde{c}_{V_i^-} \mid V_i) \cdot Pr(V_i \mid \tilde{c}_{V_i^+})$$

where $\tilde{c}_{V(G)} = \tilde{c}_{V_i^-} \wedge \tilde{c}_{V_i^+}$ and α is a normalization constant.

Proof. For the probabilities of the values of the probabilistic variable V_i, we have that

$$\begin{aligned} Pr(V_i \mid \tilde{c}_{V(G)}) &= Pr(V_i \mid \tilde{c}_{V_i^-} \wedge \tilde{c}_{V_i^+}) = \\ &= \frac{Pr(\tilde{c}_{V_i^-} \wedge \tilde{c}_{V_i^+} \mid V_i) \cdot Pr(V_i)}{Pr(\tilde{c}_{V_i^-} \wedge \tilde{c}_{V_i^+})} \end{aligned}$$

using Bayes' Rule for the last equality. Now consider Figure 2 showing a fragment of the directed tree G. From G we observe that $\langle X \mid \{V_i\} \mid Y \rangle_G^d$ for all subsets $X \subseteq V_i^-$ and $Y \subseteq V_i^+$. Since G is an I-map for the joint probability distribution Pr it follows that $I_{Pr}(X, \{V_i\}, Y)$ for all $X \subseteq V_i^-$, $Y \subseteq V_i^+$. From this observation we have

$$\begin{aligned} Pr(V_i \mid \tilde{c}_{V(G)}) &= \frac{Pr(\tilde{c}_{V_i^-} \mid V_i) \cdot Pr(\tilde{c}_{V_i^+} \mid V_i) \cdot Pr(V_i)}{Pr(\tilde{c}_{V_i^-} \wedge \tilde{c}_{V_i^+})} = \\ &= Pr(\tilde{c}_{V_i^-} \mid V_i) \cdot Pr(V_i \mid \tilde{c}_{V_i^+}) \cdot \frac{Pr(\tilde{c}_{V_i^+})}{Pr(\tilde{c}_{V_i^-} \wedge \tilde{c}_{V_i^+})} \end{aligned}$$

Now observe that the factor

$$\frac{Pr(\tilde{c}_{V_i^+})}{Pr(\tilde{c}_{V_i^-} \wedge \tilde{c}_{V_i^+})} = \frac{1}{Pr(\tilde{c}_{V_i^-} \mid \tilde{c}_{V_i^+})}$$

is dependent on the variable V_i but not on its values; this factor may therefore be viewed as a constant for V_i, which will subsequently be denoted by α. It follows that

$$Pr(V_i \mid \tilde{c}_{V(G)}) = \alpha \cdot Pr(\tilde{c}_{V_i^-} \mid V_i) \cdot Pr(V_i \mid \tilde{c}_{V_i^+})$$

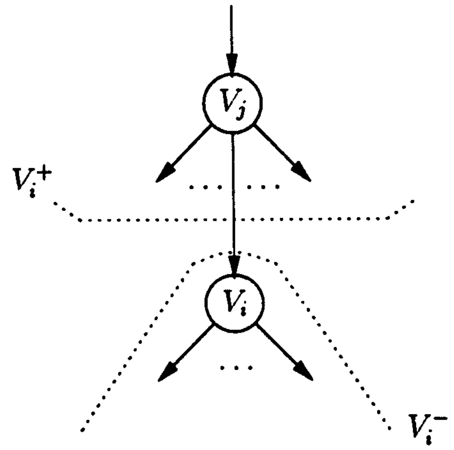

Figure 2: A Fragment of a Directed Tree.

The constant α is generally referred to as a *normalization constant* because it may be computed from $Pr(v_i \mid \tilde{c}_{V(G)}) + Pr(\neg v_i \mid \tilde{c}_{V(G)}) = 1$. □

As Pearl notes, the property stated in the previous lemma provides a generalization to Bayes' Rule for recursive updating when the pieces of evidence obtained are not conditionally independent given the updated variable, [Pearl, 1988].

The previous lemma shows that the probabilities of the values of a vertex can be written in terms of two factors describing the influence of evidence concerning the descendants of the vertex and the influence of evidence concerning the other vertices in the tree separately. The following definition introduces separate functions to describe these influences.

Definition 4.2 *Let $B = (G, \Gamma)$ be a belief network as before and let Pr be the joint probability distribution defined by B. Let $V_i \in V(G)$ be a probabilistic variable in G and let V_i^- and V_i^+ be as in the previous lemma. The* compound causal parameter *π_{V_i} for V_i is the function $\pi_{V_i}\colon \{v_i, \neg v_i\} \rightarrow [0,1]$ defined by*

$$\pi_{V_i}(V_i) = Pr(V_i \mid \tilde{c}_{V_i^+})$$

The compound diagnostic parameter *λ_{V_i} for V_i is the function $\lambda_{V_i} : \{v_i, \neg v_i\} \rightarrow [0,1]$ defined by*

$$\lambda_{V_i}(V_i) = Pr(\tilde{c}_{V_i^-} \mid V_i)$$

Sometimes, the compound causal parameter for a variable is called the *predictive support* for the variable; the compound diagnostic parameter for a variable then is called its *retrospective support*.

We reconsider the previous definition for the boundary vertices in the digraph of a belief network for which either the in-degree or the out-degree equals zero. We first address the case where these vertices are uninstantiated. Recall that in this section we confined ourselves to directed trees only. The digraph therefore has one vertex with in-degree equal to zero; this vertex will be called the *root* of the tree. Now consider the compound causal parameter π_W for the root W of a directed tree G. We observe that for W the set W^+ is empty. So, $\tilde{c}_{W^+} = true$. The compound causal parameter $\pi_W : \{w, \neg w\} \rightarrow [0,1]$ for W therefore is defined by $\pi_W(W) = Pr(W)$ where Pr is the joint probability distribution defined by the network. The directed tree may further have several vertices with out-degree equal to zero; these vertices are called the *leaves* of the tree. Considering the compound diagnostic parameter for a leaf V of the tree, we observe that the set V^- consists of V only. The compound diagnostic parameter $\lambda_V : \{v, \neg v\} \rightarrow [0,1]$ for V therefore is defined by $\lambda_V(V) = 1$. To conclude our discussion of the compound parameters, we consider instantiated variables. For a variable V_i for which the evidence $V_i = true$ is observed we find $\pi_{V_i}(V_i) = Pr(V_i \mid \tilde{c}_{V_i^+})$, and $\lambda_{V_i}(v_i) = Pr(\tilde{c}_{V_i^-} \mid v_i)$ and $\lambda_{V_i}(\neg v_i) = 0$; a similar observation holds for the case where we have observed the evidence $V_i = false$.

The following lemma shows that the compound causal and diagnostic parameters for a variable provide it with enough information for computing the probabilities of its values, that is, no further knowledge of the joint probability distribution is needed; this lemma is known as the *data fusion lemma*, [Pearl, 1988].

Lemma 4.3 *Let $B = (G, \Gamma)$ be a belief network as before and let Pr be the joint probability distribution defined by B. Let $V_i \in V(G)$ be a probabilistic variable. Let the compound causal parameter π_{V_i} for V_i and the compound diagnostic parameter λ_{V_i} for V_i be defined as in Definition 4.2. Then,*

$$Pr(V_i \mid \tilde{c}_{V(G)}) = \alpha \cdot \pi_{V_i}(V_i) \cdot \lambda_{V_i}(V_i)$$

where α is a normalization constant.

