Centrum voor Wiskunde en Informatica

Centre for Mathematics and Computer Science

L.C. van der Gaag

Belief networks in plausible reasoning

Computer Science/Department of Software Technology Report CS-R8927 July

1989

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

L.C. van der Gaag

Belief networks in plausible reasoning

Computer Science/Department of Software Technology Report CS-R8927 July

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (N.W.O.).

Belief Networks in Plausible Reasoning

L.C. van der Gaag
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

After the introduction of the (theoretically) disappointing quasi-probabilistic models for handling uncertainty in knowledge-based systems, the attention for probability theory as a mathematical foundation for dealing with uncertain information in an artificial intelligence context diminished. Recently however, a new probabilistic trend in reasoning with uncertainty in knowledge-based systems is discernable: several models have been proposed departing from so-called belief networks. Informally speaking, a belief network is a representation of a problem domain, consisting of a graphical model of the statistical variables and their causal interrelationships, and an associated joint probability distribution. In this paper we introduce the notion of a belief network in general and discuss several schemes for updating the joint probability distribution of the network as evidence becomes available.

1980 Mathematics Subject Classification (1985): 05Cxx, 06Exx, 60Bxx, 68Txx.
1987 CR Categories: I.2.3. [**Deduction and Theorem Proving**]: uncertainty, probabilistic reasoning; I.2.5. [**Programming Languages and Software**]: expert system tools and techniques.
Key Words & Phrases: probability theory, belief networks, knowledge-based systems, plausible reasoning.

1. Introduction

In the seventies several quasi-probabilistic models for dealing with uncertain information in rule-based expert systems were proposed, such as the well-known certainty factor model. The incorrectness of these models from a mathematical point of view and an analysis of the problems the researchers were confronted with, led to a world-wide discussion concerning the appropriateness of probability theory for handling uncertainty in a knowledge-based context, [1,2]. Although this discussion has not yet subdued, in the mid-eighties a new trend in probabilistic reasoning in knowledge-based systems is discernable: several (mathematically correct) probabilistic models for handling uncertainty have been proposed, each departing from a graphical representation of a knowledge base, e.g. [3,4,5]. Hereafter such a graphical representation will be called a *belief network*. Informally speaking, a belief network is a map of the statistical variables discerned in the problem domain and their causal interrelationships. The causal relationships between the statistical variables are quantified by means of 'local' probabilities, together defining a joint probability distribution. The belief network serves as an architecture for performing certain local probabilistic computations during an actual consultation of the knowledge-based system. The phrase belief network has been adopted from J. Pearl, [4]. Several other phrases are used to denote the same concept: D.J. Spiegelhalter uses the phrase *causal graph* [3], and the phrase *influence diagram* is used by R.D. Shachter [5]. Statisticians often use the phrase *recursive model* to denote similar graphical representations of a problem domain, see for example [6,7].

As reasoning with uncertain information is concerned, not only researchers who have chosen probability theory as the mathematical foundation have turned their attention to belief networks; researchers departing from Dempster-Shafer theory have begun to investigate the use of graphical representations of a problem domain as well, see for example [8,9].

Report CS-R8927
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

This paper is a theoretical introduction to belief networks. In Section 2 some preliminaries are provided. Section 3 discusses the representation of a problem domain in a belief network in general. In literature, several schemes have been proposed for updating the joint probability distribution of a belief network as evidence becomes available, in which the network serves as a 'computational architecture'; in Section 4 two such schemes are discussed.

2. Preliminaries

In this section we introduce several notions from probability theory and graph theory, that will play a central role in the remainder of this paper.

2.1. Probability Theory

In an expert system, knowledge concerning the problem domain usually is represented in a special knowledge representation formalism such as production rules or frames. In this paper we do not consider these knowledge representation schemes nor do we discuss the reasoning methods that have been associated with these formalisms. Here, we assume that knowledge is simply represented in the form of statistical variables and their probabilistic interrelationships. We assume that these variables can only take one of two states at a time, thus allowing to view them as logical, propositional variables.

In the following definition the notion of a Boolean algebra of propositions is introduced.

Definition 2.1. *Let $\mathscr{A}$ denote a finite set of atomic propositions: $\mathscr{A} = \{a_1, \ldots, a_n\}$, $n \geqslant 1$. Let $\mathscr{B}$ be the free Boolean algebra generated by $\mathscr{A}$, that is,*

(1) *for all $x \in \mathscr{A}$, $x \in \mathscr{B}$,*
(2) *for all $x_1, x_2 \in \mathscr{B}$, $x_1 \wedge x_2 \in \mathscr{B}$ and $x_1 \vee x_2 \in \mathscr{B}$, and*
(3) *for all $x \in \mathscr{B}$, $\neg x \in \mathscr{B}$.*

$\mathscr{B}$ is called the set of Boolean combinations of atomic propositions.

On $\mathscr{B}$ we define a partial order $\preccurlyeq$: *for any $x_1, x_2 \in \mathscr{B}$, we say $x_1 \preccurlyeq x_2$ if $x_2 = x_1 \vee x_2$ or (equivalently) $x_1 = x_1 \wedge x_2$.*

According to the convention in logic we denote the universal lower bound in the algebra $\mathscr{B}$ by false *and the universal upper bound by* true.

Since the set $\mathscr{B}$ of Boolean combinations of atomic propositions is a Boolean algebra we have equality according to logical truth tables. It will be obvious that a universal lower bound and upper bound exist. Furthermore, since $\mathscr{B}$ is a free Boolean algebra we have that the atomic propositions $a_i \in \mathscr{A}$ are algebraically independent, meaning that each of the 2^n conjunctions of the form $\bigwedge_{i=1}^{n} A_i$, where for $i = 1, \ldots, n$, either $A_i = a_i$ or $A_i = \neg a_i$, is different from *false*. In the sequel, we will use lowercase symbols to denote elements of $\mathscr{B}$. An uppercase symbol A_i is used to denote a logical variable taking one of the values a_i and $\neg a_i$. In the sequel, we will often view the Boolean algebra $\mathscr{B}$ as being 'spanned' by a set of logical variables.

In the following definition we introduce the notion of a probability distribution on our Boolean algebra $\mathscr{B}$. For a more general introduction, any introductory textbook on probability theory will suffice.

Definition 2.2. *Let $\mathscr{B}$ be the Boolean algebra defined as above. Let Pr be a function $Pr: \mathscr{B} \to [0,1]$ such that*

(1) *Pr is positive, that is, for all $x \in \mathscr{B}$, $Pr(x) \geqslant 0$, and furthermore $Pr(false) = 0$,*
(2) *Pr is normed, that is, $Pr(true) = 1$, and*
(3) *Pr is additive, that is, for all $x_1, x_2 \in \mathscr{B}$, 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 probability distribution *on* $\mathscr{B}$. *The pair* $(\mathscr{B},Pr)$ *is called a* probability algebra.

In probability theory we are used to associate probabilities with sets instead of logical propositions. However, 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. A probability distribution Pr on the Boolean algebra defined above therefore has the usual properties.

When $\mathscr{B}$ is viewed as being spanned by a set $A = \{A_1, \ldots, A_n\}$ of logical variables, we will speak of a *joint probability distribution*; a joint probability distribution on the algebra $\mathscr{B}' \subseteq \mathscr{B}$ spanned by $A' \subseteq A$ will be called a *marginal distribution*.

The notion of a conditional probability distribution on the Boolean algebra $\mathscr{B}$ is defined in the following definition.

DEFINITION 2.3. *Let* $(\mathscr{B},Pr)$ *be the probability algebra from Definition 2.2. For each* $x, y \in \mathscr{B}$ *with* $Pr(y) > 0$, *the* conditional probability *of x given y, notation:* $Pr(x \mid y)$, *is defined as*

$$Pr(x \mid y) = \frac{Pr(x \wedge y)}{Pr(y)}.$$

It can easily be proven that the conditional probabilities given a specific element $y \in \mathscr{B}$ different from *false* define a probability distribution, i.e. the axioms (1), (2) and (3) from Definition 2.2 hold.

We conclude this subsection with two well-known theorems.

THEOREM 2.4. *Let* $(\mathscr{B},Pr)$ *be the probability algebra from Definition 2.2. Let* $x_i \in \mathscr{B}$ *for* $i = 1, \ldots, n$. *Then,* $Pr(x_1 \wedge \cdots \wedge x_n) = Pr(x_n \mid x_1 \wedge \cdots \wedge x_{n-1}) \cdots Pr(x_2 \mid x_1) \cdot Pr(x_1)$.

Theorem 2.4 is called the *chain rule*. The following theorem is known as *Bayes' Theorem*.

THEOREM 2.5. *Let* $(\mathscr{B},Pr)$ *be the probability algebra from Definition 2.2. Let* $x,y \in \mathscr{B}$ *such that* $Pr(x) > 0$ *and* $Pr(y) > 0$. *Then,*

$$Pr(x \mid y) = \frac{Pr(y \mid x)Pr(x)}{Pr(y)}$$

2.2. *Graph Theory*

In this section we review some basic notions from graph theory. For further information on graph theory the reader is referred to [10,11].

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

DEFINITION 2.6. *An* undirected graph *G is an ordered pair* $G = (V(G),E(G))$, *where* $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$, *is a finite set of* vertices *and* $E(G)$ *is a set of unordered pairs* (V_i,V_j), $V_i \neq V_j$, $V_i,V_j \in V(G)$, *called* edges. *Two vertices* V_i *and* V_j *are called* adjacent *or* neighbouring *vertices in G if* $(V_i,V_j) \in E(G)$. *The set of all neighbours of vertex* V_i *in G is denoted by* $\nu_G(V_i)$.

A directed graph *(or* digraph, *for short) G is an ordered pair* $G = (V(G),A(G))$, *where* $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$, *is a finite set of vertices and* $A(G)$ *is a set of ordered pairs* (V_i,V_j), $V_i \neq V_j$, $V_i,V_j \in V(G)$, *called* arcs. *Vertex* V_j *is called a* successor *of vertex* V_i *if there is an arc* $(V_i,V_j) \in A(G)$. *The set of all successors of vertex* V_i *in the digraph G is denoted as* $\sigma_G(V_i)$. *Similarly, vertex* V_i *is called a* predecessor *of vertex* V_j *if there is an arc* $(V_i,V_j) \in A(G)$. *The set of all predecessors of vertex* V_i *is denoted by* $\rho_G(V_i)$. *The set of all* neighbours *of vertex* V_i *is defined as* $\nu_G(V_i) = \sigma_G(V_i) \cup \rho_G(V_i)$.

DEFINITION 2.7. *Let $G = (V(G),A(G))$ be a digraph with vertices $V(G) = \{V_1, \ldots, V_n\}$, $n \geqslant 1$. G is called a* simple digraph *if for each $V_i \in V(G)$, $V_i \notin \sigma_G(\sigma_G(V_i))$.*

Let $G = (V(G),A(G))$ be a simple digraph. The underlying graph *G' of G is the undirected graph $G' = (V(G'),E(G'))$ where $V(G') = V(G)$ and $E(G')$ is obtained from $A(G)$ by replacing each arc $(V_i,V_j) \in A(G)$ by its corresponding edge (V_i,V_j).*

In the following Definition 2.8 some notions are introduced concerning undirected graphs. These notions however can easily be extended to directed graphs by taking the directions of the arcs into account.

