

REPRESENTING CONDITIONAL INDEPENDENCE RELATIONS BY VALUATION NETWORKS

PRAKASH P. SHENOY

*School of Business
University of Kansas
Summerfield Hall*

*Lawrence, KS 66045-2003, USA
pshenoy@ukanvm.cc.ukans.edu*

Received 26 April 1993

Revised 1 March 1994

Valuation networks have been proposed as graphical representations of valuation-based systems. The axiomatic framework of valuation-based systems is able to capture many uncertainty calculi including probability theory, Dempster-Shafer's belief-function theory, Spohn's epistemic belief theory, and Zadeh's possibility theory. In this paper, we show how valuation networks encode conditional independence relations. For the probabilistic case, the class of probability models encoded by valuation networks includes undirected graph models, directed acyclic graph models, directed balloon graph models, and recursive causal graph models.

Keywords: Valuation networks, conditional independence, undirected graphs, Markov networks, directed acyclic graphs, Bayesian networks, directed balloon graphs, recursive causal graphs.

1. Introduction

Recently, we proposed valuation networks as a graphical representation of valuation-based systems [1, 2]. The axiomatic framework of valuation-based systems (VBS) is able to represent many different uncertainty calculi such as probability theory [2], Dempster-Shafer's belief-function theory [3], Spohn's epistemic belief theory [2, 4], and Zadeh's possibility theory [5]. The VBS framework is also flexible enough to include propositional logic [6, 7], discrete optimization [8], Bayesian decision analysis [9, 10], and constraint satisfaction [11]. The valuation network representation and solution techniques for Bayesian decision analysis are more expressive and computationally more efficient than influence diagram representation and solution techniques [12]. In this paper, we explore the use of valuation networks for representing conditional independence relations in probability theory and in other uncertainty theories that fit in the VBS framework.

Conditional independence has been widely studied in probability and statistics [see, for example, 13–18]. Pearl and Paz [19] have stated some basic properties of the conditional independence relation. (These properties are similar to those stated first by Dawid [13] for probabilistic conditional independence, those stated by Spohn [14] for causal independence, and those stated by Smith [17] for generalized conditional independence.) Pearl and Paz call these properties ‘graphoid axioms,’ and they call any ternary relation that satisfies these properties a ‘graphoid.’ The graphoid axioms are satisfied not only by conditional independence in probability theory, but also by vertex separation in undirected graphs (hence the term graphoids) [19], by d-separation in directed acyclic graphs [20], by partial correlation [19], by embedded multi-valued dependency models in relational databases [21], by conditional independence in Spohn’s theory of epistemic beliefs [22, 23], and by qualitative conditional independence [24]. Shenoy [25] has defined conditional independence in VBSs and shown that it satisfies the graphoid axioms. Thus the graphoid axioms are also satisfied by the conditional independence relations in all uncertainty theories that fit in the VBS framework including Dempster-Shafer’s belief-function theory and Zadeh’s possibility theory.

The use of undirected graphs and the use of directed acyclic graphs to represent conditional independence relations in probability theory have been extensively studied [see, for example, 15, 19, 26–36]. The use of graphs to represent conditional independence relations is useful since an exponential number of conditional independence statements can be represented by a graph with a polynomial number of vertices.

In undirected graphs (UGs), vertices represent variables, and edges between variables represent dependencies in the following sense. Suppose a , b , and c are disjoint subsets of variables. The conditional independence statement ‘ a is conditionally independent of b given c ,’ denoted by $a \perp b | c$, is represented in an UG if every path from a variable in a to a variable in b contains a variable in c , i.e., if c is a cut-set separating a and b . One can also represent a conditional independence relation by a set of UGs [37]. A conditional independence relation is represented by a set of UGs if each independence statement in the relation is represented in one of the UGs in the set. In general, one may not be able to represent a conditional independence relation that holds in a probability distribution by one UG. Some probability distributions may require an exponential number of UGs to represent the conditional independence relation that holds in it [38].

In directed acyclic graphs (DAGs), vertices represent variables, and arcs represent dependencies in the following sense. Pearl [39] has defined d-separation of two sets of variables by a third. Suppose a , b , and c are disjoint subsets of variables. We say c d-separates a and b iff there is no path from a variable in a to a variable in b along which (1) every vertex with an outgoing arc is not in c , and (2) every vertex with incoming arcs is either in c or has a descendant in c . The definition of d-separation takes into account the direction of the arcs in a DAG. The conditional independence statement $a \perp b | c$ is represented in a DAG if c d-separates a and b . One can also represent conditional independence relations by a set of DAGs [40]. A conditional independence relation is represented by a set of DAGs if it is represented in one of the DAGs in the set. As in the case of UGs, one may not be able to represent a conditional independence relation that holds in a probability dis-

tribution by one DAG. Some probability distributions may require an exponential number of DAGs to represent the conditional independence relations that hold in it [38].

Shafer [41] has defined directed balloon graphs (DBGs) that generalize DAGs. A DBG includes a partition of the set of all variables. Each element of the partition is called a balloon. Each balloon has a set of variables as its parents. The parents of a balloon are shown by directed arcs pointing to the balloon. A DBG is acyclic in the same sense that DAGs are acyclic. A DBG implies a probability model consisting of a conditional for each balloon given its parents. A DAG may be considered as a DBG in which each balloon is a singleton subset. Independence properties of DBGs are studied in [42].

UGs and DAGs represent conditional independence relations in fundamentally different ways. There are UGs such that the conditional independence relation represented in an UG cannot be represented by one DAG. And there are DAGs such that the conditional independence relation represented in a DAG cannot be represented by one UG. In fact, Ur and Paz [43] have shown that there is an UG such that to represent the conditional independence relation in it requires an exponential number of DAGs. And there is a DAG such that to represent the conditional independence relation in it requires an exponential number of UGs.

In valuation networks (VNs), there are two types of vertices. One set of vertices represents variables, and the other set represents valuations. Valuations are functions defined on variables. In probability theory, for example, a valuation is a factor of the joint probability distribution. In VNs, there are edges only between variables and valuations. There is an edge between a variable and a valuation if and only if the variable is in the domain of the valuation. If a valuation is a conditional for r given t , then we represent this by making the edges between the conditional and variables in r directed (pointed toward the variables). (Conditionals are defined in Section 2 and correspond to conditional probability distributions in probability theory.) Thus VNs explicitly depict a factorization of the joint valuation. Since there is a one-to-one correspondence between a factorization of the joint valuation and the conditional independence relation that holds in it, VNs also explicitly represent conditional independence relations.

The class of probability models included by VNs include UGs, DAGs and DBGs. Given a UG, there is a corresponding VN such that the conditional independence relation represented in the UG is represented in the VN. Given a DAG, there is a corresponding VN such that the conditional independence relation represented in the DAG is represented in the corresponding VN. And given a DBG, there is a corresponding VN such that the conditional independence relation represented in the DBG is represented in the corresponding VN.