Proof. The property stated in the lemma follows immediately from Lemma 4.1 and Definition 4.2. □

The compound diagnostic parameter for a vertex specifies information concerning the joint probability distribution from all its descendants combined; in general, a similar observation applies to the compound causal parameter for the vertex. To be able to exploit the graphical part of a belief network as a computational architecture, these compound parameters have to be decomposed into parameters corresponding with each of the successors and the predecessor of the vertex separately. The following definition introduces separate parameters to this end; the Lemmas 4.5 and 4.6 will show the decomposition of the compound parameters.

Definition 4.4 *Let $B = (G, \Gamma)$ be a belief network as before and let Pr be the joint probability distribution defined by B. For each vertex $V \in V(G)$, let V^- and V^+ be as in Lemma 4.1. Let V_i be a probabilistic variable in G having a successor V_k. The* causal parameter $\pi_{V_k}^{V_i}$ from V_i to V_k *is the function $\pi_{V_k}^{V_i} : \{v_i, \neg v_i\} \rightarrow [0,1]$ defined by*

$$\pi_{V_k}^{V_i}(V_i) = Pr(V_i \mid \tilde{c}_{V_k^+})$$

Now, let V_i be a probabilistic variable in G having the predecessor V_j. The diagnostic parameter $\lambda_{V_i}^{V_j}$ from V_i to V_j *is the function* $\lambda_{V_i}^{V_j} : \{v_j, \neg v_j\} \rightarrow [0,1]$ *defined by*

$$\lambda_{V_i}^{V_j}(V_j) = Pr(\tilde{c}_{V_i^-} \mid V_j)$$

The separate causal and diagnostic parameters defined above are associated with the arcs of the directed tree of the belief network as shown in Figure 3, and can be viewed as messages the objects send each other through the communication channels. It will be evident that the root of the directed tree sends no diagnostic parameter, and that the leaves of the tree send no causal parameters. Note that for a variable V_i for which the evidence $V_i = true$ is observed, we find $\pi_{V_k}^{V_i}(v_i) = 1$ and $\pi_{V_k}^{V_i}(\neg v_i) = 0$; a similar observation holds for the case where $V_i = false$ is observed.

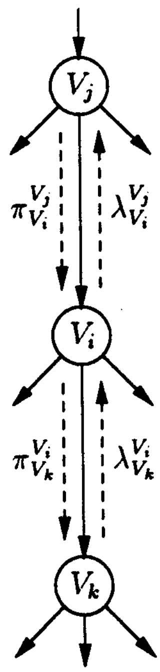

Figure 3: The Causal and Diagnostic Parameters.

The following lemma shows that a vertex can compute its compound causal parameter from its associated conditional probability assessment function and the causal parameter it receives from its predecessor.

Lemma 4.5 *Let $B = (G, \Gamma)$ be a belief network as before. Let $V_i \in V(G)$ be a probabilistic variable having the predecessor V_j. Furthermore, let π_{V_i} be the compound causal parameter for V_i, and let $\pi_{V_i}^{V_j}$ be the causal parameter from V_j to V_i. Then,*

$$\pi_{V_i}(V_i) = \sum_{c_{V_j}} \gamma_{V_i}(V_i \mid c_{V_j}) \cdot \pi_{V_i}^{V_j}(c_{V_j})$$

Proof. Let Pr be the joint probability distribution defined by the belief network B. Furthermore, let V_i^+ be as in Lemma 4.1. Then, from Definition 4.2 we have

$$\begin{aligned}\pi_{V_i}(V_i) &= Pr(V_i \mid \tilde{c}_{V_i^+}) = \\ &= Pr(V_i \mid v_j \wedge \tilde{c}_{V_i^+}) \cdot Pr(v_j \mid \tilde{c}_{V_i^+}) + Pr(V_i \mid \neg v_j \wedge \tilde{c}_{V_i^+}) \cdot Pr(\neg v_j \mid \tilde{c}_{V_i^+})\end{aligned}$$

Now consider once more Figure 2 showing a fragment of the directed tree G. From G we observe that $\langle\{V_i\}|\{V_j\}|X\rangle_G^d$ for all subsets $X \subseteq V_i^+$. Since G is an I-map for the joint probability distribution Pr, it follows that $I_{Pr}(\{V_i\}, \{V_j\}, X)$ for all $X \subseteq V_i^+$. Exploiting this observation, we find

$$\pi_{V_i}(V_i) = Pr(V_i \mid v_j) \cdot Pr(v_j \mid \tilde{c}_{V_i^+}) + Pr(V_i \mid \neg v_j) \cdot Pr(\neg v_j \mid \tilde{c}_{V_i^+})$$

The probabilities $Pr(V_i \mid V_j)$ have been specified as function values $\gamma_{V_i}(V_i \mid V_j)$ of the conditional probability assessment function γ_{V_i} associated with V_i and therefore are known to vertex V_i. Furthermore, we observe that vertex V_i receives the probabilities $Pr(V_j \mid \tilde{c}_{V_i^+})$ as function values $\pi_{V_i}^{V_j}(V_j)$ of the causal parameter $\pi_{V_i}^{V_j}$ from its predecessor V_j. Substitution yields

$$\begin{aligned}\pi_{V_i}(V_i) &= \gamma_{V_i}(V_i \mid v_j) \cdot \pi_{V_i}^{V_j}(v_j) + \gamma_{V_i}(V_i \mid \neg v_j) \cdot \pi_{V_i}^{V_j}(\neg v_j) = \\ &= \textstyle\sum_{c_{V_j}} \gamma_{V_i}(V_i \mid c_{V_j}) \cdot \pi_{V_i}^{V_j}(c_{V_j})\end{aligned}$$

□

Equally, a vertex can compute its compound diagnostic parameter from the separate diagnostic parameters it receives from its successors. This property is stated in the following lemma.

Lemma 4.6 *Let $B = (G, \Gamma)$ be a belief network as before. Let $V_i \in V(G)$ be an uninstantiated probabilistic variable with $\sigma(V_i) = \{V_{i_1}, \ldots, V_{i_m}\}$, $m \geq 1$. Furthermore, let λ_{V_i} be the compound diagnostic parameter for V_i, and for each successor $V_{i_j} \in \sigma(V_i)$ of V_i, let $\lambda_{V_{i_j}}^{V_i}$ be the diagnostic parameter from V_{i_j} to V_i. Then,*

$$\lambda_{V_i}(V_i) = \textstyle\prod_{j=1,\ldots,m} \lambda_{V_{i_j}}^{V_i}(V_i)$$

Proof. Let Pr be the joint probability distribution defined by the belief network B. For each vertex $V \in V(G)$, let V^- be as in Lemma 4.1. Then, by definition we have

$$\lambda_{V_i}(V_i) = Pr(\tilde{c}_{V_i^-} \mid V_i)$$

Since V_i is an uninstantiated variable, we have that $\tilde{c}_{V_i^-} = \tilde{c}_{V_{i_1}^-} \wedge \cdots \wedge \tilde{c}_{V_{i_m}^-}$. So,

$$\lambda_{V_i}(V_i) = Pr(\tilde{c}_{V_{i_1}^-} \wedge \cdots \wedge \tilde{c}_{V_{i_m}^-} \mid V_i)$$

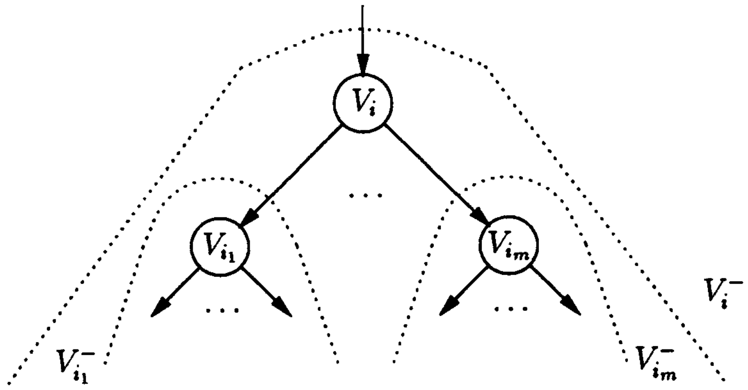

Figure 4: Exploiting d-Separation for Computing $\lambda_{V_i}(V_i)$.