DEFINITION 2.8. *Let $G = (V(G),E(G))$ be an undirected graph. A* path *of length k from vertex V_0 to vertex V_k is a sequence of vertices $V_0,V_1,\ldots,V_k$ in $V(G)$ such that $(V_{i-1},V_i) \in E(G)$, $i = 1,\ldots,k$. A* cycle *is a path $V_0,V_1,\ldots,V_k$ with $V_0 = V_k$. An* elementary cycle *is a cycle $V_0,V_1,\ldots,V_k = V_0$ such that $V_i \neq V_j$, $i \neq j$, $i,j = 1,\ldots,k$, that is, an elementary cycle is a cycle in which each node except V_0 appears only once. A* chord *or* shortcut *of an elementary cycle $V_0,V_1,\ldots,V_k = V_0$ is an edge (V_i,V_j), $i \neq j \pm 1 (\mathrm{mod}\ k+1)$.*

If for each pair of distinct vertices $V_i,V_j \in V(G)$ there is a path from V_i to V_j in G, then G is called a connected *graph; otherwise G is* disconnected. *G is called a* cyclic *graph if it contains at least one cycle; a graph without any cycles is called* acyclic.

We conclude this section with two more definitions.

DEFINITION 2.9. *Let $G = (V(G),E(G))$ be an undirected graph with vertices $V(G) = \{V_1,\ldots,V_n\}$, $n \geqslant 1$. A graph $G' = (V(G'),E(G'))$ is a* subgraph *of G if $V(G') \subseteq V(G)$ and $E(G') \subseteq E(G)$. A subgraph $G' = (V(G'),E(G'))$ is a* full *subgraph of G if for each $(V_i,V_j) \in E(G)$ such that $V_i,V_j \in V(G')$ we have $(V_i,V_j) \in E(G')$. We say that the full subgraph G' is* induced *by $V(G')$.*

The order *of G is the number of vertices in G, i.e. $|V(G)|$. The size of G is the number of edges in G, i.e. $|E(G)|$. G is a* complete *n-graph if it has order n and size $\binom{n}{2}$, that is, a graph is complete if there exists an edge between each pair of nodes.*

A clique *in G is a full subgraph $G' = (V(G'),E(G'))$ of G such that G' is complete. G' is called a* maximal clique *if there does not exist a clique G'' in G with $G'' \neq G'$, such that G' is a full subgraph of G''.*

In this paper, we will use the word clique to mean a maximal clique.

DEFINITION 2.10. *A* tree *is an undirected graph $T = (V(T),E(T))$ which is connected and has no cycles.*

Let $G = (V(G),E(G))$ be an undirected connected graph. Let $T = (V(T),E(T))$ with $V(T) = V(G)$ and $E(T) \subseteq E(G)$ be a subgraph of G such that T is a tree. Then, T is called a spanning tree *of G.*

3. KNOWLEGE REPRESENTATION IN A BELIEF NETWORK

In Section 2.1 we have remarked that in this paper we do not depart from the knowledge representation schemes generally employed in knowledge-based systems; we assume that knowledge is represented in statistical variables (taken as propositional variables) and their causal interrelationships. In this section we discuss this representation scheme in further detail. We introduce the notion of a belief network informally before giving a formal definition.

Belief networks provide a formalism for representing a problem domain. A belief network comprises two parts: a *qualitative representation* of the problem domain and an associated *quantitative*

representation. The qualitative part takes the form of an acyclic directed graph $G = (V(G), A(G))$ with vertices $V(G) = \{V_1, \ldots, V_n\}$ and arcs $A(G)$. Each vertex V_i in $V(G)$ represents a statistical variable that can take one of a set of values. In this paper we assume that the statistical variables can take only one of the truth values *true* or *false*. We will adhere to the following notational convention: the variable V_i taking the truth value *true* will be denoted by v_i; V_i = *false* will be denoted by $\neg v_i$. We take an arc $(V_i, V_j) \in A(G)$ to represent a direct 'causal' relationship between the variables V_i and V_j: the arc (V_i, V_j) is interpreted as stating that 'V_i directly causes V_j'. 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 human judgment; hence the phrase *belief* network. Note that indirect influences can be read from the figure.

Associated with the digraph is a numerical assessment of the strengths of the represented relationships: with each vertex is associated a set of (conditional) probabilities describing the influence of the values of the predecessors of the vertex, on the value of the vertex itself. So, with a vertex V_i with m predecessors are associated 2^m (conditional) probabilities of the form $P(v_i \mid c)$ for all possible combinations c of values of the predecessors of V_i in the digraph. These 'local' probabilities are assumed to contain all information necessary for defining a unique joint probability distribution on the variables.

The following example illustrates these notions. Whenever possible, this example will be used throughout this paper as the running example; it has been taken from Lauritzen and Spiegelhalter, [12].

EXAMPLE 3.1. Consider the acyclic digraph G shown in Figure 3.1. We assume that this graph has been configured by an expert who for example observed that the value of the variable V_2 is only dependent directly upon the value of the variable V_1.

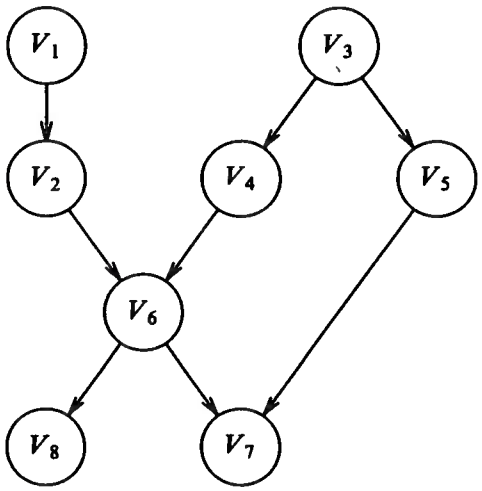

FIGURE 3.1. An acyclic digraph G.

From the figure we for example read that the value of the variable V_1 does not influence the value of the variable V_6 directly, but only through the value of V_2.

Corresponding with this digraph G the expert has assessed eighteen probabilities:

$Pr(v_1)$

$Pr(v_2 \mid v_1)$ and $Pr(v_2 \mid \neg v_1)$

$Pr(v_3)$

$Pr(v_4 \mid v_3)$ and $Pr(v_4 \mid \neg v_3)$

$Pr(v_5 \mid v_3)$ and $Pr(v_5 \mid \neg v_3)$

$Pr(v_6 \mid v_2 \wedge v_4)$, $Pr(v_6 \mid v_2 \wedge \neg v_4)$, $Pr(v_6 \mid \neg v_2 \wedge v_4)$ and $Pr(v_6 \mid \neg v_2 \wedge \neg v_4)$

$Pr(v_7 \mid v_5 \wedge v_6)$, $Pr(v_7 \mid v_5 \wedge \neg v_6)$, $Pr(v_7 \mid \neg v_5 \wedge v_6)$ and $Pr(v_7 \mid \neg v_5 \wedge \neg v_6)$

$Pr(v_8 \mid v_6)$ and $Pr(v_8 \mid \neg v_6)$

Note that from these probabilities we can uniquely compute the complementary probabilities; for example $Pr(\neg v_4 \mid v_3)$ is computed using $Pr(\neg v_4 \mid v_3) = 1 - Pr(v_4 \mid v_3)$.

The graph is assumed to represent all dependencies between the statistical variables, so that the joint probability distribution $Pr(V_1 \wedge \cdots \wedge V_8)$ can be expressed as the product

$$Pr(V_1 \wedge \cdots \wedge V_8) = Pr(V_8 \mid V_1 \wedge \cdots \wedge V_7) \cdot Pr(V_7 \mid V_1 \wedge \cdots \wedge V_6) \cdot Pr(V_6 \mid V_1 \wedge \cdots \wedge V_5) \cdot \cdots \cdot Pr(V_1) =$$
$$= Pr(V_8 \mid V_6) \cdot Pr(V_7 \mid V_5 \wedge V_6) \cdot Pr(V_6 \mid V_2 \wedge V_4) \cdot Pr(V_5 \mid V_3) \cdot Pr(V_4 \mid V_3) \cdot Pr(V_3) \cdot Pr(V_2 \mid V_1) \cdot Pr(V_1).$$

This equation is a kind of *template*: an actual probability can be obtained by filling in values for the statistical variables V_1 through V_8. So, for example $Pr(v_1 \wedge \cdots \wedge v_8)$ may be obtained by substituting the value v_i for each variable V_i, and then computing the resulting product on the right-hand side from the initially assessed probabilities. Note that only eighteen probabilities suffice for specifying a joint probability distribution over eight variables. ■

Definition 3.3 provides a formal definition of a belief network. Some preliminary notions are introduced in Definition 3.2.

DEFINITION 3.2. *Let* $G = (V(G), A(G))$ *be an acyclic directed graph with vertices* $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$, *and arcs* $A(G)$. *Let each* $V_i \in V(G)$ *have associated a* state space $\Omega_{V_i} = \{v_i, \neg v_i\}$. *A conjunction* $c_{V'} = \bigwedge_{V_i \in V'} c_{V_i}$, $c_{V_i} \in \Omega_{V_i}$, $V' \subseteq V(G)$, *is called a* configuration *of* V'. *In the sequel, we use* $C_{V'}$ *to denote the conjunction* $C_{V'} = \bigwedge_{V_i \in V'} V_i$; $C_{V'}$ *is called a* configuration template.

Note that from a configuration template $C_{V'}$ we may obtain any actual configuration $c_{V'}$ of V' by filling in an appropriate value from the state space for each variable mentioned in the template.

DEFINITION 3.3. *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 arcs* $A(G)$, *representing directed conditional dependency relationships, and*

(2) $\Gamma = \{\gamma_{V_i} \mid V_i \in V(G)\}$ *is a set of* conditional probability vectors γ_{V_i}, *where each* γ_{V_i} *is a real-valued vector with* $2^{|\rho(V_i)|+1}$ *components* $\gamma_{V_i}(v_i \mid c_{\rho(V_i)})$ *and* $\gamma_{V_i}(\neg v_i \mid c_{\rho(V_i)})$, *such that* $\gamma_{V_i}(v_i \mid c_{\rho(V_i)})$, $\gamma_{V_i}(\neg v_i \mid c_{\rho(V_i)}) \geq 0$, *and* $\gamma_{V_i}(\neg v_i \mid c_{\rho(V_i)}) = 1 - \gamma_{V_i}(v_i \mid c_{\rho(V_i)})$, *for each* $V_i \in V(G)$ *and each configuration* $c_{\rho(V_i)}$ *of* $\rho(V_i)$.

The following lemma states that under the independency assumptions represented in the graphical part of the belief network, the initially assessed conditional probability vectors define a unique joint probability distribution.

LEMMA 3.4. *Let $B = (G, \Gamma)$ be a belief network defined as above, where G is an acyclic digraph with n vertices, $n \geqslant 1$. Then,*

$$Pr(V_1 \wedge \cdots \wedge V_n) = \prod_{i=1,\ldots,n} \gamma_{V_i}(V_i \mid C_{p(V_i)})$$

constitutes a joint probability distribution on $V(G)$.

PROOF. A digraph without directed cycles allows at least one 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 statistical 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 conditional independency relationships represented in G can be exploited. By taking $Pr(V_i \mid C_{p(V_i)}) = \gamma_{V_i}(V_i \mid C_{p(V_i)})$ for each $V_i \in V(G)$, the property stated in the lemma follows immediately. For further details, see [7]. ■

Note that the representation of uncertainty in factors which are local to expressions giving a qualitative description of the domain, closely resembles the approach followed in the quasi-probabilistic models for dealing with uncertainty in rule-based systems, in which the production rules constitute the qualitative representation of the domain.

4. EVIDENCE PROPAGATION IN A BELIEF NETWORK

In the preceding section we have introduced the notion of a belief network, consisting of a qualitative representation of the problem domain and a quantitative one. We have seen that the quantitative part of such a belief network is a representation of a joint probability distribution in terms of 'local' factors: the initially given conditional probability vectors describe the joint probability distribution locally for each vertex and its predecessors in the digraph, that is, locally in terms of the qualitative representation of the problem domain.