Besides UGs, DAGs, and DBGs, there are other graphical models of probability distributions. Kiiveri, Speed, and Carlin [31] have defined recursive causal graphs (RCGs) that generalize DAGs and UGs. Recursive causal graphs have two components, an UG on one subset of variables (exogenous), and a DAG on another subset of variables (endogenous). Given a RCG, there is a corresponding VN such that the conditional independence relation represented in the UG is represented in the VN.

Lauritzen and Wermuth [33] and Wermuth and Lauritzen [30] have defined chain graphs that generalize recursive causal graphs. Conditional independence properties of chain graphs have been studied by Frydenberg [34]. It is not clear to this author whether VNs include the class of probability models captured by chain graphs.

Jirousek [44] has defined decision tree models of probability distributions. These models are particularly expressive for asymmetric conditional independence relations, i.e., relations that only hold for some configurations of the given variables, and not true for others. VNs, as defined here, do not include the class of models captured by decision trees.

Heckerman [45] has defined similarity networks as a tool for knowledge acquisition. Like Jirousek's decision tree models, similarity networks allow representations of asymmetric conditional independence relations. VNs, as defined here, do not include the class of models captured by similarity networks.

An outline of this paper is as follows. Section 2 describes the axiomatic framework of valuation-based systems and the general definition of conditional independence in VBSs. The definition of conditional independence in VBS is a generalization of the definition of conditional independence in probability theory. Most of the material in this section is a summary of [25]. Section 3 describes the valuation network representation and shows how conditional independence relations are encoded in valuation networks. Section 4 compares valuation networks to undirected graphs, directed acyclic graphs, balloon graphs, and recursive causal graphs. Finally, Section 5 contains some concluding remarks.

2. Valuation-Based Systems and Conditional Independence

In this section, we describe the axiomatic framework of valuation-based systems (VBSs). Most of the material in this section is taken from [25]. A similar axiomatic framework is defined by Cano, Delgado and Moral [46]. To help the reader follow the abstract exposition, we will illustrate all definitions using probability theory as an example. In [25], there are other examples given such as Dempster-Shafer's belief-function theory, Spohn's epistemic beliefs theory, and Zadeh's possibility theory.

In the VBS framework, we represent knowledge by entities called variables and valuations. We infer conditional independence relations using three operations called combination, marginalization, and removal. We use these operations on valuations.

Variables. We assume there is a finite set $\mathfrak{X}$ whose elements are called *variables*. Variables are denoted by upper-case Latin alphabets, X, Y, Z , etc. Subsets of $\mathfrak{X}$ are denoted by lower-case Latin alphabets, r, s, t , etc.

Valuations. For each $s \subseteq \mathfrak{X}$, there is a set $\mathcal{V}_s$. We call the elements of $\mathcal{V}_s$ *valuations for s* . Let $\mathcal{V}$ denote $\cup \{ \mathcal{V}_s \mid s \subseteq \mathfrak{X} \}$, the set of all *valuations*. If $\sigma \in \mathcal{V}_s$, then we say s is the *domain of σ* . Valuations are denoted by lower-case Greek alphabets, ρ, σ, τ , etc.

Valuations are primitives in our abstract framework and, as such, require no definition. But as we shall see shortly, they are entities that can be combined, marginalized, and removed. Intuitively, a valuation for s represents some knowledge about variables in s .

In probability theory, with each variable X , we associate a finite set $\mathcal{W}_X$ called the *frame of X* . For each $s \subseteq \mathcal{X}$, we let denote $\mathcal{W}_s = \times \{ \mathcal{W}_X \mid X \in s \}$ and we call $\mathcal{W}_s$ the *frame of s* . We call elements of $\mathcal{W}_s$ *configurations of s* , and we use lower-case bold-faced letters such as x, y , etc., to denote configurations. A valuation σ for s is a function $\sigma: \mathcal{W}_s \rightarrow [0, 1]$.

Zero Valuations. For each $s \subseteq \mathcal{X}$, there is at most one valuation $\zeta_s \in \mathcal{V}_s$ called the *zero valuation for s* . Let $\mathcal{Z}$ denote $\{ \zeta_s \mid s \subseteq \mathcal{X} \}$, the set of all *zero valuations*. Notice that we are not assuming zero valuations always exist. If zero valuations do not exist, $\mathcal{Z} = \emptyset$. We call valuations in $\mathcal{V} - \mathcal{Z}$ *nonzero valuations*.

Intuitively, a zero valuation represents knowledge that is internally inconsistent, i.e., knowledge that is a contradiction, or knowledge whose truth value is always false. The concept of zero valuations is important in the theory of consistent knowledge-based systems [7].

In probability theory, a zero valuation is a valuation whose values are identically zero, i.e., $\zeta_s(x) = 0$ for all $x \in \mathcal{W}_s$.

Proper Valuations. For each $s \subseteq \mathcal{X}$, there is a subset $\mathcal{P}_s$ of $\mathcal{V}_s - \{ \zeta_s \}$. We call the elements of $\mathcal{P}_s$ *proper valuations for s* . Let $\mathcal{P}$ denote $\cup \{ \mathcal{P}_s \mid s \subseteq \mathcal{X} \}$, the set of all *proper valuations*. Intuitively, a proper valuation represents knowledge that is partially coherent. By coherent knowledge, we mean knowledge that has well-defined semantics.

The concept of proper valuations has substance (i.e., $\mathcal{P}_s$ is a proper subset of $\mathcal{V}_s - \{ \zeta_s \}$) only in Dempster-Shafer's belief-function theory. In Dempster-Shafer's belief-function theory, a valuation for s is a function from the power set of the frame for s to the unit interval $[0, 1]$, and a proper valuation is an unnormalized commonality function. In probability theory, Spohn's epistemic-belief theory, and Zadeh's possibility theory, $\mathcal{P}_s = \mathcal{V}_s - \{ \zeta_s \}$. Proper valuations play no role either in the definition, or in the characterizations, or in the properties of conditional independence. The only role of proper valuations is in the semantics of knowledge.

Normal Valuations. For each $s \subseteq \mathcal{X}$, there is another subset $\mathcal{N}_s$ of $\mathcal{V}_s - \{ \zeta_s \}$. We call the elements of $\mathcal{N}_s$ *normal valuations for s* . Let $\mathcal{N}$ denote $\cup \{ \mathcal{N}_s \mid s \subseteq \mathcal{X} \}$, the set of all *normal valuations*. Intuitively, a normal valuation represents knowledge that is also partially coherent, but in a sense that is different from proper valuations. Normal valuations play an important role in the definition and characterization of conditional independence.

In probability theory, normal valuations are valuations whose values add to 1, i.e., a valuation σ for s is normal iff $\sum \{ \sigma(x) \mid x \in \mathcal{W}_s \} = 1$.

We call the elements of $\mathcal{P} \cap \mathcal{N}$ *proper normal valuations*. Intuitively, a proper normal valuation represents knowledge that is completely coherent, i.e., knowledge that has well-defined semantics. In probability theory, proper normal valuations correspond to probability mass functions.