Now consider Figure 4 showing a fragment of the directed tree G. From G we observe that $\langle X \mid \{V_i\} \mid Y \rangle_G^d$ for all subsets $X \subseteq V_{i_j}^-$ and $Y \subseteq \bigcup_{k=1,\ldots,m,k\neq j} V_{i_k}^-$, $j = 1,\ldots,m$. Since G is an I-map for Pr, it follows that $I_{Pr}(X, \{V_i\}, Y)$ for $X \subseteq V_{i_j}^-$, $Y \subseteq \bigcup_{k=1,\ldots,m,k\neq j} V_{i_k}^-$, $j = 1,\ldots,m$. From this observation, we find

$$\lambda_{V_i}(V_i) = Pr(\tilde{c}_{V_{i_1}^-} \mid V_i) \cdot \ldots \cdot Pr(\tilde{c}_{V_{i_m}^-} \mid V_i)$$

Vertex V_i receives the probabilities $Pr(\tilde{c}_{V_{i_j}^-} \mid V_i)$ as function values $\lambda_{V_{i_j}}^{V_i}(V_i)$ of the diagnostic parameter $\lambda_{V_{i_j}}^{V_i}$ from its successor V_{i_j}, $j = 1,\ldots,m$. Substitution yields

$$\begin{aligned} \lambda_{V_i}(V_i) &= \lambda_{V_{i_1}}^{V_i}(V_i) \cdot \ldots \cdot \lambda_{V_{i_m}}^{V_i}(V_i) = \\ &= \prod_{j=1,\ldots,m} \lambda_{V_{i_j}}^{V_i}(V_i) \end{aligned}$$

□

Note that the previous lemma applies to *uninstantiated* variables only. However, the property mentioned in the lemma can be taken to hold for an instantiated variable V_i as well if we model entering the evidence $V_i = true$ into the network by adding a dummy successor D to V_i that sends the diagnostic parameter $\lambda_D^{V_i}$ with $\lambda_D^{V_i}(v_i) = 1$ and $\lambda_D^{V_i}(\neg v_i) = 0$ to V_i; a similar observation is made for the case where the evidence $V_i = false$ is entered into the network.

We have shown that a vertex can compute the probabilities of its values from its local conditional probability assessment function and the diagnostic parameters and the causal parameter it receives from its neighbours. Now observe that the vertex in turn has to compute causal and diagnostic parameters to send to its respective neighbours. The following lemma shows that a vertex can compute the diagnostic parameter to send to its predecessor from its own assessment function and the diagnostic parameters it receives from its successors; in other words, to this purpose it combines its own information about the joint probability distribution with the information it receives concerning the evidence obtained sofar for its descendants.

Lemma 4.7 *Let $B = (G, \Gamma)$ be a belief network as before. Let $V_i \in V(G)$ be a probabilistic variable having the predecessor V_j. Let λ_{V_i} be the compound diagnostic parameter for V_i, and let $\lambda_{V_i}^{V_j}$ be the diagnostic parameter from V_i to V_j. Then,*

$$\lambda_{V_i}^{V_j}(V_j) = \sum_{c_{V_i}} \lambda_{V_i}(c_{V_i}) \cdot \gamma_{V_i}(c_{V_i} \mid V_j)$$

Proof. Let Pr be the joint probability distribution defined by the belief network B. Furthermore, let V_i^- be as before. Then, from Definition 4.4, we have

$$\begin{aligned} \lambda_{V_i}^{V_j}(V_j) &= Pr(\tilde{c}_{V_i^-} \mid V_j) = \\ &= Pr(\tilde{c}_{V_i^-} \mid v_i \wedge V_j) \cdot Pr(v_i \mid V_j) + Pr(\tilde{c}_{V_i^-} \mid \neg v_i \wedge V_j) \cdot Pr(\neg v_i \mid V_j) \end{aligned}$$

The reader is referred once more to Figure 2 showing a fragment of the directed tree G. From $\langle X | \{V_i\} | \{V_j\} \rangle_G^d$ for any subset $X \subseteq V_i^-$ and G being an I-map for Pr it follows that $I_{Pr}(X, \{V_i\}, \{V_j\})$ for $X \subseteq V_i^-$. So,

$$\lambda_{V_i}^{V_j}(V_j) = Pr(\tilde{c}_{V_i^-} \mid v_i) \cdot Pr(v_i \mid V_j) + Pr(\tilde{c}_{V_i^-} \mid \neg v_i) \cdot Pr(\neg v_i \mid V_j)$$

The probabilities $Pr(V_i \mid V_j)$ have been specified as function values $\gamma_{V_i}(V_i \mid V_j)$ of the conditional probability assessment function γ_{V_i} associated with vertex V_i and hence are available to V_i. From Definition 4.2 we have that the probabilities $Pr(\tilde{c}_{V_i^-} \mid V_i)$ equal the function values $\lambda_{V_i}(V_i)$ of the compound diagnostic parameter λ_{V_i} for V_i. Substitution yields

$$\begin{aligned} \lambda_{V_i}^{V_j}(V_j) &= \lambda_{V_i}(v_i) \cdot \gamma_{V_i}(v_i \mid V_j) + \lambda_{V_i}(\neg v_i) \cdot \gamma_{V_i}(\neg v_i \mid V_j) = \\ &= \sum_{c_{V_i}} \lambda_{V_i}(c_{V_i}) \cdot \gamma_{V_i}(c_{V_i} \mid V_j) \end{aligned}$$

□

Similarly, the causal parameter a vertex has to send to a successor can be computed from its compound causal parameter and the diagnostic parameters it receives from its other successors. The following lemma states this property.

Lemma 4.8 *Let $B = (G, \Gamma)$ be a belief network as before. Let $V_i \in V(G)$ be an uninstantiated probabilistic variable with $\sigma(V_i) = \{V_{i_1}, \ldots, V_{i_m}\}$, $m \geq 1$. Furthermore, let π_{V_i} be the compound causal parameter for V_i. For each successor $V_{i_j} \in \sigma(V_i)$ of V_i, let $\pi_{V_{i_j}}^{V_i}$ be the causal parameter from V_i to V_{i_j}, and let $\lambda_{V_{i_j}}^{V_i}$ be the diagnostic parameter from V_{i_j} to V_i. Then,*

$$\pi_{V_{i_j}}^{V_i}(V_i) = \alpha \cdot \pi_{V_i}(V_i) \cdot \prod_{k=1,\ldots,m,k\neq j} \lambda_{V_{i_k}}^{V_i}(V_i)$$

where α is a normalization constant.