Since a joint probability distribution on the variables is uniquely defined by the conditional probability vectors, the impact of a value of a specific variable becoming known, on each of the other variables can be computed from these local factors. Calculation of an updated probability from the joint probability distribution in a straightforward manner however, will generally not be restricted to performing computations which are local in terms of the causal graph. In literature therefore, several less naive schemes for updating a joint probability distribution as evidence becomes available have been proposed. Although all schemes proposed for evidence propagation are based on (Bayesian) probability theory, they differ considerably in concept and in computational efficiency. It should be noted that in general probabilistic inference in belief networks without any restrictions is NP-hard, [13].

R.D. Shachter has presented a method for propagating the impact of a specified set of observed variables to a set of variables of interest. The general idea of his method is to eliminate vertices from the original graph without changing the (updated) joint probability distribution; the topology of the graph is modified using a sequence of arc reversals, and vertex removals and additions, [5]. For each subsequent propagation of evidence again such a sequence of graph modifications has to be performed. The problem of optimizing a sequence of graph modifications has been further investigated, see for example [14].

J. Pearl and J.H. Kim in [15] depart from a singly connected digraph, which is a restricted type of acyclic digraph. Their method for propagating evidence leaves the original graphical representation of the problem domain unchanged. Updating the joint probability distribution essentially entails each variable, that is, each vertex, updating the joint probability distribution locally from messages it receives from its neighbours in the digraph, that is, from its predecessors as well as its successors, and then in turn sending new, updated messages to them. In his later work [4], Pearl proposes additional schemes for coping with (undirected) cycles.

D.J. Spiegelhalter and S.L. Lauritzen have presented another, elegant scheme for evidence

propagation, [3,12]. They have observed that updating the joint probability distribution as evidence becomes available will entail going against the 'directed' conditional probabilities. They concluded that the directed graphical representation of a belief network is not suitable as an architecture for propagating evidence. This observation amongst other ones motivated an initial transformation of the belief network into an undirected graphical and probabilistic representation of the problem domain. This new representation allows for an efficient scheme for evidence propagation in which the computations to be performed are local to small sets of variables. For this purpose, Spiegelhalter and Lauritzen make use of the existing statistical theory of Markov random fields, [16].

In this paper we do not discuss Shachter's work. The work of Pearl and Kim will be dealt with briefly in Subsection 4.1. The work of Spiegelhalter and Lauritzen will be treated in some detail in Subsection 4.2.

4.1. The Scheme For Evidence Propagation Proposed by Pearl and Kim

One of the earliest schemes for propagating evidence in a belief network was proposed by J. Pearl and J.H. Kim. Their method as presented in [15], however, is not able to deal with belief networks as general as defined in the preceding section: Pearl and Kim depart from a restricted type of belief network in which the graphical part is a *singly connected digraph.*

DEFINITION 4.1. *A directed graph G is called* singly connected *if the underlying graph of G is acyclic.*

A singly connected digraph sometimes is called a *generalized Chow tree*, [15], or a *causal polytree*, [4]. Note that the digraph shown in Figure 3.1 is not singly connected.

For evidence propagation in such a restricted type of belief network, Pearl and Kim have exploited the following topological property of a singly connected digraph explicitly: the removal of any arc from a singly connected digraph G splits it up into two separate components. This property is stated more formally in the following lemma.

LEMMA 4.2. *Let G be a singly connected digraph. Any graph G' obtained by deleting an arbitrary arc from G is disconnected.*

From this lemma we have that in a singly connected digraph G we can identify for a vertex V with m neighbours, m connected subgraphs of G each containing a neighbour of V such that after removal of V from G there does not exist a path from one such subgraph to another one.

DEFINITION 4.3. *Let* $G = (V(G),A(G))$ *be a singly connected digraph. Let* $V \in V(G)$. *For each* $V_i \in \rho(V)$, *let* $G^o_{V_i} = (V(G^o_{V_i}),A(G^o_{V_i}))$ *such that* $V(G^o_{V_i}) = V(G)$ *and* $A(G^o_{V_i}) = A(G) \setminus \{(V_i,V)\}$. *The component of* $G^o_{V_i}$ *containing* V_i, *denoted by* $G^+_{(V_i,V)}$, *is called an* upper graph *of V.*

For each $V_j \in \sigma(V)$, *let* $G^o_{V_j} = (V(G^o_{V_j}),A(G^o_{V_j}))$ *such that* $V(G^o_{V_j}) = V(G)$ *and* $A(G^o_{V_j}) = A(G) \setminus \{(V,V_j)\}$. *The component of* $G^o_{V_j}$ *containing* V_j, *denoted by* $G^-_{(V,V_j)}$, *is called a* lower graph *of V.*

We say that V partitions *G into* $\{G^+_{(V_i,V)} \mid V_i \in \rho(V)\} \cup \{G^-_{(V,V_j)} \mid V_j \in \sigma(V)\}$.

The following example illustrates the idea; the example has been taken from [15] and adapted to our notational convention.

EXAMPLE 4.4. Consider Figure 4.1 showing a part of a singly connected digraph G. The vertex V_0 has four neighbours $V_1, \ldots, V_4$. V_0 therefore partitions G into four subgraphs: the two upper graphs $G^+_{(V_1,V_0)}$ and $G^+_{(V_2,V_0)}$, and the two lower graphs $G^-_{(V_0,V_3)}$ and $G^-_{(V_0,V_4)}$. ■

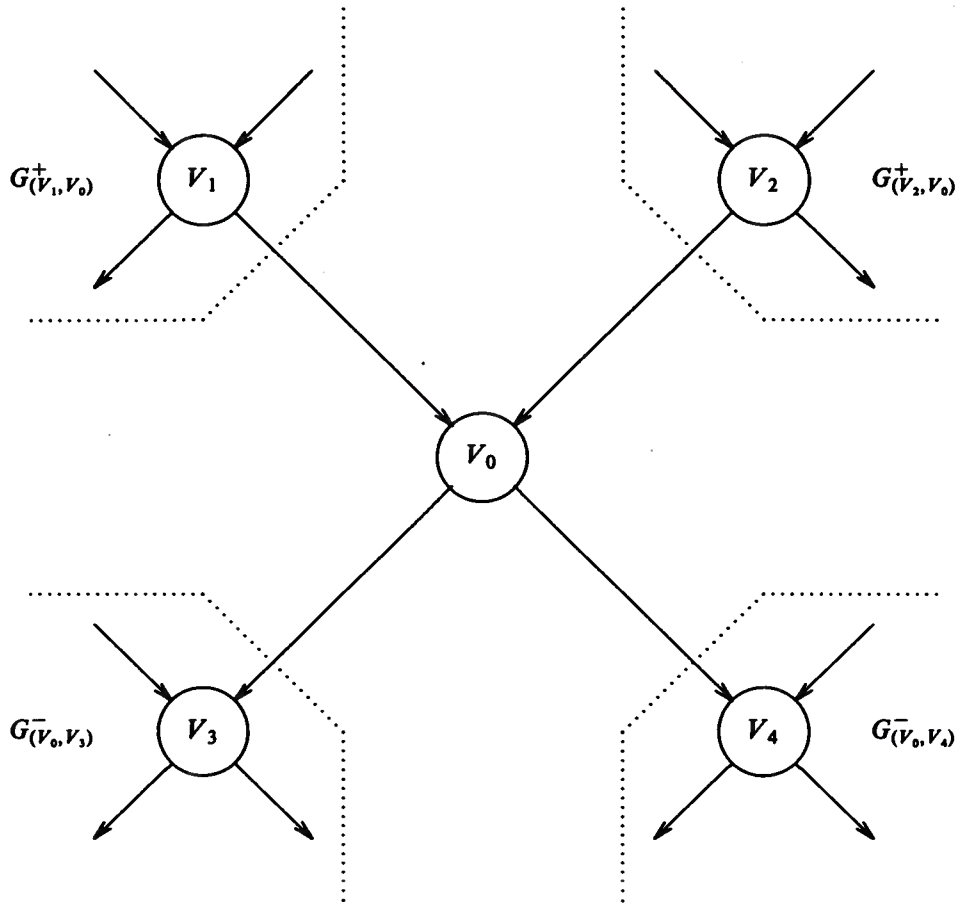

FIGURE 4.1. V_0 partitioning the singly connected digraph G.

In general we have at any time during a consultation in which evidence becomes available, that the actual probabilities of the values of a vertex V in a singly connected digraph G are dependent upon the information from its upper and lower graphs, that is, upon all data observed so far. Before discussing this in detail, we introduce some new notations.

In the remainder of this subsection, we will frequently encounter the situation in which some of the statistical variables in the digraph are known to have a specific value. In these situations we (ambiguously) use the notation C_V to denote a *partial configuration* of a set V. In such a partial configuration, some variables have been 'instantiated' with a value and some have not, that is, we take $C_V = \left[\bigwedge_{V_i \in V'} c_{V_i}\right] \wedge \left[\bigwedge_{V_j \in V''} V_j\right]$, $V' \cap V'' = \varnothing$, $V' \cup V'' = V$; we define $[C_V]$ by $[C_V] = c_{V'}$ if $V' \neq \varnothing$, and $[C_V] = \textit{true}$ otherwise. Furthermore, we will frequently identify upper and lower graphs with their vertex sets, as long as ambiguity cannot occur.

It will be evident that at any time during a consultation the probability distribution for the values of an 'interior' uninstantiated vertex $V \in V(G)$ equals the conditional probability distribution $Pr(V|[C_{G^+} \wedge C_{G^-}])$ where $C_{G^+} = \bigwedge_{V_j \in p(V)} C_{G^+_{(V_j,V)}}$ and $C_{G^-} = \bigwedge_{V_i \in s(V)} C_{G^-_{(V,V_i)}}$. From Bayes' Theorem and the independence relationships represented in G we have

$$Pr(V|[C_{G^+} \wedge C_{G^-}]) = \alpha \cdot Pr([C_{G^-}]|V) \cdot Pr(V|[C_{G^+}]) =$$
$$= \alpha \cdot \prod_{V_i \in \sigma(V)} Pr([C_{G^-_{(V,V_i)}}]|V) \cdot Pr(V|[C_{G^+}]) =$$
$$= \alpha \cdot \prod_{V_i \in \sigma(V)} Pr([C_{G^-_{(V,V_i)}}]|V) \cdot \left[\sum_{c_{\rho(V)}} Pr(V|c_{\rho(V)}) \cdot Pr(c_{\rho(V)}|[C_{G^+}])\right] =$$
$$= \alpha \cdot \prod_{V_i \in \sigma(V)} Pr([C_{G^-_{(V,V_i)}}]|V) \cdot \left[\sum_{c_{\rho(V)}} Pr(V|c_{\rho(V)}) \cdot \prod_{c_{V_j} \text{ in } c_{\rho(V)}} Pr(c_{V_j}|[C_{G^+_{(V_j,V)}}])\right],$$

where α is a normalization constant. The last two equalities are obtained by first conditioning on all configurations of the set $\rho(V)$ of predecessors of V and then using the independency relationships shown in the graph.

Recall that the values $Pr(V|c_{\rho(V)})$ have been specified initially in the conditional probability vector γ_V. So, if V obtains the probabilities $Pr([C_{G^-_{(V,V_i)}}]|V)$ from its successors V_i, and the probability $Pr(c_{V_j}|[C_{G^+_{(V_j,V)}}])$ from each of its predecessors V_j, then the vertex V is able to locally compute the probability distribution for its values.

DEFINITION 4.5. *Let $G = (V(G),A(G))$ be a singly connected digraph. Let Pr be a joint probability distribution on $V(G)$.*

Let $V \in V(G)$ such that $\rho(V) \neq \emptyset$. Furthermore, let $G^+_{(V_j,V)}$, $V_j \in \rho(V)$, be an upper graph of V. Then, we define $\pi_V(V_j) = Pr(V_j|[C_{G^+_{(V_j,V)}}])$. $\pi_V(V_j)$ is called the causal evidence parameter *from V_j to V.*

Now let $V \in V(G)$ such that $\sigma(V) \neq \emptyset$. Let $G^-_{(V,V_i)}$, $V_i \in \sigma(V)$, be a lower graph of V. Then, we define $\lambda_{V_i}(V) = Pr([C_{G^-_{(V,V_i)}}]|V)$. $\lambda_{V_i}(V)$ is called the diagnostic evidence parameter *from V_i to V.*

Note that before any evidence has been propagated, the causal evidence parameter that a vertex having no predecessors sends to its successors, equals the prior probability distribution for its values; furthermore, the diagnostic evidence parameter that a vertex having no successors sends to its predecessors, is a constant function yielding the function value 1 for all arguments.