Combination. We assume there is a mapping $\oplus: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{N} \cup \mathcal{Z}$, called *combination*, that satisfies the following four axioms:

Axiom C1 (Domain): If $\rho \in \mathcal{V}_r$ and $\sigma \in \mathcal{V}_s$, then $\rho \oplus \sigma \in \mathcal{V}_{r \cup s}$;

Axiom C2 (Associative): $\rho \oplus (\sigma \oplus \tau) = (\rho \oplus \sigma) \oplus \tau$;

Axiom C3 (Commutative): $\rho \oplus \sigma = \sigma \oplus \rho$; and

Axiom C4 (Zero): Suppose zero valuations exist, and suppose $\sigma \in \mathcal{V}_s$. Then $\zeta_r \oplus \sigma = \zeta_{r \cup s}$.

If $\rho \oplus \sigma$, read as ρ plus σ , is a zero valuation, then we say that ρ and σ are *inconsistent*. If $\rho \oplus \sigma$ is a normal valuation, then we say that ρ and σ are *consistent*.

Intuitively, combination corresponds to aggregation of knowledge. If ρ and σ are valuations for r and s representing knowledge about variables in r and s , respectively, then $\rho \oplus \sigma$ represents the aggregated knowledge about variables in $r \cup s$.

In probability theory, combination is pointwise multiplication followed by normalization (if normalization is possible). If $x \in \mathcal{W}_{r \cup s}$, let $x^{\downarrow r}$ denote the projection of x to r , where projection simply means dropping extra coordinates. Suppose $\rho \in \mathcal{V}_r$ and $\sigma \in \mathcal{V}_s$. Let $K = \sum \{\rho(x^{\downarrow r})\sigma(x^{\downarrow s}) \mid x \in \mathcal{W}_{r \cup s}\}$. The *combination of ρ and σ* is the valuation for $r \cup s$ given by

$$(\rho \oplus \sigma)(x) = \begin{cases} K^{-1} \rho(x^{\downarrow r}) \sigma(x^{\downarrow s}) & \text{if } K > 0 \\ 0 & \text{if } K = 0 \end{cases}$$

for all $x \in \mathcal{W}_{r \cup s}$. If $K = 0$, $\rho \oplus \sigma = \zeta_{r \cup s}$. If $K > 0$, then K is a normalization constant that ensures $\rho \oplus \sigma$ is a normal valuation. It is easy to see that this definition satisfies Axioms C1–C4.

An implication of Axiom C2 is that when we have multiple combinations of valuations, we can write it without using parenthesis. For example, $((\sigma_1 \oplus \sigma_2) \oplus \sigma_3) \oplus \dots \oplus \sigma_m$ can be written simply as $\sigma_1 \oplus \dots \oplus \sigma_m$ without parenthesis. Further, by Axiom C3, we can write $\sigma_1 \oplus \dots \oplus \sigma_m$ simply as $\oplus\{\sigma_1, \dots, \sigma_m\}$, i.e., not only do we not need parenthesis, we need not indicate the order in which the valuations are combined.

An implication of Axioms C1, C2, and C3 is that the set $\mathcal{N}_s \cup \{\zeta_s\}$ together with the combination operation $\oplus$ is a commutative semigroup [47]. (If zero valuations do not exist, then $\mathcal{N}_s \cup \{\zeta_s\} = \mathcal{N}_s$.) If zero valuations exist, then Axiom C4 defines the valuation ζ_s as the zero of the semigroup $\mathcal{N}_s \cup \{\zeta_s\}$.

Identity Valuations. We assume the following identity axiom.

Axiom C5 (Identity): For each $s \subseteq \mathcal{X}$, the commutative semigroup $\mathcal{N}_s \cup \{\zeta_s\}$ has an identity denoted by ι_s .

In other words, Axiom C5 assumes there exists $\iota_s \in \mathcal{N}_s \cup \{\zeta_s\}$ such that for each $\sigma \in \mathcal{N}_s \cup \{\zeta_s\}$, $\sigma \oplus \iota_s = \sigma$. Notice that a commutative semigroup may have at most one identity. From Axiom C4, it follows that $\iota_s \neq \zeta_s$, therefore $\iota_s \in \mathcal{N}_s$. Intuitively, identity valuations represent knowledge that is completely vacuous, i.e., they have no substantive content. In probability theory, the identity ι_s for $\mathcal{N}_s \cup \{\zeta_s\}$ is given by $\iota_s(x) = 1/|\mathcal{W}_s|$ for all $x \in \mathcal{W}_s$.

It follows from Axiom C5 that for each $s \subseteq \mathcal{X}$, and for each $\sigma \in \mathcal{N}_s \cup \{\zeta_s\}$, there exists at least one identity for it, i.e., there exists a $\delta_\sigma \in \mathcal{N}_s \cup \{\zeta_s\}$ such that $\sigma \oplus \delta_\sigma = \sigma$. For example, ι_s is an identity for each element of $\mathcal{N}_s \cup \{\zeta_s\}$. A valuation may have more than one identity. For example, Axiom C4 states that every element of $\mathcal{N}_s \cup \{\zeta_s\}$ is an identity for ζ_s . Notice that if $\sigma \in \mathcal{N}_s$, then $\delta_\sigma \in \mathcal{N}_s$. Also, notice that ι_s has only one identity, namely itself.

Positive Normal Valuations. Let $\mathcal{U}_s$ denote the subset of $\mathcal{N}_s$ consisting of all valuations in $\mathcal{N}_s$ that have unique identities. We call elements of $\mathcal{U}_s$ *positive normal valuations* for s . Let $\mathcal{U}$ denote $\cup\{\mathcal{U}_s \mid s \subseteq \mathcal{X}\}$, the set of all *positive normal valuations*. The concept of positive normal valuations is important because the intersection property of conditional independence holds only for positive normal valuations. Positive normal valuations correspond to strictly positive probability distributions in probability theory. Figure 1 shows the relation between different types of valuations.

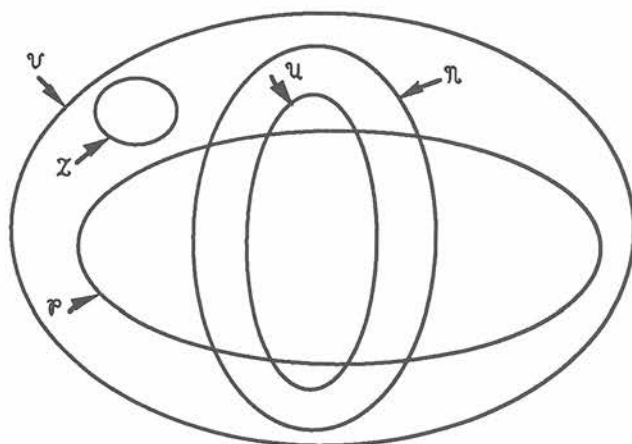


Figure 1. The relation between different types of valuations.

We assume the following axiom regarding the normal valuation for the empty set.