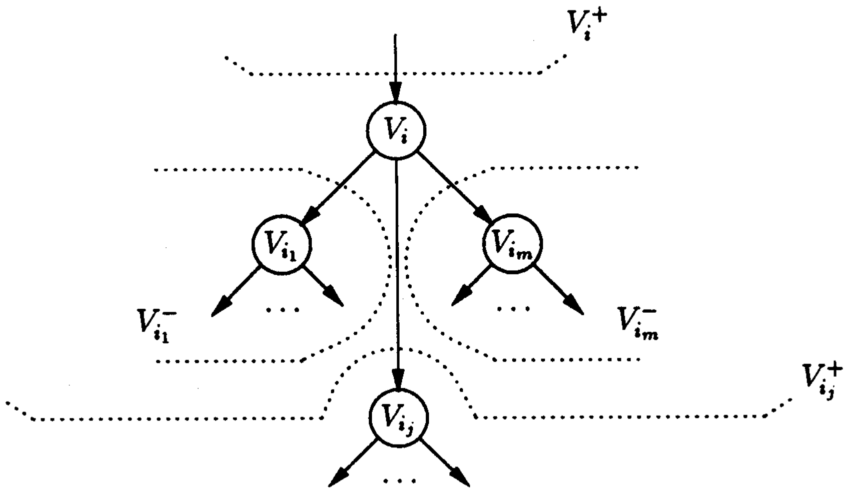

Figure 5: Exploiting d-Separation for Computing $\pi_{V_{i_j}}^{V_i}(V_i)$.

Proof. Let Pr be the joint probability distribution defined by the belief network B. For each vertex $V \in V(G)$, let V^+ and V^- be as before. Then, by definition we have

$$\begin{aligned}\pi_{V_{i_j}}^{V_i}(V_i) &= Pr(V_i \mid \tilde{c}_{V_{i_j}^+}) = \\ &= \frac{Pr(\tilde{c}_{V_{i_j}^+} \mid V_i) \cdot Pr(V_i)}{Pr(\tilde{c}_{V_{i_j}^+})}\end{aligned}$$

using Bayes' Rule for the last equality. Now consider Figure 5 showing a fragment of the directed tree G. Since V_i is an uninstantiated variable, we have that $\tilde{c}_{V_{i_j}^+} = \tilde{c}_{V_i^+} \wedge (\bigwedge_{k=1,\ldots,m,k\neq j} \tilde{c}_{V_{i_k}^-})$. So,

$$\pi_{V_{i_j}}^{V_i}(V_i) = \frac{Pr(\tilde{c}_{V_i^+} \wedge (\bigwedge_{k=1,\ldots,m,k\neq j} \tilde{c}_{V_{i_k}^-}) \mid V_i) \cdot Pr(V_i)}{Pr(\tilde{c}_{V_{i_j}^+})}$$

From G we observe that $\langle X|\{V_i\}|Y\rangle_G^d$ for $X \subseteq V_i^+$ and $Y \subseteq V_{i_k}^-$, $k = 1, \ldots, m$, and for $X \subseteq V_{i_k}^-$ and $Y \subseteq V_{i_l}^-$, $k = 1, \ldots, m$, $l = 1, \ldots, m$, $k \neq l$. Now, exploiting the observation that G is an I-map for Pr, it follows that

$$\begin{aligned}\pi_{V_{i_j}}^{V_i}(V_i) &= \frac{Pr(\tilde{c}_{V_i^+} \mid V_i) \cdot \prod_{k=1,\ldots,m,k\neq j} Pr(\tilde{c}_{V_{i_k}^-} \mid V_i) \cdot Pr(V_i)}{Pr(\tilde{c}_{V_{i_j}^+})} = \\ &= \frac{Pr(V_i \mid \tilde{c}_{V_i^+}) \cdot \prod_{k=1,\ldots,m,k\neq j} Pr(\tilde{c}_{V_{i_k}^-} \mid V_i) \cdot Pr(\tilde{c}_{V_i^+})}{Pr(\tilde{c}_{V_{i_j}^+})}\end{aligned}$$

using Bayes' Rule once more for the last equality. From Definition 4.2 we have that the probabilities $Pr(V_i \mid \tilde{c}_{V_i^+})$ equal the function values $\pi_{V_i}(V_i)$ of the compound causal parameter π_{V_i} for V_i. The probabilities $Pr(\tilde{c}_{V_{i_k}^-} \mid V_i)$ equal the function values $\lambda_{V_{i_k}}^{V_i}(V_i)$ of the diagnostic parameter $\lambda_{V_{i_k}}^{V_i}$ vertex V_i receives from its successor V_{i_k}. In addition, we observe that the factor

$$\frac{Pr(\tilde{c}_{V_i^+})}{Pr(\tilde{c}_{V_{i_j}^+})} = \frac{1}{Pr(\tilde{c}_{V_{i_j}^+} \mid \tilde{c}_{V_i^+})}$$

is dependent on the variables V_i and V_{i_j} but not on their values; this factor may therefore be viewed as a normalization constant for V_i and V_{i_j}, which will subsequently be denoted by α. Substitution yields

$$\pi_{V_{i_j}}^{V_i}(V_i) = \alpha \cdot \pi_{V_i}(V_i) \cdot \prod_{k=1,\ldots,m,k\neq j} \lambda_{V_{i_k}}^{V_i}(V_i)$$

□

The previous lemma applies to uninstantiated variables only. The property, however, can be taken to hold for instantiated variables in the way described before.

Now all parameters and their computation have been considered, we will take a closer look at how the influence of new evidence will spread through the network. Initially, the belief network is in an equilibrium state: recomputing the parameters will not result in a change in any of them. When a piece of evidence for a specific variable is entered into the belief network, this equilibrium is perturbed: the parameters from that variable to its neighbours are modified to reflect the entered evidence. These modifications activate updating parameters throughout the network: after receiving modified parameters, the neighbours in turn compute new parameters to send to their neighbours. The way the causal and diagnostic parameters are computed enforces that the influence of the entered evidence is passed on correctly. We note that the neighbour from which the modified parameter originated will not receive a new parameter since a causal parameter or a diagnostic parameter to a vertex is not affected by the diagnostic parameter or the causal parameter, respectively, from that vertex. This property guarantees that feedback and circular reasoning are prevented and that the evidence is propagated through the network in a single pass. The belief network will therefore reach a new equilibrium after a finite number of steps.

We conclude this section by stating some additional properties concerning the compound diagnostic parameter. These properties are useful when investigating the spreading of evidence.

Lemma 4.9 *Let $B = (G, \Gamma)$ be a belief network as before. For each vertex $V_i \in V(G)$, let $\lambda_{V_i}(V_i)$ be the compound diagnostic parameter for V_i. If $\tilde{c}_{V(G)} = true$, then $\lambda_{V_i}(V_i) = 1$ for all $V_i \in V(G)$.*

Proof. The property stated in the lemma is proven by (reverse) induction on the depth of the directed tree G. Let n be the maximal depth of the tree.

Induction Basis
The property holds for every leaf of the tree on depth n by definition.

Induction Hypothesis
For a specific $d \leq n$, we assume that $\lambda_{V_i}(V_i) = 1$ for all vertices V_i at depth $d, d+1, \ldots, n$.