The π and λ parameters can be viewed as being associated with the arcs in the singly connected digraph; Figure 4.2 for example shows the parameters which are associated with the digraph from Figure 4.1.

Lemma 4.6 now follows from our observations and the preceding definition.

LEMMA 4.6. *Let $B = (G,\Gamma)$ be a belief network defined as in Definition 3.3, such that $G = (V(G),A(G))$ is a singly connected digraph. Let the π and λ parameters be defined according to Definition 4.5. Now, let $V \in V(G)$ be an uninstantiated vertex partitioning G into $\{G^+_{(V_j,V)} \mid V_j \in \rho(V)\} \cup \{G^-_{(V,V_i)} \mid V_i \in \sigma(V)\}$. Then,*

$$Pr(V|[C_{G^+} \wedge C_{G^-}]) = \alpha \cdot \prod_{V_i \in \sigma(V)} \lambda_{V_i}(V) \cdot \left[\sum_{c_{\rho(V)}} \gamma_V(V|c_{\rho(V)}) \cdot \prod_{c_{V_j} \text{ in } c_{\rho(V)}} \pi_V(c_{V_j})\right],$$

where α is a normalization constant.

Consider the statement from Lemma 4.6 once more. If the vertex V has no upper graphs, then it will be evident that the probability distribution for the values of V does not involve the π parameters; for a vertex V having no lower graphs the probability distribution for its values does not involve the λ parameters.

Now suppose that evidence becomes available that a certain variable $V \in V(G)$ has a certain

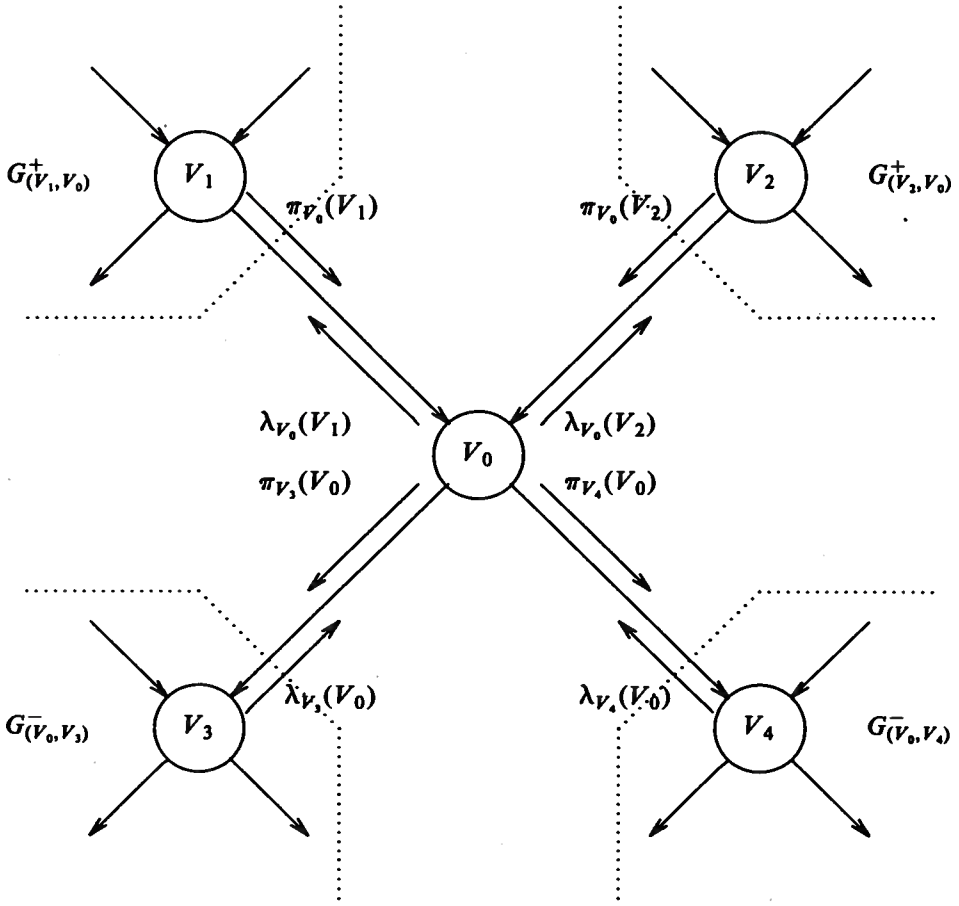

FIGURE 4.2. The π - λ parameters associated with the singly connected digraph G.

value. This evidence causes that variable V to update the joint probability distribution for its values locally. From its new knowledge about the joint probability distribution, V computes the proper π and λ parameters to be sent to its neighbours. V's neighbours subsequently are forced to update their knowledge about the joint probability distribution and to send new parameters to their neighbours in turn. After the initial updating of the joint probability distribution by V, the updating of all parameters is governed by the properties stated in Lemma 4.7.

LEMMA 4.7. *Let $B = (G, \Gamma)$ be a belief network such that $G = (V(G), A(G))$ is a singly connected digraph. Let the π and λ parameters be defined according to Definition 4.5. Let $V \in V(G)$. Then, the following properties hold:*

(1) $$\pi_{V_k}(V) = \alpha \cdot \prod_{\substack{V_i \in \sigma(V) \\ i \neq k}} \lambda_{V_i}(V) \cdot \left[\sum_{c_{\beta(V)} = \bigwedge_{v_j \in \beta(V)} c_{V_j}} \gamma_V(V \mid c_{\beta(V)}) \cdot \prod_{c_{V_j} \text{ in } c_{\beta(V)}} \pi_V(c_{V_j}) \right],$$

for each $V_k \in \sigma(V)$, where α is a normalization constant.

$$(2)\quad \lambda_V(V_k) = \alpha \cdot \sum_{c_V} \left[\prod_{V_i \in \sigma(V)} \lambda_{V_i}(c_V) \cdot \left[\sum_{c_{V'} = \bigwedge_{V_j \in \rho(V)} c_{V_j}} \gamma_V(c_V \mid c_{V'} \wedge V_k) \cdot \prod_{c_{V_j} \text{ in } c_{V'}} \pi_V(c_{V_j}) \right] \right],$$

for each $V_k \in \rho(V)$, *where* α *is a normalization constant.*

PROOF. The lemma can easily be proven using Definition 4.5, the topological properties of the singly connected digraph G and the independency relationships between the statistical variables represented in G. For a more detailed discussion of the properties stated in the lemma, see [4]. ■

Lemma 4.7 shows how evidence, once entered, is propagated through the belief network. The properties mentioned in the lemma show that any change in the causal parameter π associated with a specific arc of the digraph does not affect the diagnostic parameter λ on the same arc, and vice versa. Pearl argues [4], that any perturbation in the joint probability distribution in response to new evidence spreads through the digraph in a single pass, since the perturbation is absorbed without reflection at the 'boundary' vertices with either one outgoing or one incoming arc (note that every singly connected digraph must have at least two such vertices).

It remains to be discussed how a piece of evidence may be entered into the network. In [4], Pearl suggests an elegant way for entering evidence: if evidence has become available that the variable V has the value v (or $\neg v$ alternatively), then a dummy successor W of V is temporarily added to the digraph sending a diagnostic parameter $\lambda_W(V)$ to V such that $\lambda_W(v) = 1$ and $\lambda_W(\neg v) = 0$ (or vice versa if the value $\neg v$ has been observed).

We have mentioned before that the method for reasoning with belief networks presented by Pearl and Kim is not as general as the one presented by Spiegelhalter and Lauritzen which is to be discussed in the following subsection; the method is not able to deal with *loops* (cycles in the underlying graph of the acyclic digraph of the belief network). For a non-trivial application however, one may expect that the digraph contains one or more such loops. In fact, Jensen et al. [17] have encountered this problem in developing the MUNIN system (a medical expert system for electromyography), in which the method of Pearl and Kim is employed: the originally assessed digraph contained loops. Jensen et al. solved the problem by transforming the original graph into a singly connected digraph by combining several vertices into compound vertices, a method known as *clustering*, and by removing 'weak' arcs. Pearl also suggests clustering as a method of handling loops; he further introduces the method of *conditioning* on value combinations of vertices that decomposes the digraph into a singly connected digraph and then averaging over the results obtained. For further details the reader is referred once more to [4].

4.2. The Scheme Presented by Spiegelhalter and Lauritzen

The scheme for evidence propagation presented by Spiegelhalter and Lauritzen departs from an *undirected* graphical and probabilistic representation of the problem domain. Their scheme has been inspired by the existing statistical theory of Markov random fields, and more in specific, by the theory of graphical models (i.e. probabilistic models that can be represented by an undirected graph) in contingency tables, see for example [16]. To be able to exploit this theory, the original *directed* belief network (or recursive model as probabilistic models that can be represented by a directed graph are generally called in statistics) is transformed initially into an *undirected* so-called decomposable belief network, consisting of a decomposable graph and a set of marginal distributions on the cliques of this graph. In Definition 4.8 we introduce the notion of a decomposable graph; a decomposable belief network is defined in Definition 4.9.

DEFINITION 4.8. *Let G be an undirected graph. G is a* decomposable graph *if all elementary cycles of length $k \geqslant 4$ possess a chord. The set of all cliques in G, denoted by $Cl(G)$, is called the* clique set *of G.*

Decomposable graphs are also called *triangulated graphs*, [11,12]. The term decomposable has been

adopted from [18]. In the sequel, we will identify a clique $Cl_i \in Cl(G)$ with its vertex set, and write Cl_i instead of $V(Cl_i)$, as long as ambiguity cannot occur.

DEFINITION 4.9. *A* decomposable belief network *is a tuple* $B = (G, \mathbf{M})$ *such that*

(1) *G is a decomposable graph with a clique set* $Cl(G) = \{Cl_1, \ldots, Cl_m\}$, $m \geq 1$, *and*
(2) $\mathbf{M} = \{\mu_{Cl_i} \mid Cl_i \in Cl(G)\}$ *is a set of marginal distributions* $\mu_{Cl_i}(C_{Cl_i})$ *on the cliques* Cl_i *of G, such that for each* $Cl_i, Cl_j \in Cl(G)$ *with* $Cl_i \cap Cl_j \neq \emptyset$ *we have* $\mu_{Cl_i}(C_{Cl_i \cap Cl_j}) = \mu_{Cl_j}(C_{Cl_i \cap Cl_j})$.

Consider the previous definition once more. Note that for each $Cl_i, Cl_j \in Cl(G)$ with $Cl_i \cap Cl_j \neq \emptyset$, we may obtain $\mu_{Cl_i}(C_{Cl_i \cap Cl_j})$ from $\mu_{Cl_i}(C_{Cl_i})$ by *further marginalization*, that is, by summing out the variables from $Cl_i \setminus (Cl_i \cap Cl_j)$: $\mu_{Cl_i}(C_{Cl_i \cap Cl_j}) = \sum \mu_{Cl_i}(C_{Cl_i \cap Cl_j} \wedge c_{Cl_i \setminus (Cl_i \cap Cl_j)})$. Darroch et al. [16] have proven that in a decomposable belief network, the set M of clique marginals uniquely defines a joint probability distribution. Here, we merely mention this property; in Subsection 4.2.2 we will discuss the property in further detail.