Axiom C6 (Normal Valuations for the Empty Set): The set $\mathcal{N}_\emptyset$ consists of exactly one element.

Axiom C6 implies that $\mathcal{U}_\emptyset = \mathcal{N}_\emptyset = \{\iota_\emptyset\}$ where $\iota_\emptyset$ is the identity valuation for the semigroup $\mathcal{N}_\emptyset \cup \{\zeta_\emptyset\}$.

Marginalization. We assume that for each nonempty $s \subseteq \mathcal{X}$, and for each $X \in s$, there is a mapping $\downarrow_{(s-\{X\})}: \mathcal{V}_s \rightarrow \mathcal{V}_{s-\{X\}}$, called *marginalization to $s-\{X\}$* , that satisfies the following six axioms:

Axiom M1 (Order of Deletion): Suppose $\sigma \in \mathcal{V}_s$, and suppose $X, Y \in s$. Then $(\sigma \downarrow_{(s-\{X\})}) \downarrow_{(s-\{X,Y\})} = (\sigma \downarrow_{(s-\{Y\})}) \downarrow_{(s-\{X,Y\})}$;

Axiom M2 (Zero): If zero valuations exist, then $\zeta_s \downarrow_{(s-\{X\})} = \zeta_{s-\{X\}}$;

Axiom M3 (Normal): $\sigma \downarrow_{(s-\{X\})} \in \mathcal{N}$ if and only if $\sigma \in \mathcal{N}$;

Axiom M4 (Positive Normal): If $\sigma \in \mathcal{U}$, then $\sigma^{\downarrow(s-\{X\})} \in \mathcal{U}$;

Axiom CM1 (Combination and Marginalization 1): Suppose $\rho \in \mathcal{V}_r$ and $\sigma \in \mathcal{V}_s$. Suppose $X \notin r$, and $X \in s$. Then $(\rho \oplus \sigma)^{\downarrow((r \cup s) - \{X\})} = \rho \oplus (\sigma^{\downarrow(s - \{X\})})$; and

Axiom CM2 (Combination and Marginalization 2): Suppose $\sigma \in \mathcal{V}_s$, suppose $r \subseteq s$, and suppose $\delta_{\sigma \downarrow r}$ is an identity for $\sigma^{\downarrow r}$ in $\mathcal{V}_r$. Then $\delta_{\sigma \downarrow r}$ is an identity for σ , i.e., $\sigma \oplus \delta_{\sigma \downarrow r} = \sigma$.

We call $\sigma^{\downarrow(s-\{X\})}$ the *marginal of σ for $s-\{X\}$* .

Intuitively, marginalization corresponds to coarsening of knowledge. If σ is a valuation for s representing some knowledge about variables in s , and $X \in s$, then $\sigma^{\downarrow(s-\{X\})}$ represents the knowledge about variables in $s-\{X\}$ implied by σ if we disregard variable X .

In probability theory, marginalization is addition. Suppose $\sigma \in \mathcal{V}_s$, and $X \in s$. The *marginal of σ for $s-\{X\}$* is the valuation for $s-\{X\}$ defined as follows:

$$\sigma^{\downarrow(s-\{X\})}(y) = \sum \{ \sigma(y, x) \mid x \in \mathcal{W}_X \}$$

for all $y \in \mathcal{W}_{s-\{X\}}$.

If we regard marginalization as a coarsening of a valuation by deleting variables, then Axiom M2 says that the order in which the variables are deleted does not matter. One implication of this axiom is that $(\sigma^{\downarrow(s-\{X\})})^{\downarrow(s-\{X, Y\})}$ can be written simply as $\sigma^{\downarrow(s-\{X, Y\})}$, i.e., we need not indicate the order in which the variables are deleted.

Axioms M2, M3 and M4 state that marginalization preserves the coherence of knowledge. An implication of Axiom M4 is that a valuation σ for s is normal if and only if $\sigma^{\downarrow \emptyset} = \iota_{\emptyset}$.

Axiom CM1 states that the computation of $(\rho \oplus \sigma)^{\downarrow((r \cup s) - \{X\})}$ can be accomplished without having to compute $\rho \oplus \sigma$. The combination $\rho \oplus \sigma$ is a valuation for $r \cup s$ whereas the combination $\rho \oplus (\sigma^{\downarrow(s-\{X\})})$ is a valuation for $(r \cup s) - \{X\}$. Axiom CM2 states an important property of identity valuations. It follows from Axiom CM2 that $\iota_s \oplus \iota_r = \iota_{r \cup s}$. Also, if $r \subseteq s$, then $\iota_s^{\downarrow r} = \iota_r$ [25].

Axioms C1, C2, C3, M1, and CM1 make local computation of marginals possible. Suppose $\{\sigma_1, \dots, \sigma_m\}$ is a collection of valuations, and suppose $\sigma_i \in \mathcal{V}_{s_i}$. Suppose $\mathfrak{X} = s_1 \cup \dots \cup s_m$, and suppose $X \in \mathfrak{X}$. Suppose we wish to compute $(\sigma_1 \oplus \dots \oplus \sigma_m)^{\downarrow \{X\}}$. We can do so by successively deleting all variables but X from the collection of valuation $\{\sigma_1, \dots, \sigma_m\}$. Each time we delete a variable, we do a fusion operation defined as follows. Consider a set of k valuations $\rho_1, \dots, \rho_k$. Suppose $\rho_i \in \mathcal{V}_{r_i}$. Let $\text{Fus}_Y\{\rho_1, \dots, \rho_k\}$ denote the collection of valuations after fusing the valuations in the set $\{\rho_1, \dots, \rho_k\}$ with respect to variable $Y \in r_1 \cup \dots \cup r_k$. Then

$$\text{Fus}_Y\{\rho_1, \dots, \rho_k\} = \{ \rho^{\downarrow(r-\{Y\})} \mid \rho_i \mid Y \notin r_i \}$$

where $\rho = \oplus \{ \rho_i \mid Y \in r_i \}$, and $r = \cup \{ r_i \mid Y \in r_i \}$. After fusion, the set of valuations is changed as follows. All valuations whose domains include Y are combined, and the resulting valuation is marginalized such that Y is eliminated from its domain. The valuations

whose domains do not include Y remain unchanged. The following lemma describes an important consequence of the fusion operation, and it follows directly from Axiom CM1.

Lemma 2.1 [2]. Suppose $\{\rho_1, \dots, \rho_k\}$ is a collection of valuations such that $\rho_i \in \mathcal{V}_{r_i}$. Suppose Axioms C1, C2, C3, and CM1 hold. Let $\mathfrak{X}$ denote $r_1 \cup \dots \cup r_k$. Suppose $Y \in \mathfrak{X}$. Then $\oplus \text{Fus}_Y\{\rho_1, \dots, \rho_k\} = (\rho_1 \oplus \dots \oplus \rho_k)^{\downarrow(\mathfrak{X} - \{Y\})}$.

Next, we define another binary operation called removal. The removal operation is an inverse of the combination operation.