Induction Step
Now consider a vertex V_i at depth $d-1$ in the tree. We distinguish two cases. If V_i is a leaf of the tree, then $\lambda_{V_i}(V_i) = 1$ by definition. Now, suppose that V_i has m successors $V_{i_1}, \ldots, V_{i_m}$, $m \geq 1$. From $\tilde{c}_{V(G)} = true$, it follows that V_i is an uninstantiated variable. Therefore, it follows from Lemma 4.6 that

$$\lambda_{V_i}(V_i) \;=\; \prod_{j=1,\ldots,m} \lambda_{V_{i_j}}^{V_i}(V_i)$$

For each parameter $\lambda_{V_{i_j}}^{V_i}(V_i)$ we have from Lemma 4.7 that

$$\lambda_{V_{i_j}}^{V_i}(V_i) \;=\; \sum_{c_{V_{i_j}}} \lambda_{V_{i_j}}(c_{V_{i_j}}) \cdot \gamma_{V_{i_j}}(c_{V_{i_j}} \mid V_i)$$

Since vertex V_{i_j} is a successor of V_i, it is at depth d in the directed tree G. From the Induction Hypothesis we have that $\lambda_{V_{i_j}}(V_{i_j}) = 1$. So,

$$\lambda_{V_{i_j}}^{V_i}(V_i) \;=\; \sum_{c_{V_{i_j}}} \gamma_{V_{i_j}}(c_{V_{i_j}} \mid V_i) \;=\; 1$$

From $\lambda_{V_{i_j}}^{V_i}(V_i) = 1$ for all successors V_{i_j} of V_i, it follows that $\lambda_{V_i}(V_i) = 1$. Since vertex V_i was chosen arbitrarily, it follows that for each vertex $V_i \in V(G)$, we have $\lambda_{V_i}(V_i) = 1$. □

It will be evident that the previous lemma can be taken to apply to subtrees of a directed tree, yielding the following property.

Corollary 4.10 *Let $B = (G, \Gamma)$ be a belief network as before. Let $V_i \in V(G)$ be a probabilistic variable and let $\lambda_{V_i}(V_i)$ be its compound diagnostic parameter. If $\tilde{c}_{V_i^-} = true$, then $\lambda_{V_i}(V_i) = 1$.*

4.2 Singly Connected Digraphs

In the previous section we have discussed Pearl's algorithms for computing probabilities and for processing evidence in a belief network comprising a directed tree. In this section, these algorithms are extended to apply to belief networks where the qualitative part is a singly connected digraph.

The absence of loops in a singly connected digraph allows for algorithms for computing probabilities and for processing evidence that are based on the same basic ideas as the

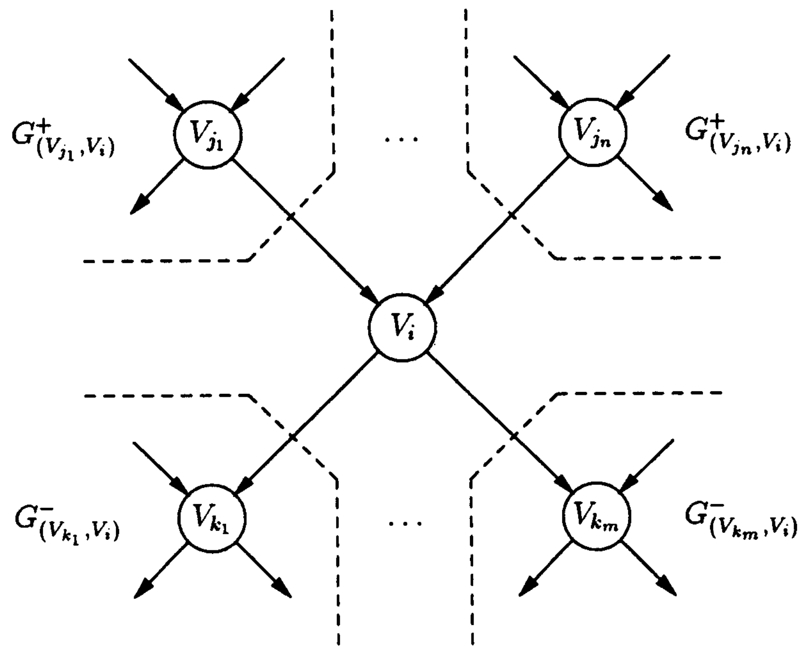

Figure 6: Upper and Lower Graphs.

algorithms applying to directed trees. More in specific, the following topological property can be exploited: the removal of any arc from a singly connected digraph splits the graph into two separate components. From this property we have that in a singly connected digraph G we can identify with a vertex V_i having m neighbours, m subgraphs of G each containing a neighbour of V_i such that after removal of V_i and all arcs incident on V_i there does not exist a path from one such subgraph to another. The following definition introduces these notions more formally; Figure 6 illustrates the basic idea.

Definition 4.11 *Let* $G = (V(G), A(G))$ *be a singly connected digraph. For each arc* $(V_i, V_j) \in A(G)$, *let* $G_{(V_i,V_j)} = (V(G), A(G) \setminus \{(V_i, V_j)\})$. *Now, consider vertex* $V_i \in V(G)$. *For each predecessor* $V_j \in \rho(V_i)$ *of* V_i, *let* $G^+_{(V_j,V_i)}$ *be the component of* $G_{(V_j,V_i)}$ *containing* V_j; $G^+_{(V_j,V_i)}$ *is called an* upper graph *of* V_i. *For each successor* $V_k \in \sigma(V_i)$ *of* V_i, *let* $G^-_{(V_i,V_k)}$ *be the component of* $G_{(V_i,V_k)}$ *containing* V_k; $G^-_{(V_i,V_k)}$ *is called a* lower graph *of* V_i.

Now consider computing probabilities from a belief network comprising a singly connected digraph. It will be evident that the probabilities of the values of a specific vertex in the digraph are dependent upon the evidence entered into the graph. As was true for the vertices in a directed tree, the probabilities of the values of a vertex in a singly connected digraph can be written in terms of factors describing the influence of evidence entered into the upper graphs of the vertex and the influence of evidence entered into the lower graphs of the vertex. In fact, the data fusion lemma presented in the previous section for directed trees applies to singly connected digraphs as well. Before discussing this in further detail, we redefine the compound causal and diagnostic parameters for a vertex.

Definition 4.12 *Let $B = (G, \Gamma)$ be a belief network where $G = (V(G), A(G))$ is a singly connected digraph, and let Pr be the joint probability distribution defined by B. Furthermore, let $V_i \in V(G)$ be a probabilistic variable in G and let $V_i^+ = \bigcup_{V_j \in \rho(V_i)} V(G^+_{(V_j,V_i)})$ and $V_i^- = V(G) \setminus V_i^+$. The* compound causal parameter *π_{V_i} for V_i is the function $\pi_{V_i} \colon \{v_i, \neg v_i\} \to [0,1]$ defined by*

$$\pi_{V_i}(V_i) \;=\; Pr(V_i \mid \tilde{c}_{V_i^+})$$

The compound diagnostic parameter *λ_{V_i} for V_i is the function $\lambda_{V_i} : \{v_i, \neg v_i\} \to [0,1]$ defined by*

$$\lambda_{V_i}(V_i) \;=\; Pr(\tilde{c}_{V_i^-} \mid V_i)$$

Note that this redefinition of the compound parameters differs from Definition 4.2 only with respect to the sets V_i^+ and V_i^- for a vertex $V_i \in V(G)$; the basic idea is the same. For a vertex W with an in-degree equal to zero, we once more find $\pi_W(W) = Pr(W)$; for a vertex V with an out-degree equal to zero, we have $\lambda_V(V) = 1$. The observations we made in Section 4.1 concerning instantiated variables also hold here.

The data fusion lemma, that is, Lemma 4.3 from the previous section, now is taken to apply to belief networks comprising a singly connected digraph G: for each probabilistic variable $V_i \in V(G)$, we have

$$Pr(V_i \mid \tilde{c}_{V(G)}) \;=\; \alpha \cdot \pi_{V_i}(V_i) \cdot \lambda_{V_i}(V_i)$$

where α is a normalization constant.