The transformation of the originally assessed belief network into a decomposable belief network comprises several steps, as shown in Figure 4.3: the graphical representation of the belief network is transformed into a decomposable graph and from the probabilistic part of the network a new representation of the joint probability distribution in terms of marginal distributions on the cliques of the decomposable graph is obtained. The overall transformation essentially comprises two steps:

(1) Transform the initial belief network into a moral belief network consisting of a moral graph and an associated local representation of the joint probability distribution in evidence potentials.
(2) Transform the moral belief network into a decomposable belief network consisting of a decomposable graph and an associated local representation of the joint probability distribution in clique marginals.

These steps will be discussed in the Subsections 4.2.1 and 4.2.2 respectively. Subsection 4.2.3 will deal with the actual scheme for evidence propagation.

4.2.1. The moral belief network

Spiegelhalter has pointed out that 'the class of recursive models and the class of graphical models intersect in the class of decomposable models, and a recursive model is a member of this intersection provided it does not have two non-adjacent vertices both preceding the same vertex', ([3], p. 51; the statement has been proven in [6]), thus providing a motivation and means for the construction of the graphical part of the moral belief network.

Consider a belief network $B = (G, \Gamma)$ as defined in Section 3. Informally speaking, the *moral graph* G_M of the digraph G is obtained by first adding arcs to G such that no vertex in $V(G)$ has nonadjacent predecessors, and subsequently dropping the directions of the arcs. The moral graph of an acyclic digraph is defined formally in the following definition.

DEFINITION 4.10. *Let* $G = (V(G), A(G))$ *be an acyclic directed graph with vertices* $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$, *and arcs* $A(G)$. *Let* G' *be the simple digraph* $G' = (V(G'), A(G'))$ *such that* $V(G') = V(G)$ *and* $A(G') = A(G) \cup \{(V_i, V_j) \mid$ *there is an index k such that* $V_i, V_j \in \rho(V_k)$ *and* $(V_i, V_j), (V_j, V_i) \notin A(G)\}$. *The* moral graph G_M *of G is defined as the underlying graph of* G'.

The construction of the moral graph for our running example is demonstrated below.

EXAMPLE 4.11. Consider the acyclic digraph G from Figure 3.1 once more. Upon successively examining the vertices V_1 through V_8 we find that the predecessors of vertex V_6, that is, the vertices V_2 and V_4, are not adjacent, and that the same holds for the predecessors of vertex V_7. We

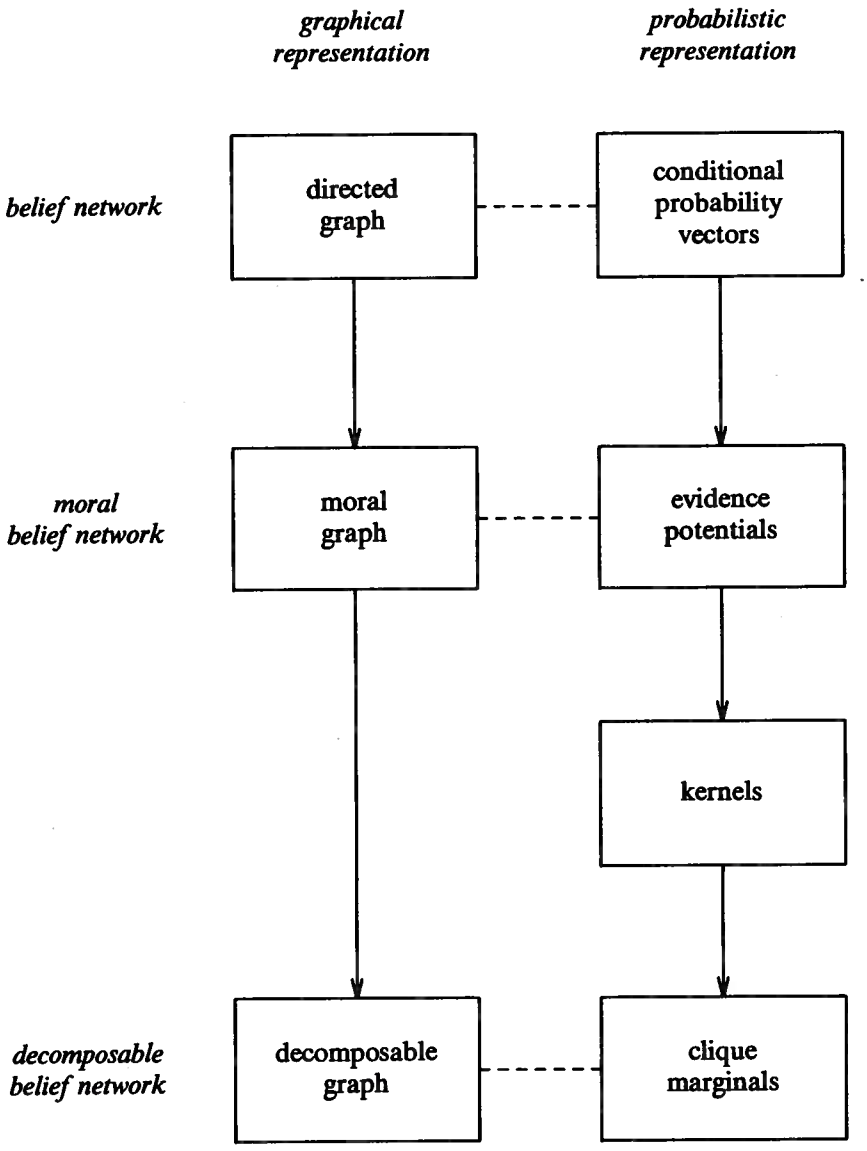

FIGURE 4.3. Transformation of the original belief network.

therefore add the arcs (V_2, V_4) and (V_5, V_6) to G. Note that the directions of these arcs are irrelevant since we will drop all directions subsequently. The construction of the moral graph G_M of G is shown in Figure 4.4. ■

LEMMA 4.12. *Let* $G = (V(G), A(G))$ *be an acyclic directed graph with vertices* $V(G) = \{V_1, \ldots, V_n\}$, $n \geqslant 1$. *Let* $V_i^\rho = V_i \cup \rho(V_i)$ *for each* $V_i \in V(G)$. *Furthermore, let* G_M *be the moral graph of* G *as defined above. Then for each* $V_i \in V(G)$, *the full subgraph of* G_M *induced by* V_i^ρ *is complete.*

For the qualitative part of the original belief network, we now have obtained an undirected

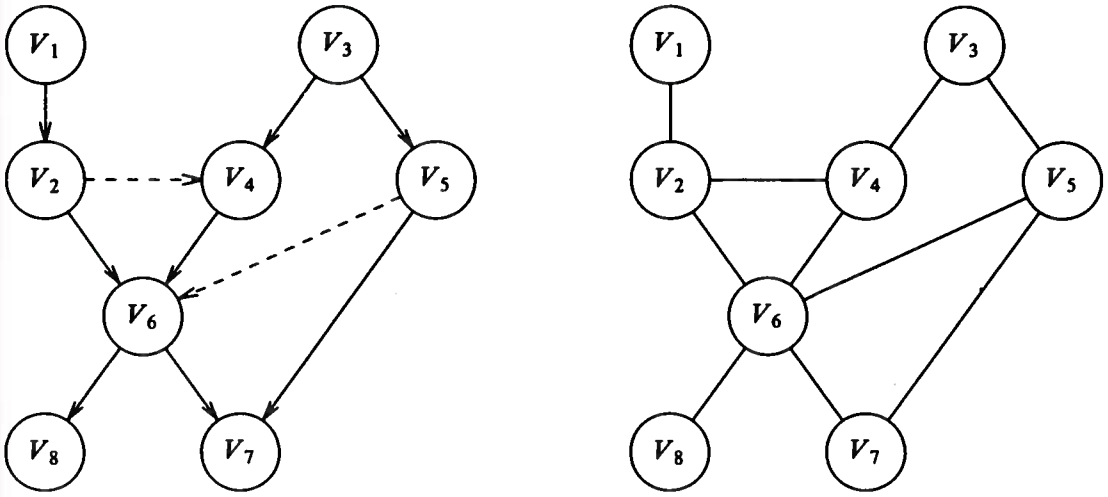

FIGURE 4.4. Construction of the moral graph G_M.

representation. This undirected graph again demonstrates certain independency relationships between the statistical variables. The following example shows that some of the initially assessed independencies are no longer visible explicitly in the moral graph. The arcs that were added to the original graph therefore should be taken as a kind of 'dummy' relationships.

EXAMPLE 4.13. Consider the graphs shown in Figure 4.5(a) - (d). The digraph (a) for example represents independency of the variables V_1 and V_2; this independency however is no longer represented explicitly in the corresponding moral graph (d). The three digraphs (a) - (c) represent different probabilistic relationships between the variables V_1, V_2 and V_3: (b) for example shows independency of V_1 and V_3. All three digraphs nevertheless have the same moral graph. ■

For the moral graph, we now obtain an 'undirected' representation of the initially given joint probability distribution. This new representation is based on the notion of an *evidence potential*, a real-valued nonnegative function the values of which only depend on the configurations of small sets of vertices.

Definition 4.14 introduces these evidence potentials and the resulting moral belief network.

DEFINITION 4.14. *Let $B = (G, \Gamma)$ be a belief network as defined in Definition 3.3. Let for each $V_i \in V(G)$, V_i^ρ be defined as $V_i^\rho = V_i \cup \rho(V_i)$. The* moral belief network *B_M derived from B is the tuple $B_M = (G_M, \Psi)$ where*

(1) *G_M is the moral graph of G as defined in Definition 4.10, and*
(2) *$\Psi = \{\psi_{V_i^\rho} \mid V_i \in V(G)\}$ is the set of* evidence potentials *$\psi_{V_i^\rho}$, where each $\psi_{V_i^\rho}$ is the real-valued non-negative function defined by $\psi_{V_i^\rho}(C_{V_i^\rho}) = \gamma_{V_i}(V_i \mid C_{\rho(V_i)})$, $C_{V_i^\rho} = V_i \wedge C_{\rho(V_i)}$.*

An evidence potential can be viewed as the proportional contribution of the indicated set of variables to the joint probability distribution.