Removal. We assume there is a mapping $\ominus: \mathcal{V} \times (\mathcal{N} \cup \mathcal{Z}) \rightarrow (\mathcal{N} \cup \mathcal{Z})$, called *removal*, that satisfies the following three axioms:

Axiom R1 (Domain): Suppose $\sigma \in \mathcal{V}_s$, and $\rho \in \mathcal{N}_{r \cup \mathcal{Z}_r}$. Then $\sigma \ominus \rho \in \mathcal{N}_{r \cup s \cup \mathcal{Z}_{r \cup s}}$.

Axiom CR1 (Combination and Removal 1): For each $\rho \in \mathcal{N} \cup \mathcal{Z}$, there exists an identity for ρ , denoted by, say, ι_ρ , such that $\rho \ominus \rho = \iota_\rho$; and

Axiom CR2 (Combination and Removal 2): Suppose $\pi, \theta \in \mathcal{V}$, and $\rho \in \mathcal{N} \cup \mathcal{Z}$. Then, $(\pi \oplus \theta) \ominus \rho = \pi \oplus (\theta \ominus \rho)$.

We call $\sigma \ominus \rho$, read as σ minus ρ , the *valuation resulting after removing ρ from σ* . Intuitively, $\sigma \ominus \rho$ can be interpreted as follows. If σ and ρ represent some knowledge, and if we remove the knowledge represented by ρ from σ , then $\sigma \ominus \rho$ describes the knowledge that remains.

In probability theory, removal is division followed by normalization (if normalization is possible). Division by zero can be defined arbitrarily. For example, we define division of any real number by zero as resulting in zero. Suppose $\sigma \in \mathcal{V}_s$, and $\rho \in \mathcal{N}_{r \cup \mathcal{Z}_r}$. Let $K = \sum \{\sigma(x^{\downarrow s})/\rho(x^{\downarrow r}) \mid x \in \mathcal{W}_{r \cup s} \text{ such that } \rho(x^{\downarrow r}) > 0\}$. Then the valuation resulting from the removal of ρ from σ is the valuation for $r \cup s$ given by

$$(\sigma \ominus \rho)(x) = \begin{cases} K^{-1} \sigma(x^{\downarrow s}) / \rho(x^{\downarrow r}) & \text{if } K > 0 \text{ and } \rho(x^{\downarrow r}) > 0 \\ 0 & \text{if } K = 0 \text{ or } \rho(x^{\downarrow r}) = 0 \end{cases}$$

for all $x \in \mathcal{W}_{r \cup s}$.

Axioms CR1 and CR2 define the removal operation as an “inverse” of the combination operation in the sense that arithmetic division is inverse of arithmetic multiplication, and in the sense that arithmetic subtraction is inverse of arithmetic multiplication.

Conditionals. Suppose $\sigma \in \mathcal{N}_s$, and suppose a and b are disjoint subsets of s . The valuation $\sigma^{\downarrow(a \cup b)} \ominus \sigma^{\downarrow a}$ for $a \cup b$ plays an important role in the theory of conditional independence. Borrowing terminology from probability theory, we call $\sigma^{\downarrow(a \cup b)} \ominus \sigma^{\downarrow a}$ the *conditional for b given a with respect to σ* . Let $\sigma(b|a)$ denote $\sigma^{\downarrow(a \cup b)} \ominus \sigma^{\downarrow a}$. We call b the *head* of the domain of $\sigma(b|a)$, and we call a the *tail* of the domain of $\sigma(b|a)$. Also, if $a = \emptyset$, let $\sigma(b)$ denote $\sigma(b|\emptyset)$. The following theorem states some important properties of conditionals.

Theorem 2.1 [25]. Suppose $\sigma \in \mathcal{N}_s$, and suppose a, b , and c are disjoint subsets of s .

- (i). $\sigma(a) = \sigma \downarrow_a$.
- (ii). $\sigma(a) \oplus \sigma(b|a) = \sigma(a \cup b)$.
- (iii). $\sigma(b|a) \oplus \sigma(c|a \cup b) = \sigma(b \cup c|a)$.
- (iv). Suppose $b' \subseteq b$. Then $\sigma(b|a) \downarrow_{(a \cup b')} = \sigma(b'|a)$.
- (v). $(\sigma(b|a) \oplus \sigma(c|a \cup b)) \downarrow_{(a \cup c)} = \sigma(c|a)$.
- (vi). $\sigma(b|a) \downarrow_a = \iota_{\sigma(a)}$.
- (vii). $\sigma(b|a) \in \mathcal{N}_{a \cup b}$.

Conditional Independence. Suppose $\tau \in \mathcal{N}_w$, and suppose r, s , and v are disjoint subsets of w . We say r and s are *conditionally independent given v with respect to τ* , written as $r \perp_\tau s | v$, if and only if $\tau(r \cup s \cup v) = \alpha_{r \cup v} \oplus \alpha_{s \cup v}$, where $\alpha_{r \cup v} \in \mathcal{V}_{r \cup v}$, and $\alpha_{s \cup v} \in \mathcal{V}_{s \cup v}$.

When it is clear that all conditional independence statements are with respect to τ , we simply say ' r and s are conditionally independent given v ' instead of ' r and s are conditionally independent given v with respect to τ ,' and use the simpler notation $r \perp s | v$ instead of $r \perp_\tau s | v$. Also, if $v = \emptyset$, we say ' r and s are independent' instead of ' r and s are conditionally independent given $\emptyset$ ' and use the simpler notation $r \perp s$ instead of $r \perp s | \emptyset$.

Shenoy [25] shows that the conditional independence relation generalizes the conditional independence relation in probability theory. In particular, all characterizations of it given by Dawid [13] (including the graphoid axioms) follow from the above definition.

3. Valuation Networks

In this section, we define a valuation network representation of a VBS and explain how a valuation network encodes conditional independence statements.

A valuation network (VN) consists of a four-tuple $\Gamma = \{\mathcal{X}, \mathcal{V}, \mathcal{E}, \mathcal{Q}\}$ such that

- (i) $\mathcal{X} \neq \emptyset$,
- (ii) $\mathcal{V} \neq \emptyset$,
- (iii) $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{X}$,
- (iv) $\mathcal{Q} \subseteq \mathcal{V} \times \mathcal{X}$,
- (v) $\cup\{\alpha | (\alpha, X) \in \mathcal{E} \cup \mathcal{Q} \text{ for some } X \in \mathcal{X}\} = \mathcal{V}$, and
- (vi) $\cup\{X | (\alpha, X) \in \mathcal{E} \cup \mathcal{Q} \text{ for some } \alpha \in \mathcal{V}\} = \mathcal{X}$.

We call the elements of $\mathcal{X}$ *vertices* and they represent variables. We call the elements of $\mathcal{V}$ *nodes* and they represent valuations. We call the elements of $\mathcal{E}$ *edges*, and they denote either domains of valuations, or tails of domains of conditionals. We call the elements of $\mathcal{Q}$ *arcs* and they denote the heads of domains of conditionals. Condition (v) requires each valuation $\alpha \in \mathcal{V}$ to have at least one variable in its domain. Condition (vi) requires each variable $X \in \mathcal{X}$ to be in the domain of at least one valuation in $\mathcal{V}$. The six conditions that define a VN are quite weak, but is necessarily so since, as we will see, the class of VNs includes a wide class of models.