The two compound parameters for a vertex specify information concerning the joint probability distribution from all its descendants combined and from all its ancestors combined. Once more we observe that to be able to exploit the graphical part of a belief network as a computational architecture, these compound parameters have to be decomposed into separate causal and diagnostic parameters corresponding with the neighbours of the vertex. We redefine these separate parameters before discussing the decomposition of the compound ones.

Definition 4.13 *Let $B = (G, \Gamma)$ be a belief network as before and let Pr be the joint probability distribution defined by B. Let V_i be a probabilistic variable in G having a successor V_k. The* causal parameter *$\pi^{V_i}_{V_k}$* from *V_i* to *V_k is the function $\pi^{V_i}_{V_k} \colon \{v_i, \neg v_i\} \to [0,1]$ defined by*

$$\pi^{V_i}_{V_k}(V_i) \;=\; Pr(V_i \mid \tilde{c}_{V(G^+_{(V_i,V_k)})})$$

Now, let V_i be a probabilistic variable in G having a predecessor V_j. The diagnostic parameter *$\lambda^{V_j}_{V_i}$* from *V_i* to *V_j is the function $\lambda^{V_j}_{V_i} \colon \{v_j, \neg v_j\} \to [0,1]$ defined by*

$$\lambda^{V_j}_{V_i}(V_j) \;=\; Pr(\tilde{c}_{V(G^-_{(V_j,V_i)})} \mid V_j)$$

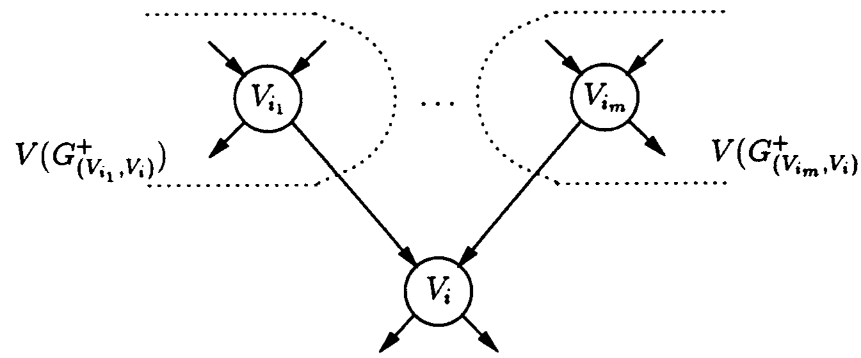

Figure 7: Exploiting d-Separation for Computing $\pi_{V_i}(V_i)$.

The separate causal and diagnostic parameters defined above can once more be viewed as associated with the arcs of the graphical part of a belief network.

The following lemma now shows how a vertex can compute its compound causal parameter from its associated conditional probability assessment function and the causal parameters it receives from each of its predecessors.

Lemma 4.14 *Let $B = (G, \Gamma)$ be a belief network as before. Let $V_i \in V(G)$ be a probabilistic variable with $\rho(V_i) = \{V_{i_1}, \ldots, V_{i_m}\}$, $m \geq 1$. Furthermore, let π_{V_i} be the compound causal parameter for V_i, and for each $V_{i_j} \in \rho(V_i)$, let $\pi_{V_i}^{V_{i_j}}$ be the causal parameter from V_{i_j} to V_i. Then,*

$$\pi_{V_i}(V_i) = \sum_{c_{\rho(V_i)}} \gamma_{V_i}(V_i \mid c_{\rho(V_i)}) \cdot \prod_{j=1,\ldots,m} \pi_{V_i}^{V_{i_j}}(c_{V_{i_j}})$$

where $c_{\rho(V_i)} = \bigwedge_{j=1,\ldots,m} c_{V_{i_j}}$.

Proof. Let Pr be the joint probability distribution defined by the belief network B. Furthermore, let $V_i^+ = \bigcup_{j=1,\ldots,m} V(G^+_{(V_{i_j},V_i)})$ as before. Then, from Definition 4.12 we have

$$\begin{aligned}\pi_{V_i}(V_i) &= Pr(V_i \mid \tilde{c}_{V_i^+}) = \\ &= Pr(V_i \mid \tilde{c}_{V(G^+_{(V_{i_1},V_i)})} \wedge \cdots \wedge \tilde{c}_{V(G^+_{(V_{i_m},V_i)})}) = \\ &= \sum_{c_{\rho(V_i)}} Pr(V_i \mid c_{\rho(V_i)} \wedge \tilde{c}_{V(G^+_{(V_{i_1},V_i)})} \wedge \cdots \wedge \tilde{c}_{V(G^+_{(V_{i_m},V_i)})}) \cdot \\ &\quad \cdot Pr(c_{\rho(V_i)} \mid \tilde{c}_{V(G^+_{(V_{i_1},V_i)})} \wedge \cdots \wedge \tilde{c}_{V(G^+_{(V_{i_m},V_i)})})\end{aligned}$$

Now consider Figure 7 showing a fragment of the singly connected digraph G. We observe that $\langle \{V_i\} | \rho(V_i) | X \rangle_G^d$ for all subsets $X \subseteq V_i^+$. Since G is an I-map for the joint probability distribution Pr, it follows that $I_{Pr}(\{V_i\}, \rho(V_i), X)$ for all $X \subseteq V_i^+$. Exploiting this observation, we find

$$\pi_{V_i}(V_i) = \sum_{c_{\rho(V_i)}} Pr(V_i \mid c_{\rho(V_i)}) \cdot Pr(c_{\rho(V_i)} \mid \tilde{c}_{V(G^+_{(V_{i_1},V_i)})} \wedge \cdots \wedge \tilde{c}_{V(G^+_{(V_{i_m},V_i)})})$$

In addition, we observe that $I_{Pr}(X, \{V_{i_j}\}, Y)$ for all subsets $X \subseteq V(G^+_{(V_{i_j}, V_i)})$, $Y \subseteq V(G^+_{(V_{i_k}, V_i)})$, $k = 1, \ldots, m$, $j = 1, \ldots, m$, $k \neq j$. Exploiting this observation, it follows that

$$\pi_{V_i}(V_i) = \sum_{c_{\rho(V_i)}} Pr(V_i \mid c_{\rho(V_i)}) \cdot Pr(c_{V_{i_1}} \mid \tilde{c}_{V(G^+_{(V_{i_1}, V_i)})}) \cdot \ldots \cdot Pr(c_{V_{i_m}} \mid \tilde{c}_{V(G^+_{(V_{i_m}, V_i)})})$$

where $c_{\rho(V_i)} = \bigwedge_{j=1,\ldots,m} c_{V_{i_j}}$. The probabilities $Pr(V_i \mid C_{\rho(V_i)})$ have been specified as function values $\gamma_{V_i}(V_i \mid C_{\rho(V_i)})$ of the conditional probability assessment function γ_{V_i} associated with vertex V_i, and hence are available to V_i. In addition, the vertex receives the probabilities $Pr(V_{i_j} \mid \tilde{c}_{V(G^+_{(V_{i_j}, V_i)})})$ as function values $\pi_{V_i}^{V_{i_j}}(V_{i_j})$ of the causal parameter $\pi_{V_i}^{V_{i_j}}$ from its predecessor V_{i_j}, $j = 1, \ldots, m$. Substitution yields