EXAMPLE 4.15. Consider the acyclic digraph G from Figure 3.1 once more and its corresponding

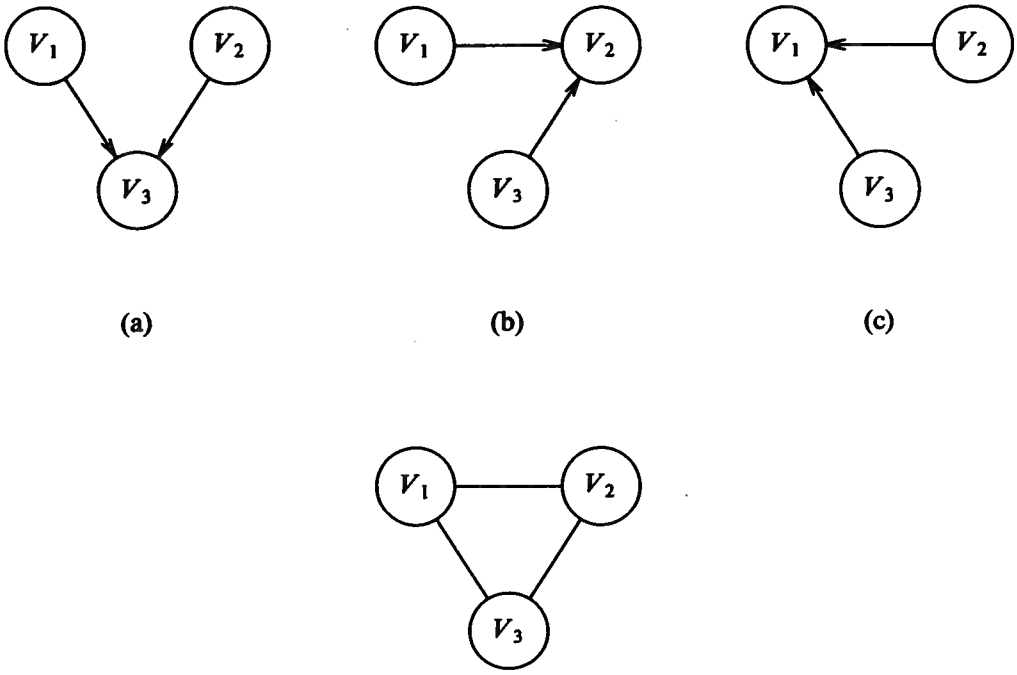

FIGURE 4.5. Three digraphs having the same moral graph.

moral graph G_M as shown in Figure 4.4(b). We obtain the following representation of the joint probability distribution in terms of evidence potentials from the conditional probability vectors associated with G:

$$
\begin{aligned}
\psi_{\{V_1\}}(V_1) &= \gamma_{V_1}(V_1) \\
\psi_{\{V_1,V_2\}}(V_1 \wedge V_2) &= \gamma_{V_2}(V_2 \mid V_1) \\
\psi_{\{V_3\}}(V_3) &= \gamma_{V_3}(V_3) \\
\psi_{\{V_3,V_4\}}(V_3 \wedge V_4) &= \gamma_{V_4}(V_4 \mid V_3) \\
\psi_{\{V_3,V_5\}}(V_3 \wedge V_5) &= \gamma_{V_5}(V_5 \mid V_3) \\
\psi_{\{V_2,V_4,V_6\}}(V_2 \wedge V_4 \wedge V_6) &= \gamma_{V_6}(V_6 \mid V_2 \wedge V_4) \\
\psi_{\{V_5,V_6,V_7\}}(V_5 \wedge V_6 \wedge V_7) &= \gamma_{V_7}(V_7 \mid V_5 \wedge V_6) \\
\psi_{\{V_6,V_8\}}(V_6 \wedge V_8) &= \gamma_{V_8}(V_8 \mid V_6)
\end{aligned}
$$

■

LEMMA 4.16. *Let $B = (G, \Gamma)$ be a belief network where $G = (V(G), A(G))$ is an acyclic digraph with n vertices, $n \geq 1$. Let $Pr(V_1 \wedge \cdots \wedge V_n) = \prod_{i=1,\ldots,n} \gamma_{V_i}(V_i \mid C_{p(V_i)})$ constitute a joint probability distribution on $V(G)$. Furthermore, let $B_M = (G_M, \Psi)$ be the moral belief network derived from B defined as above. Then, we have*

$$Pr(V_1 \wedge \cdots \wedge V_n) = \prod_{i=1,\ldots,n} \psi_{V_i^*}(C_{V_i^*}).$$

PROOF. The property follows immediately from Definition 4.14 and Lemma 3.4. ■

From Lemma 4.16 we have that although some of the originally assessed independence relationships are no longer explicitly represented in the moral graph, they still are represented implicitly in the joint probability distribution. Note that the representation of the joint probability distribution in terms of evidence potentials again is a local representation of uncertainty.

4.2.2. The Decomposable Belief Network

We recall that the transformation of the initial belief network into a decomposable belief network comprises two steps, the first of which we have discussed in the preceding subsection. In this subsection we discuss the transformation of the moral belief network resulting from the first transformation step into a corresponding decomposable belief network. We first consider the transformation of the moral graph into a decomposable graph.

The moral graph can be made decomposable by *filling-in*, that is, (again) by adding certain 'dummy' edges. Lauritzen and Spiegelhalter use an efficient algorithm by R.E. Tarjan and M. Yannakakis [19], for doing so. Before discussing this algorithm we provide some preliminary definitions.

DEFINITION 4.17. *Let* $G = (V(G),E(G))$ *be an undirected graph with vertices* $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$. *Let* $\iota: V(G) \leftrightarrow \{1, \ldots, n\}$ *denote a total ordering of the vertices of G. The ordering* ι *is called a* perfect ordering *of the vertices* $V(G)$ *if for each number* $i = \iota(V_j)$, $i,j \in \{1, \ldots, n\}$, *the full subgraph of G induced by the set of vertices* $v_G(V_j) \cap \{V_{j_1}, \ldots, V_{j_{i-1}}\}$, $\iota(V_{j_k}) = k$, $k = 1, \ldots, i-1$, *is complete.*

Note that an undirected graph may allow more than one perfect ordering. In the sequel, we will identify a vertex with its number according to an ordering ι as long as ambiguity cannot occur. The notion of a perfect ordering and its definition have been taken from [12]; in discussing the algorithm for filling-in by Tarjan and Yannakakis, Pearl uses a notion similar to the one defined above, [4]. Tarjan and Yannakakis themselves however define the notion of a *zero fill-in numbering* and show in a lemma that an ordering is a zero fill-in numbering if and only if it has the property we have used for a definition, [19].

The following lemma now is of major importance (for a proof of the lemma, see [19]).

LEMMA 4.18. *Let G be an undirected graph. G is decomposable if and only if it permits a perfect ordering of its vertices.*

There are several ways to compute a perfect ordering of the vertices $V(G)$ of a decomposable graph G, one of which is known as maximum cardinality search.

ALGORITHM 4.19. *Let* $G = (V(G),E(G))$ *be an undirected graph. The* maximum cardinality search *algorithm for computing an ordering* ι *of* $V(G)$ *is the following:*

1. *Assign the number 1 to an arbitrary vertex.*
2. *Number the remaining vertices from 2 to* $|V(G)|$ *in increasing order such that the next number is assigned to the vertex having the largest set of previously numbered neighbours.*

Note that in Algorithm 4.19 the next vertex to be numbered need not be a neighbour of the last numbered vertex. Furthermore, it will be evident that the algorithm is non-deterministic.

Tarjan and Yannakakis have proven that when applied to a decomposable graph, maximum cardinality search renders a perfect ordering:

LEMMA 4.20. *Let G be a decomposable graph. Any ordering ι of the vertices of G obtained from maximum cardinality search is perfect.*

The following algorithm is known as the fill-in algorithm for obtaining a decomposable graph from an arbitrary undirected graph.

ALGORITHM 4.21. *Let $G = (V(G), E(G))$ be an undirected graph. The fill-in algorithm is the following:*

1. *Compute an ordering ι for the vertices of G, using maximum cardinality search.*
2. *From $|V(G)|$ to 1, for each vertex numbered i add edges between any nonadjacent neighbours of i that are assigned a lower number than i in ι.*

The set of edges added to G is called the fill-in.

If by applying Algorithm 4.21 no edges are added to an undirected graph G, then G was already decomposable; the phrase 'zero fill-in numbering' used by Tarjan and Yannakakis instead of a 'perfect ordering' emerges from this observation. Otherwise, the new graph obtained from applying the algorithm is decomposable. This property is stated in the following lemma.

LEMMA 4.22. *Let G be an undirected graph. Let G' be an undirected graph obtained from G using the fill-in algorithm shown above. Then, G' is decomposable.*

EXAMPLE 4.23. Consider the moral graph G_M from Figure 4.4(b). We use Algorithm 4.21 to transform G_M into a decomposable graph G_D. Using maximum cardinality search the vertices of G_M may be numbered as shown in Figure 4.6(a).

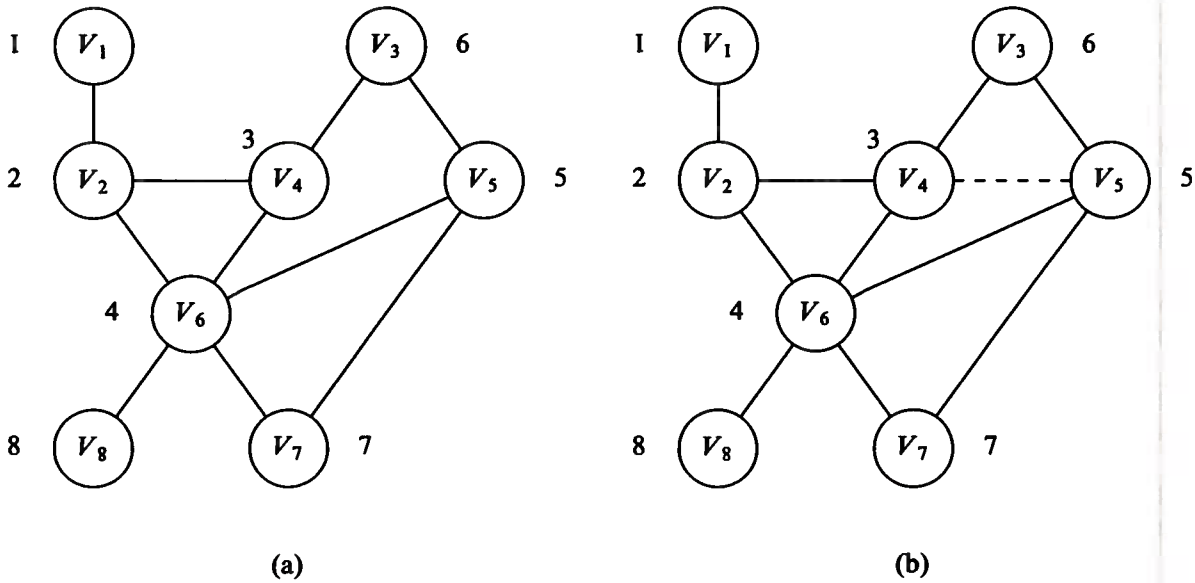

FIGURE 4.6. Construction of the decomposable graph G_D.

Examining the vertices from 8 to 1 we find that the vertex numbered 6 has two nonadjacent neighbours that are assigned a lower number than 6 in the ordering: the full subgraph generated by

$\{3,5\} \cap \{1,2,...,5\} = \{3,5\}$ is not complete. Therefore the edge (V_4, V_5) is added to G_M, yielding the decomposable graph G_D shown in Figure 4.5(b). Note that the alternative addition of (V_3, V_6) would have yielded a decomposable graph as well. ■

In [12], Spiegelhalter and Lauritzen point out that the fill-in should be computed very carefully, since the maximal order of the cliques of the decomposable graph resulting from the fill-in determines the computational complexity of their method. It should be noted that the problem of computing a fill-in containing a minimum number of edges is NP-complete, [20].

The graphical part of the original belief network has now been transformed into a decomposable graph. We have mentioned before that Darroch et al. [16] have shown that a joint probability distribution on a decomposable graph can be expressed in terms of marginal distributions on the cliques of the graph. Exploiting this property, Spiegelhalter and Lauritzen transform the representation of the joint probability distribution in terms of evidence potentials into such a representation in terms of clique marginals.

Before discussing this transformation in detail, we introduce several new notions.

DEFINITION 4.24. *Let $G = (V(G), E(G))$ be a decomposable graph with n vertices, $n \geqslant 1$. Let $\iota: V(G) \leftrightarrow \{1, \ldots, n\}$ be a perfect ordering of $V(G)$ obtained from maximum cardinality search. Let $Cl(G) = \{Cl_1, \ldots, Cl_m\}$, $m \geqslant 1$, be the clique set of G. We define the total ordering $\iota': Cl(G) \leftrightarrow \{1, \ldots, m\}$ of $Cl(G)$ such that for each $Cl_i, Cl_j \in Cl(G)$ we have $\iota'(Cl_i) < \iota'(Cl_j)$ if $max\{\iota(V_i) \mid V_i \in Cl_i\} < max\{\iota(V_j) \mid V_j \in Cl_j\}$.*

Note that in the ordering ι' the cliques of the graph G are numbered in the order of their highest numbered vertex according to ι. It will be evident that ι' is uniquely determined by the ordering ι.