In VNs, vertices are denoted by circles, nodes by diamonds, edges by lines joining the respective nodes and vertices, and arcs by a directed edge pointing to the corresponding vertex. When a VN contains conditionals, we assume that all conditionals are with respect to the valuation τ obtained by combining all valuations in the network, i.e., $\tau = \oplus \mathcal{V}$.

Example 1. Consider a VBS consisting of variables W , X , Y , and Z , and valuations α for $\{W, X\}$, β for $\{X, Y\}$, γ for $\{Y, Z\}$, and δ for $\{W, Z\}$. Figure 2 shows the VN for this VBS. Vertices (variables) are depicted by circles, nodes (valuations) are depicted by diamonds, and edges are depicted by lines. The edges (α, W) and (α, X) incident to node α indicate that $\{W, X\}$ is the domain of α .

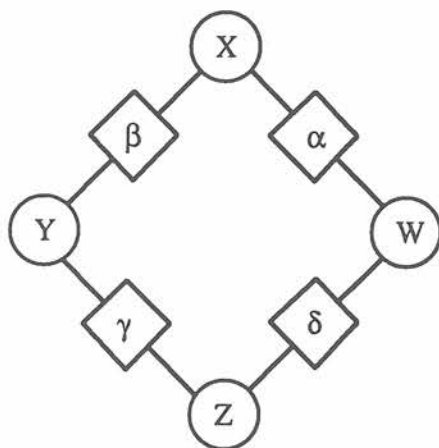


Figure 2. The VN for the VBS of Example 1.

Example 2. Consider a VBS consisting of variables V , W , X , Y , and Z , and conditionals α for $\{V\}$, β for $\{W\}$ given $\{V\}$, γ for $\{X\}$ given $\{V\}$, δ for $\{Y\}$ given $\{W, X\}$, and ϵ for $\{Z\}$ given $\{Y\}$. Figure 3 shows the VN for this VBS. The arc (δ, Y) incident to δ indicates that Y is the head of the domain of δ , and the edges (δ, W) and (δ, X) incident to δ indicate that $\{W, X\}$ is the tail of the domain of δ . Further, if τ denotes $\alpha \oplus \beta \oplus \gamma \oplus \delta \oplus \epsilon$, then $\alpha = \tau(V)$, $\beta = \tau(W | V)$, $\gamma = \tau(X | V)$, $\delta = \tau(Y | W, X)$, $\epsilon = \tau(Z | Y)$.¹

Example 3. Consider a VBS consisting of variables $X_1, \dots, X_{10}$, and conditionals α_1 for X_1 given $\emptyset$, α_2 for $\{X_2, X_3\}$ given X_1 , α_3 for X_4 given X_2 , α_4 for $\{X_5, X_6, X_7\}$ given X_2 , α_5 for X_8 given X_3 , α_6 for X_9 given X_5 , and α_7 for X_{10} given $\{X_6, X_7\}$. Figure 4 shows the VN for this VBS. If τ denotes $\alpha_1 \oplus \dots \oplus \alpha_7$, then $\alpha_1 = \tau(X_1)$, $\alpha_2 = \tau(X_2, X_3 | X_1)$, $\alpha_3 = \tau(X_4 | X_2)$, $\alpha_4 = \tau(X_5, X_6, X_7 | X_2)$, $\alpha_5 = \tau(X_8 | X_3)$, $\alpha_6 = \tau(X_9 | X_5)$, $\alpha_7 = \tau(X_{10} | X_6, X_7)$.

¹ For simplicity, we drop braces around subsets in conditional valuations. Thus we write $\tau(V)$ instead of $\tau(\{V\})$, $\tau(Y | W, X)$ instead of $\tau(\{Y\} | \{W, X\})$, etc.

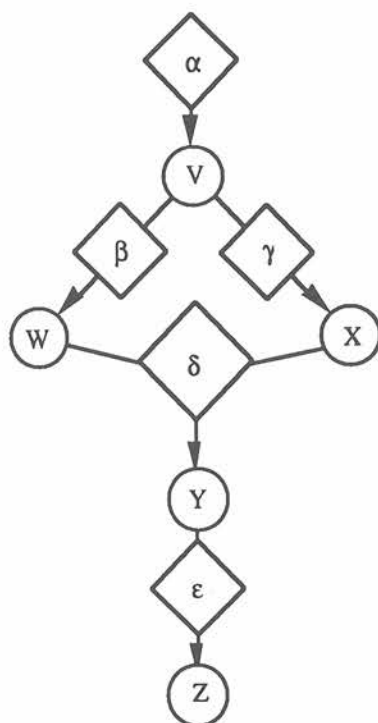


Figure 3. The VN for the VBS of Example 2.

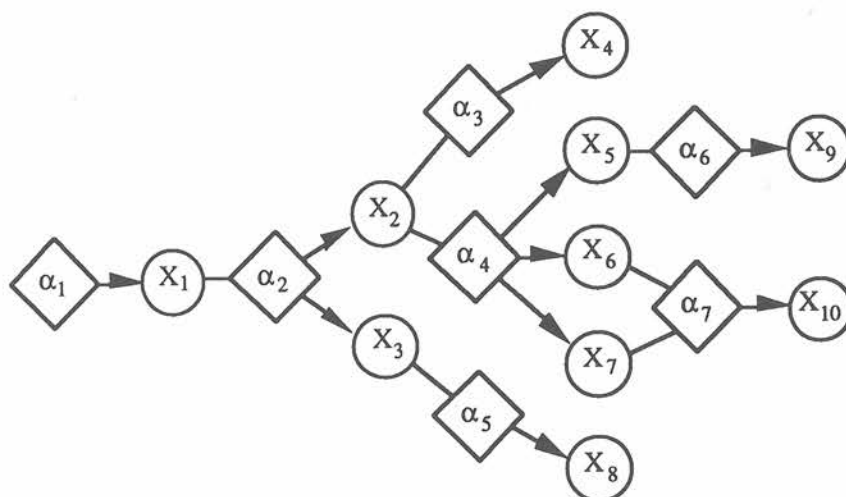


Figure 4. The VN for the VBS of Example 3.

Example 4. Consider a VBS consisting of variables V, W, X, Y and Z , valuations α for $\{V, W\}$, and β for $\{V, X\}$, and conditionals γ for Y given $\{W, X\}$, and δ for Z given X . Figure 5 shows the VN for this VBS. If τ denotes $\alpha \oplus \beta \oplus \gamma \oplus \delta$, then $\alpha \oplus \beta = \tau(V, W, X)$, $\gamma = \tau(Y | W, X)$, and $\delta = \tau(Z | X)$.