$$\begin{aligned}\pi_{V_i}(V_i) &= \sum_{c_{\rho(V_i)}} \gamma_{V_i}(V_i \mid c_{\rho(V_i)}) \cdot \pi_{V_i}^{V_{i_1}}(c_{V_{i_1}}) \cdot \ldots \cdot \pi_{V_i}^{V_{i_m}}(c_{V_{i_m}}) = \\ &= \sum_{c_{\rho(V_i)}} \gamma_{V_i}(V_i \mid c_{\rho(V_i)}) \cdot \prod_{j=1,\ldots,m} \pi_{V_i}^{V_{i_j}}(c_{V_{i_j}})\end{aligned}$$

where $c_{\rho(V_i)} = \bigwedge_{j=1,\ldots,m} c_{V_{i_j}}$. □

Equally, a vertex can compute its compound diagnostic parameter from the separate diagnostic parameters it receives from its successors. This property is similar to Lemma 4.6 from the previous section, that is, for an uninstantiated probabilistic variable $V_i \in V(G)$ with $\sigma(V_i) = \{V_{i_1}, \ldots, V_{i_m}\}$, $m \geq 1$, we have

$$\lambda_{V_i}(V_i) = \prod_{j=1,\ldots,m} \lambda_{V_{i_j}}^{V_i}(V_i)$$

This property can be taken to apply to instantiated variables also as suggested in Section 4.1.

A vertex in turn has to compute the proper parameters to send to its neighbours. The following lemma states how a vertex can compute the diagnostic parameter to send to a predecessor from its own conditional probability assessment function, its compound diagnostic parameter and the causal parameters it receives from its other predecessors.

Lemma 4.15 *Let $B = (G, \Gamma)$ be a belief network as before. Let $V_i \in V(G)$ be a probabilistic variable with $\rho(V_i) = \{V_{i_1}, \ldots, V_{i_m}\}$, $m \geq 1$. Let λ_{V_i} be the compound diagnostic parameter for V_i. Furthermore, let $\lambda_{V_i}^{V_{i_j}}$ be the diagnostic parameter from V_i to V_{i_j}, and let $\pi_{V_i}^{V_{i_j}}$ be the causal parameter from V_{i_j} to V_i, $j = 1, \ldots, m$. Then,*

$$\lambda_{V_i}^{V_{i_j}}(V_{i_j}) = \alpha \cdot \sum_{c_{V_i}} \lambda_{V_i}(c_{V_i}) \cdot \left[\sum_{c_{\rho(V_i) \setminus \{V_{i_j}\}}} \gamma_{V_i}(c_{V_i} \mid c_{\rho(V_i) \setminus \{V_{i_j}\}} \wedge V_{i_j}) \cdot \prod_{k=1,\ldots,m, k \neq j} \pi_{V_i}^{V_{i_k}}(c_{V_{i_k}}) \right]$$

where $c_{\rho(V_i) \setminus \{V_{i_j}\}} = \bigwedge_{k=1,\ldots,m,k\neq j} c_{V_{i_k}}$ and α is a normalization constant.

Proof. Let Pr be the joint probability distribution defined by the belief network B. Then, from Definition 4.13 we have

$$\lambda_{V_i}^{V_{i_j}}(V_{i_j}) = Pr(\tilde{c}_{V(G^-_{(V_{i_j},V_i)})} \mid V_{i_j})$$

Now consider Figure 8 showing a fragment of the singly connected digraph G; observe that $\tilde{c}_{V(G^-_{(V_{i_j},V_i)})} = \tilde{c}_{V_i^-} \wedge (\bigwedge_{k=1,\ldots,m,k\neq j} \tilde{c}_{V(G^+_{(V_{i_k},V_i)})})$. It follows that

$$\begin{aligned}
\lambda_{V_i}^{V_{i_j}}(V_{i_j}) &= Pr(\tilde{c}_{V_i^-} \wedge (\textstyle\bigwedge_{k=1,\ldots,m,k\neq j} \tilde{c}_{V(G^+_{(V_{i_k},V_i)})}) \mid V_{i_j}) = \\
&= \textstyle\sum_{c_{V_i}} \sum_{c_{\rho(V_i)\setminus\{V_{i_j}\}}} Pr(\tilde{c}_{V_i^-} \wedge (\bigwedge_{k=1,\ldots,m,k\neq j} \tilde{c}_{V(G^+_{(V_{i_k},V_i)})}) | c_{V_i} \wedge c_{\rho(V_i)\setminus\{V_{i_j}\}} \wedge V_{i_j}) \cdot \\
&\quad \cdot Pr(c_{V_i} \wedge c_{\rho(V_i)\setminus\{V_{i_j}\}} \mid V_{i_j})
\end{aligned}$$

Since G is an I-map for the joint probability distribution Pr, we may exploit the independencies read from G. We find

$$\begin{aligned}
\lambda_{V_i}^{V_{i_j}}(V_{i_j}) &= \textstyle\sum_{c_{V_i}} \sum_{c_{\rho(V_i)\setminus\{V_{i_j}\}}} Pr(\tilde{c}_{V_i^-} | c_{V_i}) \cdot Pr(\bigwedge_{k=1,\ldots,m,k\neq j} \tilde{c}_{V(G^+_{(V_{i_k},V_i)})} | c_{\rho(V_i)\setminus\{V_{i_j}\}}) \cdot \\
&\quad \cdot Pr(c_{V_i} \wedge c_{\rho(V_i)\setminus\{V_{i_j}\}} \mid V_{i_j}) = \\
&= \textstyle\sum_{c_{V_i}} Pr(\tilde{c}_{V_i^-} \mid c_{V_i}) \cdot \Big[\sum_{c_{\rho(V_i)\setminus\{V_{i_j}\}}} Pr(c_{V_i} \mid c_{\rho(V_i)\setminus\{V_{i_j}\}} \wedge V_{i_j}) \cdot \\
&\quad \cdot Pr(c_{\rho(V_i)\setminus\{V_{i_j}\}} \mid V_{i_j}) \cdot Pr(\textstyle\bigwedge_{k=1,\ldots,m,k\neq j} \tilde{c}_{V(G^+_{(V_{i_k},V_i)})} \mid c_{\rho(V_i)\setminus\{V_{i_j}\}})\Big] = \\
&= \alpha \cdot \textstyle\sum_{c_{V_i}} Pr(\tilde{c}_{V_i^-} \mid c_{V_i}) \cdot \Big[\sum_{c_{\rho(V_i)\setminus\{V_{i_j}\}}} Pr(c_{V_i} \mid c_{\rho(V_i)\setminus\{V_{i_j}\}} \wedge V_{i_j}) \cdot \\
&\quad \cdot Pr(c_{\rho(V_i)\setminus\{V_{i_j}\}} \mid \textstyle\bigwedge_{k=1,\ldots,m,k\neq j} \tilde{c}_{V(G^+_{(V_{i_k},V_i)})})\Big]
\end{aligned}$$

using Bayes's Rule for the last equality; α is a normalization constant. Exploiting the independencies portrayed by G, we find

$$\begin{aligned}
\lambda_{V_i}^{V_{i_j}}(V_{i_j}) &= \alpha \cdot \textstyle\sum_{c_{V_i}} Pr(\tilde{c}_{V_i^-} \mid c_{V_i}) \cdot \Big[\sum_{c_{\rho(V_i)\setminus\{V_{i_j}\}}} Pr(c_{V_i} | c_{\rho(V_i)\setminus\{V_{i_j}\}} \wedge V_{i_j}) \cdot \\
&\quad \cdot \textstyle\prod_{k=1,\ldots,m,k\neq j} Pr(c_{V_{i_k}} \mid \tilde{c}_{V(G^+_{(V_{i_k},V_i)})})\Big]
\end{aligned}$$