EXAMPLE 4.25. Consider the decomposable graph G_D and the ordering ι of its vertices as shown in Figure 4.6(b) once more. We number the six cliques in the order of their highest numbered vertex. Let Cl_i be the clique assigned number i. Then, we have obtained the following ordering ι':

$$
\begin{aligned}
Cl_1 &= \{V_1, V_2\} \\
Cl_2 &= \{V_2, V_4, V_6\} \\
Cl_3 &= \{V_4, V_5, V_6\} \\
Cl_4 &= \{V_3, V_4, V_5\} \\
Cl_5 &= \{V_5, V_6, V_7\} \\
Cl_6 &= \{V_6, V_8\}
\end{aligned}
$$

■

The following lemma states an important property of the ordering ι' of the cliques of a decomposable graph. In [18,19] further details are provided. The lemma is known as the *running intersection property*.

LEMMA 4.26. *Let G be a decomposable graph. Let $Cl(G)$ be the set of cliques of G numbered $Cl_1, \ldots, Cl_m$, $m \geqslant 1$, according to the ordering ι' from Definition 4.24. Then, ι' has the following property: for all $i \geqslant 2$ there is a $j < i$ such that $Cl_j \supset Cl_i \cap (Cl_1 \cup \cdots \cup Cl_{i-1})$.*

The previous lemma states, in other words, that the vertices a clique has in common with the lower numbered cliques are all contained in one such clique.

DEFINITION 4.27. *Let G be a decomposable graph. Let $Cl(G)$ be the set of cliques of G numbered $Cl_1, \ldots, Cl_m$, $m \geqslant 1$, according to the ordering ι' from Definition 4.24. From Lemma 4.26 we have that ι' has the running intersection property. For each $i \leqslant m$, we define*

$S_i = Cl_i \cap (Cl_1 \cup \cdots \cup Cl_{i-1})$ *where* $S_1 = \varnothing$, *and* $R_i = Cl_i \setminus S_i$. S_i *is called the* separator *of clique* Cl_i; R_i *is called the* residue *of clique* Cl_i.

We recall that associated with the moral graph G_M we departed from for constructing the decomposable belief network, we had a representation Ψ of the joint probability distribution in terms of evidence potentials ψ_V, $V \subseteq V(G_M)$. We now are interested in a representation of the joint probability distribution in terms of marginal distributions on the cliques of the decomposable graph G_D obtained from G_M. Since the evidence potentials ψ_V not necessarily are defined on cliques or clique-intersections, they do not give rise to a representation of the joint probability distribution in clique marginals in a straightforward manner. In Lemma 4.28 we introduce an intermediary representation of the joint probability distribution in terms of so-called kernels; Lemma 4.30 gives the representation in terms of clique marginals.

LEMMA 4.28. *Let* $B_M = (G_M, \Psi)$ *be a moral belief network as defined in Definition 4.14, such that* $Pr(C_{V(G_M)}) = \prod_{\psi_V \in \Psi} \psi_V(C_V)$ *constitutes a joint probability distribution on* $V(G_M)$. *Let* G_D *be the decomposable graph obtained from* G_M *by using Algorithm 4.21. Furthermore, let* $Cl(G_D)$ *be the set of cliques in* G_D, *numbered* $Cl_1, \ldots, Cl_m$, $m \geqslant 1$, *according to the ordering* i' *from Definition 4.24. Let for each clique* Cl_i, *the separator* S_i *and the residue* R_i *be defined according to Definition 4.27. Then, there exists a set* $\mathrm{K} = \{\kappa_{Cl_i} \mid Cl_i \in Cl(G_D)\}$ *of functions* $\kappa_{Cl_i}(C_{R_i} \mid C_{S_i})$, *called* kernels, *such that* $Pr(C_{V(G_D)}) = \prod_{i=1,\ldots,m} \kappa_{Cl_i}(C_{R_i} \mid C_{S_i})$.

PROOF. The lemma has been proven by Spiegelhalter and Lauritzen, [12]. Since the proof gives a construction of K, we repeat their argument in the present paper.

Let $D_{m+1} = \{V \mid V \subseteq V(G_M), \psi_V \in \Psi\}$ be the set of (initial) evidence potential domains. We have that $Pr(C_{V(G_M)}) = \prod_{V \in D_{m+1}} \psi_V(C_V)$ is an evidence potential representation of the marginal distribution on $V(G_M)$. We recursively repeat the following computation for $i = m, \ldots, 1$:

Consider clique Cl_i and its residue R_i. Let $V(Cl)_i = Cl_1 \cup \cdots \cup Cl_i$. Now assume that $Pr(C_{V(Cl)_i}) = \prod_{V \in D_{i+1}} \psi_V(C_V)$ is an evidence potential representation of the marginal distribution on $V(Cl)_i$ (note that the assumption holds for $i = m$). We split the set D_{i+1} of evidence potential domains into two disjoint subsets: the set $D_i^+ = \{V \mid V \in D_{i+1}, V \cap R_i \neq \varnothing\}$ consisting of those domains that contain variables from R_i and the set $D_i^- = D_{i+1} \setminus D_i^+$ consisting of those domains that do not. We have

$$Pr(C_{V(Cl)_i \setminus R_i}) = \sum_{c_{R_i}} Pr(C_{V(Cl)_i \setminus R_i} \wedge c_{R_i}) =$$

$$= \prod_{V \in D_i^-} \psi_V(C_V) \cdot \sum_{c_{R_i}} \left[\prod_{V \in D_i^+} \psi_V(C_{V \setminus R_i} \wedge c_{R_i}) \right].$$

Note that the first equality merely states that $Pr(C_{V(Cl)_i \setminus R_i})$ is obtained from $Pr(C_{V(Cl)_i})$ by marginalization; the second equality follows from the earlier mentioned assumption $Pr(C_{V(Cl)_i}) = \prod_{V \in D_{i+1}} \psi_V(C_V)$.

We now have obtained a representation of the marginal distribution on $Cl_1 \cup \ldots \cup Cl_{i-1}$. We define a new potential domain $D_i = \bigcup_{V \in D_i^+} V \setminus R_i$; we furthermore define the function

$$\phi_{D_i}(C_{D_i}) = \sum_{c_{R_i}} \left[\prod_{V \in D_i^+} \psi_V(C_{V \setminus R_i} \wedge c_{R_i}) \right].$$

It follows that

$$Pr(C_{R_i} \mid C_{S_i}) = Pr(C_{R_i} \mid C_{Cl_1 \cup \cdots \cup Cl_{i-1}}) = \frac{Pr(C_{Cl_1 \cup \cdots \cup Cl_i})}{Pr(C_{Cl_1 \cup \cdots \cup Cl_{i-1}})} =$$

$$= \frac{\prod_{V \in D_{i+1}} \psi_V(C_V)}{\prod_{V \in D_i^-} \psi_V(C_V) \cdot \phi_{\overline{D}_i}(C_{\overline{D}_i})} = \prod_{V \in D_i^+} \frac{\psi_V(C_V)}{\phi_{\overline{D}_i}(C_{\overline{D}_i})}.$$

We add the new evidence potential domain $\overline{D}_i$ to D_i^- as an initialization for the following computation step of the recursion. So, we let $D_i = D_i^- \cup \{\overline{D}_i\}$; note that the elements of D_i do not contain variables from R_i. We furthermore define a new set of evidence potentials

$$\overline{\psi}_{\overline{D}_i}(C_{\overline{D}_i}) = \begin{cases} \psi_{\overline{D}_i}(C_{\overline{D}_i}) \cdot \phi_{\overline{D}_i}(C_{\overline{D}_i}) & \text{if } \overline{D}_i \in D_{i+1} \\ \phi_{\overline{D}_i}(C_{\overline{D}_i}) & \text{otherwise} \end{cases}$$

and

$$\overline{\psi}_V = \psi_V \text{ for all } V \in D_{i+1},\ V \neq \overline{D}_i,$$

resulting in an evidence potential representation of the marginal distribution on $Cl_1 \cup \cdots \cup Cl_{i-1}$. Subsequently, we rename $\overline{\psi}_V$ into ψ_V, and we repeat the computation for $i - 1$. Note that we have established the property $Pr(C_{V(Cl)_{i-1}}) = \prod_{V \in D_i} \psi_V(C_V)$.

Taking $\kappa_{Cl_i}(C_{R_i} \mid C_{S_i}) = Pr(C_{R_i} \mid C_{S_i})$, it will be evident that $Pr(C_{V(G_D)}) = \prod_{i=1,\ldots,m} \kappa_{Cl_i}(C_{R_i} \mid C_{S_i})$. ■

From the intermediary set of kernels introduced in the previous lemma, we obtain a set of clique marginals. Lemma 4.29 provides a means for constructing such a set.

LEMMA 4.29. *Let $B_M = (G_M, \Psi)$ be a moral belief network as defined in Definition 4.14, such that $Pr(C_{V(G_M)}) = \prod_{\psi_V \in \Psi} \psi_V(C_V)$ constitutes a joint probability distribution on $V(G_M)$. Let G_D be the decomposable graph obtained from G_M by using Algorithm 4.21. Furthermore, let $Cl(G_D)$ be the set of cliques in G_D, numbered $Cl_1, \ldots, Cl_m$, $m \geqslant 1$, according to the ordering ι'. Let for each Cl_i, the separator S_i and the residue R_i be defined according to Definition 4.27. Let $\mathbf{K} = \{\kappa_{Cl_i}(C_{R_i} \mid C_{S_i}) \mid Cl_i \in Cl(G_D)\}$ be the set of kernels as constructed in the previous lemma. Now for each Cl_i, $i = 1, \ldots, m$, we recursively define $\mu_{Cl_i}(C_{Cl_i}) = \kappa_{Cl_i}(C_{R_i} \mid C_{S_i}) \cdot \mu_{Cl_j}(C_{S_i})$, where j is chosen such that $Cl_j \supset Cl_i \cap (Cl_1 \cup \cdots \cup Cl_{i-1})$ and $\mu_{Cl_j}(C_{S_i})$ is obtained by marginalization. Let $\mathbf{M} = \{\mu_{Cl_i} \mid Cl_i \in Cl(G)\}$. Then, $B_D = (G_D, \mathbf{M})$ is a decomposable belief network.*

The following lemma states that the set of clique marginals as constructed in the previous lemma again defines a unique joint probability distribution.

LEMMA 4.30. *Let $B = (G, \mathbf{M})$ be a decomposable belief network as defined in Definition 4.9. Let $Cl(G)$ be the clique set of G, numbered $Cl_1, \ldots, Cl_m$, $m \geqslant 1$, according to the ordering ι'. Then,*

$$Pr(C_{V(G)}) = \prod_{i=1,\ldots,m} \frac{\mu_{Cl_i}(C_{Cl_i})}{\mu_{Cl_i}(C_{S_i})}$$

constitutes a joint probability distribution on $V(G)$.