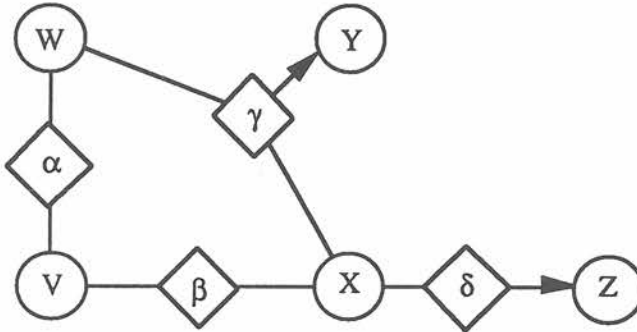


Figure 5. The VN for the VBS of Example 4.

Fusion in Valuation Networks. Next, we will illustrate fusion in VNs. Consider the VBS described in Example 1. If we fuse the valuations in the set $\{\alpha, \beta, \gamma, \delta\}$ with respect to X , we get $\text{Fus}_X(\alpha, \beta, \gamma, \delta) = \{(\alpha \oplus \beta)^{\downarrow\{Y, Z\}}, \gamma, \delta\}$. Figure 6 illustrates this fusion operation. Lemma 2.1 tells us that $(\alpha \oplus \beta)^{\downarrow\{Y, Z\}} \oplus \gamma \oplus \delta = (\alpha \oplus \beta \oplus \gamma \oplus \delta)^{\downarrow\{Y, Z, W\}}$.

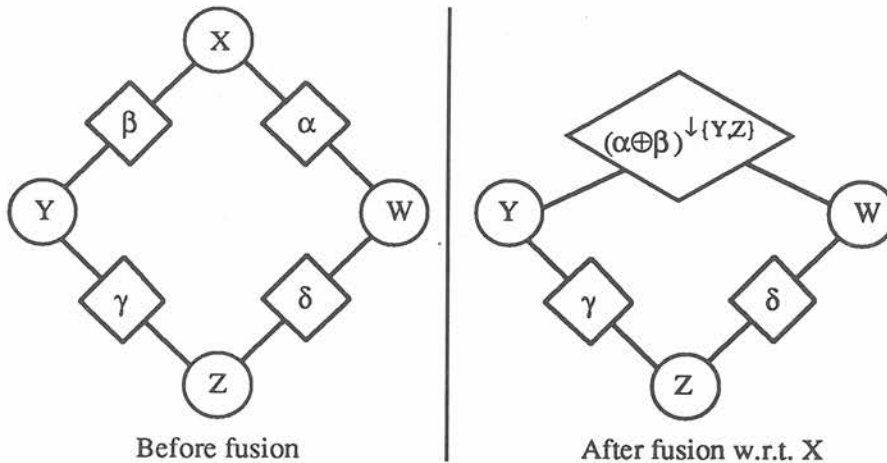


Figure 6. Fusion in VNs.

Next, we will illustrate fusion in VNs when we have conditionals. The various statements of Theorem 2.1 are useful here. Consider a VBS consisting of four variables W, X, Y , and Z , and three conditionals, α for W given $\emptyset$, β for X given W , and γ for

$\{Y, Z\}$ given X . Figure 7 shows the corresponding VN. Suppose $\tau = \alpha \oplus \beta \oplus \gamma$. Then $\alpha = \tau(W)$, $\beta = \tau(X|W)$, and $\gamma = \tau(Y, Z|X)$. After fusion with respect to X , we have two conditionals $\alpha = \tau(W)$, and $(\beta \oplus \gamma)^{\downarrow\{W, Y, Z\}} = \tau(Y, Z|W)$. This result is justified by statement (v) of Theorem 2.1. After further fusion with respect to Z , we only have two conditionals $\alpha = \tau(X)$, and $(\beta \oplus \gamma)^{\downarrow\{W, Y\}} = \tau(Y|W)$. This result is justified by statement (iv) of Theorem 2.1. Finally after further fusion with respect to Y , we have only one conditional $\alpha = \tau(W)$. This is because from statement (vi) of Theorem 2.1, $\tau(Y|W)^{\downarrow\{W\}}$ is an identity for $\tau(W)$, and this identity can be absorbed in any conditional that has W in the head of its domain.

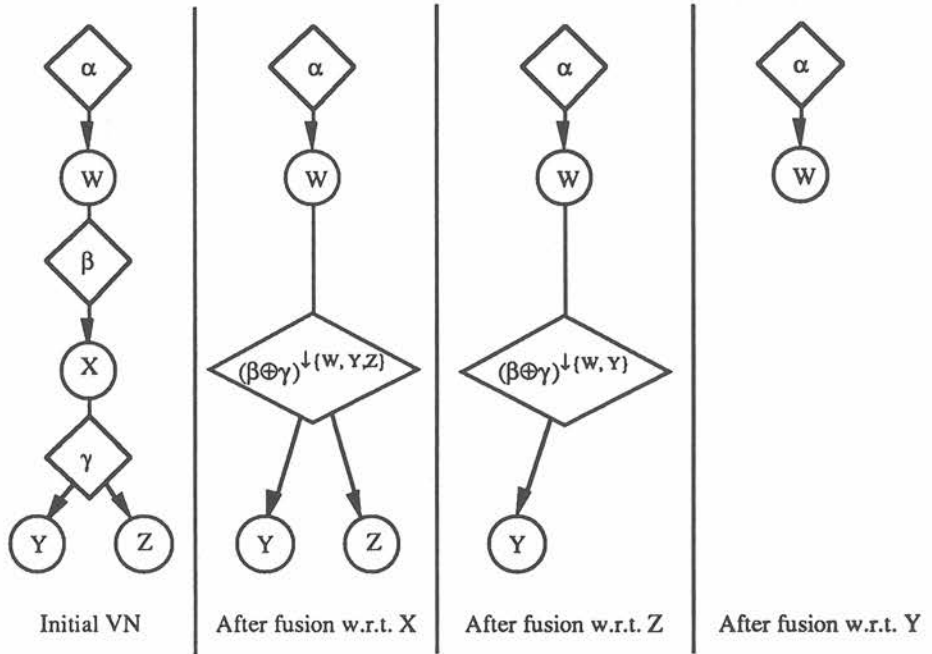


Figure 7. Fusion in VNs with conditionals.

Conditional Independence in Valuation Networks. How is conditional independence encoded in VNs? Let us examine the definition of conditional independence graphically. Suppose r , s , and v are disjoint subsets of variables, and suppose τ is a normal valuation for $r \cup s \cup v$. Our definition of conditional independence states that $\tau = \alpha_{r \cup v} \oplus \alpha_{s \cup v}$ iff $r \perp_{\tau} s | v$, where $\alpha_{r \cup v} \in \mathcal{V}_{r \cup v}$, and $\alpha_{s \cup v} \in \mathcal{V}_{s \cup v}$. Suppose $\tau = \alpha_{r \cup v} \oplus \alpha_{s \cup v}$ is a normal valuation for $r \cup s \cup v$, where $\alpha_{r \cup v} \in \mathcal{V}_{r \cup v}$, and $\alpha_{s \cup v} \in \mathcal{V}_{s \cup v}$. Figure 8 shows the VN representation of this situation. Notice that all paths from a variable in r to a variable in s go through a variable in v , i.e., v is a cut-set separating r from s . This suggests the following theorem.