The probabilities $Pr(\tilde{c}_{V_i^-} \mid V_i)$ equal the function values $\lambda_{V_i}(V_i)$ of the compound diagnostic parameter λ_{V_i} for vertex V_i. Furthermore, the probabilities $Pr(V_i \mid C_{\rho(V_i)\setminus\{V_{i_j}\}} \wedge V_{i_j}) =$

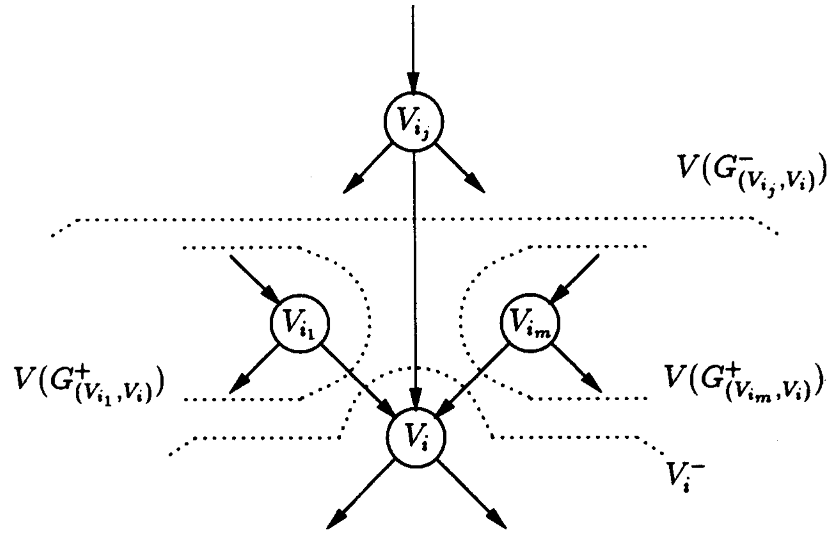

Figure 8: Exploiting d-Separation for Computing $\lambda_{V_i}^{V_{i_j}}(V_{i_j})$.

$Pr(V_i \mid C_{\rho(V_i)})$ have been specified as function values $\gamma_{V_i}(V_i \mid C_{\rho(V_i)})$ of the conditional probability assessment function γ_{V_i} associated with V_i and therefore are known to V_i. To conclude, we note that vertex V_i receives the probabilities $Pr(V_{i_k} \mid \tilde{c}_{V(G^+_{(V_{i_k},V_i)})})$ as function values $\pi_{V_i}^{V_{i_k}}(V_{i_k})$ of the causal parameter $\pi_{V_i}^{V_{i_k}}$ from its predecessor V_{i_k}, $k = 1,\ldots,m, k \neq j$. Substitution yields

$$\lambda_{V_i}^{V_{i_j}}(V_{i_j}) = \alpha \cdot \sum_{c_{V_i}} \lambda_{V_i}(c_{V_i}) \cdot \left[\sum_{c_{\rho(V_i)\setminus\{V_{i_j}\}}} \gamma_{V_i}(c_{V_i} \mid c_{\rho(V_i)\setminus\{V_{i_j}\}} \wedge V_{i_j}) \cdot \prod_{k=1,\ldots,m,k\neq j} \pi_{V_i}^{V_{i_k}}(c_{V_{i_k}}) \right]$$

where $c_{\rho(V_i)\setminus\{V_{i_j}\}} = \bigwedge_{k=1,\ldots,m,k\neq j} c_{V_{i_k}}$ and α is a normalization constant. □

Equally, the causal parameter a vertex has to send to a successor can be computed from its compound causal parameter and the diagnostic parameters it receives from its other successors. This property is similar to Lemma 4.8 from the previous section, that is, for an uninstantiated probabilistic variable $V_i \in V(G)$ with $\sigma(V_i) = \{V_{i_1},\ldots,V_{i_m}\}$, $m \geq 1$, we have

$$\pi_{V_{i_j}}^{V_i}(V_i) \;=\; \alpha \cdot \pi_{V_i}(V_i) \cdot \prod\nolimits_{k=1,\ldots,m,k\neq j} \lambda_{V_{i_k}}^{V_i}(V_i)$$

where α is a normalization constant. Again, this property can be taken to apply to instantiated variables as well.

5 Conclusion

The belief network framework provides a formalism for representing knowledge concerning a joint probability distribution on a problem domain and a set of algorithms for manipulating the knowledge represented. Both aspects of the framework have been covered by this paper. The belief network formalism has been introduced and has subsequently been taken to constitute a computational architecture for reasoning with a belief network. We have presented an in-depth discussion of the set of algorithms for efficiently propagating evidence and computing probabilities of interest proposed by J. Pearl, including full proofs. The rigorousness of the approach provides a point of departure for further investigation of these algorithms.

References

[Andreassen et al., 1987] S. Andreassen, M. Woldbye, B. Falck, S.K. Andersen (1987). MUNIN - A causal probabilistic network for interpretation of electromyographic findings, *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pp. 366 - 372.

[Andreassen et al., 1991] S. Andreassen, R. Hovorka, J. Benn, K.G. Olesen, E.R. Carson (1991). A model-based approach to insulin adjustment, in: M. Stefanelli, A. Hasman, M. Fieschi (eds.), *Proceedings of the Third Conference on Artificial Intelligence in Medicine*, Lecture Notes in Medical Informatics 44, Springer-Verlag, Berlin, pp. 239 - 248.

[Bellazzi et al., 1991] R. Bellazzi, C. Berzuini, S. Quaglini, D. Spiegelhalter, M. Leaning (1991). Cytotoxic chemotherapy monitoring using stochastic simulation on graphical models, in: M. Stefanelli, A. Hasman, M. Fieschi (eds.), *Proceedings of the Third Conference on Artificial Intelligence in Medicine*, Lecture Notes in Medical Informatics 44, Springer-Verlag, Berlin, pp. 227 - 238.

[Bruza & van der Gaag, 1992] P.D. Bruza, L.C. van der Gaag (1992). *Index Expression Belief Networks for Information Disclosure*, Technical Report RUU-CS-92-21, Utrecht University.

[Cooper, 1990] G.F. Cooper (1990). The computational complexity of probabilistic inference using Bayesian belief networks, *Artificial Intelligence*, vol. 42, pp. 393 - 405.

[Gaag, 1990] L.C. van der Gaag (1990). *Probability-Based Models for Plausible Reasoning*, Ph.D. thesis, University of Amsterdam.

[Harary, 1969] F. Harary (1969). *Graph Theory*, Addison-Wesley Publishing Company, Reading Massachusetts.

[Jensen et al., 1990] F.V. Jensen, J. Nielsen, H.I. Christensen (1990). *Use of Causal Probabilistic Networks as High Level Models in Computer Vision*, Technical Report R-90-39, University of Aalborg.

[Lauritzen & Spiegelhalter, 1988] S.L. Lauritzen, D.J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society, Series B*, vol. 50, pp. 157 – 224.

[Morawski, 1989] P. Morawski (1989). Understanding Bayesian belief networks, *AI Expert*, may, pp. 44 - 48.

[Pearl, 1988] J. Pearl (1988). *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*, Morgan Kaufmann, Palo Alto.

[Shwe et al., 1991] M.A. Shwe, B. Middleton, D.E. Heckerman, M. Henrion, E.J. Horvitz, H.P. Lehmann, G.F. Cooper (1991). Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base. The probabilistic model and inference algorithms, *Methods of Information in Medicine*, vol. 30, pp. 241 – 255.