PROOF. The lemma has been proven in [16] for a general decomposable belief network. The validity of the statement for the decomposable belief network resulting from the transformation of the originally assessed belief network can be seen from the Lemmas 4.28 and 4.29, and the observation that for each $Cl_i, Cl_j \in Cl(G)$ such that $Cl_i \cap Cl_j \neq \emptyset$, we have $\mu_{Cl_i}(C_{Cl_i \cap Cl_j}) = \mu_{Cl_j}(C_{Cl_i \cap Cl_j})$. ■

EXAMPLE 4.31. Consider the decomposable graph G_D from Figure 4.6(b) once more. Associated with G_D we obtain the set of clique marginals

$$\mathrm{M} = \{\mu_{\{V_1,V_2\}}, \mu_{\{V_2,V_4,V_6\}}, \mu_{\{V_4,V_5,V_6\}}, \mu_{\{V_3,V_4,V_5\}}, \mu_{\{V_5,V_6,V_7\}}, \mu_{\{V_6,V_8\}}\}$$

giving rise to the following representation of the joint probability distribution:

$$Pr(V_1 \wedge \cdots \wedge V_8) =$$

$$= \mu_{\{V_1,V_2\}}(V_1 \wedge V_2) \cdot \frac{\mu_{\{V_2,V_4,V_6\}}(V_2 \wedge V_4 \wedge V_6)}{\mu_{\{V_2,V_4,V_6\}}(V_2)} \cdot \frac{\mu_{\{V_4,V_5,V_6\}}(V_4 \wedge V_5 \wedge V_6)}{\mu_{\{V_4,V_5,V_6\}}(V_4 \wedge V_6)} \cdot$$

$$\cdot \frac{\mu_{\{V_3,V_4,V_5\}}(V_3 \wedge V_4 \wedge V_5)}{\mu_{\{V_3,V_4,V_5\}}(V_4 \wedge V_5)} \cdot \frac{\mu_{\{V_5,V_6,V_7\}}(V_5 \wedge V_6 \wedge V_7)}{\mu_{\{V_5,V_6,V_7\}}(V_5 \wedge V_6)} \cdot \frac{\mu_{\{V_6,V_8\}}(V_6 \wedge V_8)}{\mu_{\{V_6,V_8\}}(V_6)}.$$

■

4.2.3. Propagation of Evidence in the Decomposable Belief Network

In the previous two subsections we have discussed a transformation of the initially assessed belief network into a decomposable belief network. The scheme for evidence propagation proposed by Spiegelhalter and Lauritzen operates on this decomposable belief network. We emphasize that for a specific problem domain the transformation has to be performed only once: each consultation proceeds from the obtained decomposable belief network. In this subsection we discuss the propagation of a single piece of evidence through a decomposable belief network.

Suppose that evidence becomes available that a statistical variable V has adopted a certain value. For ease of exposition, we assume that the variable V occurs in precisely one clique of the decomposable graph. Informally speaking, propagation of this evidence amounts to the following. The vertices and the cliques of the decomposable graph are ordered anew as described in Subsection 4.2.2, this time starting with the 'observed' vertex. The ordering of the cliques then is taken as the order in which the evidence is propagated through the cliques. For each subsequent clique, the updated marginal distribution is computed locally. Then, the observed vertex is removed from the graph, and the updated marginal distributions are taken as the marginal distributions on the cliques of the remaining graph. The process may then be repeated for a new piece of evidence.

In the following definition we introduce the notions of an updated joint probability distribution and an updated graph.

DEFINITION 4.32. *Let $B = (G, \mathrm{M})$ be a decomposable belief network as defined in Definition 4.9. Now, let the evidence $V = v$ (or $V = \neg v$ alternatively) be observed for a vertex $V \in V(G)$. Let $Pr(C_{V(G)})$ be the joint probability distribution on $V(G)$ defined by* M. *Now, we define the* updated probability distribution *Pr^* by $Pr^*(C_{V'}) = Pr(C_{V'} \mid v)$ for any $V' \subseteq V(G)$. Furthermore, we define the* updated graph *G^* of G by $G^* = (V(G^*), E(G^*))$ where $V(G^*) = V(G) \setminus V$ and $E(G^*) = E(G) \setminus \{(V_k, V) \mid V_k \in V(G^*)\}$.*

The following lemma will be evident.

LEMMA 4.33. *Let $B = (G, \mathrm{M})$ be a decomposable belief network. Let the evidence $V = v$ (or $V = \neg v$ alternatively) be observed for a vertex $V \in V(G)$. Furthermore, let G^* be the updated graph of G defined as above. Then, G^* is decomposable.*

The following lemma provides a method for computing the updated joint probability distribution in which the probabilistic computations to be performed are restricted to the cliques of the decomposable graph only. The lemma has been proven in [12].

LEMMA 4.34. *Let $B = (G, \mathbf{M})$ be a decomposable belief network. Let $Cl(G)$ be the clique set of G. Now, let the evidence $V = v$ (or $V = \neg v$ alternatively) be observed for a vertex $V \in V(G)$. Let ι be the ordering of $V(G)$ obtained from maximum cardinality search starting with V, that is, $\iota(V) = 1$. Let $Cl_1, \ldots, Cl_m$, $m \geqslant 1$, be the ordering of $Cl(G)$ according to ι' obtained from the ordering ι as in Definition 4.24. Let $Pr(C_{V(G)}) = \prod_{i=1,\ldots,m} \frac{\mu_{Cl_i}(C_{Cl_i})}{\mu_{Cl_i}(C_{S_i})}$ be the joint probability distribution on $V(G)$ defined by $\mathbf{M}$. Let Pr^* be the updated probability distribution. We define the following functions for the cliques of G:*

(1) $\mu^*_{Cl_1 \setminus V}(C_{Cl_1 \setminus V}) = \mu_{Cl_1}(C_{Cl_1 \setminus V} | v)$, *and*

(2) $\mu^*_{Cl_i}(C_{Cl_i}) = \mu_{Cl_i}(C_{Cl_i}) \cdot \frac{\mu^*_{Cl_i}(C_{S_i})}{\mu_{Cl_i}(C_{S_i})}$ *for* $i = 2, \ldots, m$.

Let $\mathbf{M}^* = \{\mu^*_{Cl_1 \setminus V}\} \cup \{\mu^*_{Cl_i} | i = 2, \ldots, m\}$. Then, we have that $\mathbf{M}^*$ defines Pr^*, that is, we have $Pr^*(C_{V(G)}) = \mu^*_{Cl_1 \setminus V}(C_{Cl_1 \setminus V}) \cdot \prod_{i=2,\ldots,m} \frac{\mu^*_{Cl_i}(C_{Cl_i})}{\mu^*_{Cl_i}(C_{S_i})}$.

COROLLARY 4.35. *Let $B = (G, \mathbf{M})$ be a decomposable belief network. Let the evidence $V = v$ (or $V = \neg v$ alternatively) be observed for a vertex $V \in V(G)$. Let Pr^* be the updated probability distribution and let the set $\mathbf{M}^*$ from Lemma 4.34 be its marginal representation. Furthermore, let G^* be the updated graph of G. Then, $B^* = (G^*, \mathbf{M}^*)$ is a decomposable belief network.*

5. SUMMARY AND CONCLUSION

In this paper we have introduced the notion of a belief network as a formalism for representing a problem domain. Such a belief network consists of a qualitative part representing statistical variables and their probabilistic interrelationships, and a quantitative part representing a joint probability distribution on the problem domain. We have discussed two different schemes for reasoning with such a belief network in which the qualitative part acts as an architecture for propagating evidence. We conclude this paper with a comparison of these network models with the quasi-probabilisic models mentioned in the introduction, since both depart from probability theory.

An apparent similarity between these types of models for reasoning with uncertain information in a knowledge-based system, is that both are local in their approach: uncertainty is represented in factors which are local in terms of the knowledge representation formalism, and propagation of evidence entails only local updating of the joint probability distribution.

One of the problems in applying probability theory in a model for handling uncertainty in a knowledge-based context frequently postulated by researchers in artificial intelligence, is the difficulty of obtaining a fully and consistently specified probability distribution on the problem domain: often only a few probabilities can be obtained from an expert in the field. In such a case, we are confronted with the problem of calculating certain probabilities from a partially and often inconsistently specified probability distribution. The quasi-probabilistic models are able to deal with such partially specified probability distributions, although not in a mathematically sound way. We have mentioned before that the quasi-probabilistic models are rather disappointing from a mathematical point of view: the schemes for propagating evidence employed in these models are incorrect and even as an approximation technique they are far from convincing. In contrast, the schemes for propagating evidence employed in the network models are mathematically sound. These models however, are not capable of dealing with partially and inconsistently specified probability distributions. In each of the network models the originally assessed belief network has to be fully and

consistently quantified: a joint probability distribution on the statistical variables has to be uniquely determined by the conditional probability vectors. In a subsequent paper we will address this problem, [21].

6. References

[1] P. Cheeseman (1985). In defense of probability, *Proceedings of the 9th International Joint Conference on Articial Intelligence*, pp. 1002 - 1009.

[2] L.A. Zadeh (1986). Is probability theory sufficient for dealing with uncertainty in AI: a negative view, in: L.N. Kanal, J. Lemmer (eds.) *Uncertainty in Artificial Intelligence*, North-Holland, Amsterdam, pp. 103 - 116.

[3] D.J. Spiegelhalter (1986). Probabilistic reasoning in predicative expert systems, in: L.N. Kanal, J.F. Lemmer (eds.) *Uncertainty in Artificial Intelligence*, North-Holland, Amsterdam, pp. 47 - 67.

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

[5] R.D. Shachter (1986). Intelligent probabilistic inference, in: L.N. Kanal, J.F. Lemmer (eds.) *Uncertainty in Artificial Intelligence*, North-Holland, Amsterdam, pp. 371 - 382.

[6] N. Wermuth, S.L. Lauritzen (1983). Graphical and recursive models for contingency tables, *Biometrika*, vol. 70, pp. 537 - 552.

[7] H. Kiiveri, T.P. Speed, J.B. Carlin (1984). Recursive causal models, *Journal of the Australian Mathematical Society (Series A)*, vol. 36, pp. 30 - 52.

[8] P. Shenoy, G. Shafer (1986). Propagating belief functions with local computations, *IEEE Expert*, vol. 1, pp. 43 - 52.

[9] A.P. Dempster, A. Kong (1988). Uncertain evidence and artificial analysis, *Journal of Statistical Planning and Inference*, vol. 20, pp. 355 - 368.

[10] R.J. Wilson (1979). *Introduction to Graph Theory, 2nd Edition*, Longman, Harlow, Essex.

[11] C. Berge (1973). *Graphs and Hypergraphs*, North-Holland, Amsterdam.

[12] S.L. Lauritzen, D.J. Spiegelhalter (1987). *Fast Manipulation of Probabilities with Local Computations - With Applications to Expert Systems*, Technical Report R-87-7, Aalborg University, Aalborg, Denmark.

[13] G.F. Cooper (1987). *Probabilistic Inference Using Belief Networks is NP-Hard*, Memo KLS-87-27, Knowledge Systems Laboratory, Medical Computer Science Group, Stanford University, Stanford, California.

[14] L.M. Trung (1988). The order of node removals in influence diagrams, *Proc. of the Workshop on Uncertainty Processing in Expert Systems*, Alšovice, Csechoslovakia, pp. 20 - 24.

[15] J.H. Kim, J. Pearl (1983). A computational model for causal and diagnostic reasoning in inference systems, *Proc. of the 8th IJCAI*, Karlsruhe, pp. 190 -193.

[16] J.N. Darroch, S.L. Lauritzen, T.P. Speed (1980). Markov fields and log-linear interaction models for contingency tables, *The Annals of Statistics*, vol. 8, pp. 522 - 539.

[17] F.V. Jensen, S.K. Andersen, U. Kjaerulff, S. Andreassen (1987). MUNIN - On the case for probabilities in medical expert systems - a practical exercise, in: J. Fox, M. Fieschi, R. Engelbrecht (eds.) *AIME 87*, Lecture Notes in Medical Informatics, vol. 33, Springer-Verlag, Berlin, pp. 149 - 160.

[18] S.L. Lauritzen, T.P. Speed, K. Vijayan (1984). Decomposable graphs and hypergraphs, *Journal of the Australian Mathematical Society (Series A)*, vol. 36, pp. 12 - 19.

[19] R.E. Tarjan, M. Yannakakis (1984). Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs and selectively reduce acyclic hypergraphs, *SIAM Journal of Computing*, vol. 13, no. 3, pp. 566 - 579.

[20] M. YANNAKAKIS (1981). Computing the minimum fill-in is NP-complete, *SIAM Journal of Algebraic Discrete Methods*, vol. 2, pp. 77 - 79.

[21] L.C. VAN DER GAAG (1989). *Computing Probabilities From a Partially Quantified Belief Network*, Centre for Mathematics and Computer Science (to appear).