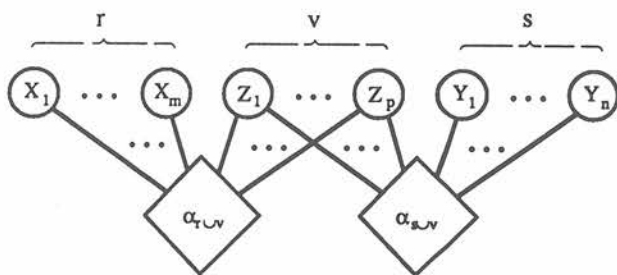


Figure 8. The VN representation of $r \perp_{\tau} s \mid v$, where $\tau = \alpha_{\tau \cup v} \oplus \alpha_{s \cup v}$.

Theorem 3.1. Suppose r , s , and v are disjoint subsets of w . Suppose $\tau \in \mathcal{N}_w$. Consider the VN representation of $\tau(r \cup s \cup v)$ after marginalizing all variables in $w - (r \cup s \cup v)$ out of τ . Suppose v is a cut-set separating r and s . Then $r \perp_{\tau} s \mid v$.

Proof. Suppose r , s , and v are disjoint subsets of w . Suppose $\tau \in \mathcal{N}_w$. Consider the VN representation of $\tau(r \cup s \cup v)$ after marginalizing all variables in $w - (r \cup s \cup v)$ out of τ . Suppose v is a cut-set separating r and s . Then there is no valuation that contains a variable in r and a variable in s . Consider all valuations whose domain includes a variable in r . Let ρ denote the combination of these valuations. Notice that the domain of ρ does not contain a variable in s . Now consider all valuations whose domain includes a variable in s . Let σ denote the combination of these valuations. Notice that the domain of σ does not include a variable in r . Finally let θ denote the combination of all valuations not included in either ρ or σ . Clearly, the domain of θ does not contain variables in either r or s . Since $\tau(r \cup s \cup v) = \rho \oplus \sigma \oplus \theta$, it follows from the definition of conditional independence that $r \perp_{\tau} s \mid v$. $\square$

A special case of the definition of conditional independence is when $v = \emptyset$. In this case, $r \perp_{\tau} s$ iff $\tau = \alpha_r \oplus \alpha_s$, where $\alpha_r \in \mathcal{V}_r$ and $\alpha_s \in \mathcal{V}_s$. Suppose $\tau = \alpha_r \oplus \alpha_s$, where $\alpha_r \in \mathcal{V}_r$ and $\alpha_s \in \mathcal{V}_s$. Figure 9 illustrates this situation. Notice that there is no path from a variable in r to a variable in s . As in the case of conditional independence, it is easy to show that if we have a VN for variables in $r \cup s$ such that there is no path from a variable in r to a variable in s , then $r \perp_{\tau} s$, where τ is the combination of all valuation in the VN.

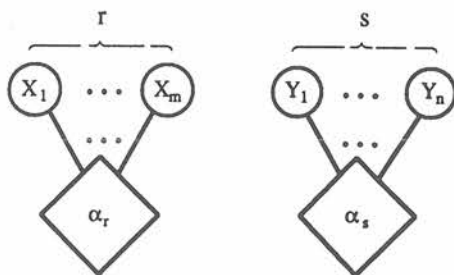


Figure 9. The VN representation of $r \perp_{\tau} s$, where $\tau = \alpha_r \oplus \alpha_s$.

To summarize, suppose we are given a VN representation of $\tau \in \mathcal{T}_W$. Suppose v is a cut-set separating r and s in the marginalized network for variables in $r \cup s \cup v$. Then $r \perp_{\tau} s \mid v$.

4. Comparison with UGs, DAGs, DBGs, and RCGs

In this section, we briefly compare VNs with UGs, DAGs, DBGs, and RCGs. We start with UGs.

In UGs, the cliques of the graph (maximal completely connected vertices) denote the factors of the joint valuation. For example, consider the UG shown in Figure 10. This graph has 4 cliques, $\{W, X\}$, $\{X, Y\}$, $\{Y, Z\}$, and $\{Z, X\}$. This undirected graph models a joint probability distribution for $\{W, X, Y, Z\}$ that factors (multiplicatively) into 4 components, α with domain $\{W, X\}$, β with domain $\{X, Y\}$, γ with domain $\{Y, Z\}$, and δ with domain $\{Z, W\}$. The VN representation of this distribution is also shown in Figure 10. Notice that for this distribution, $\{X\} \perp \{Z\} \mid \{Y, W\}$, and $\{Y\} \perp \{W\} \mid \{X, Z\}$.

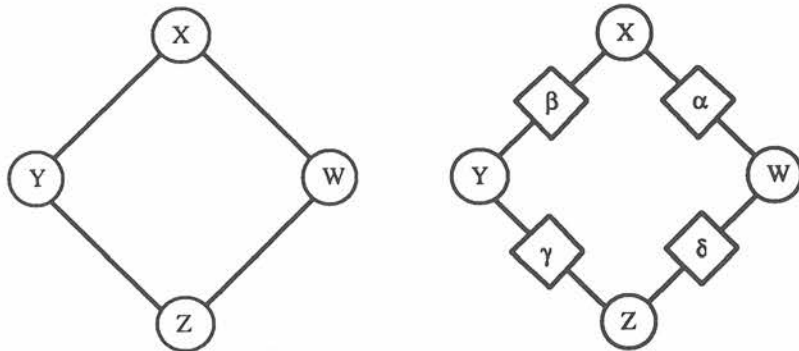


Figure 10. An UG and a corresponding VN.

Figure 11 shows another example of an undirected graph model of a probability distribution. In this UG, there is one clique consisting of $\{X, Y, Z\}$. The joint probability distribution corresponding to this UG does not decompose. The VN representation of this probability model is also shown in Figure 11. Notice that there are no non-trivial conditional independence relations in this distribution (the trivial ones being of the form $r \perp \emptyset$).

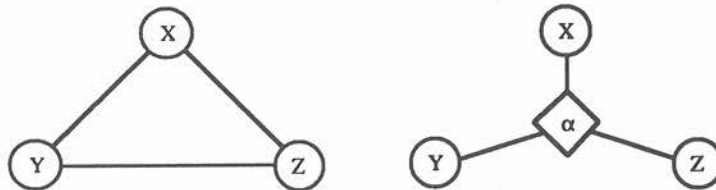


Figure 11. An UG and a corresponding VN.

Next, we consider DAGs. A DAG model of a probability distribution consists of an ordering of the variables, and a conditional for each variable given a subset of the variables that precede it in the given ordering. Figure 12 shows an example of a DAG with 5 variables. An ordering consistent with this DAG is VWXYZ. The DAG implies we have a conditional for V given $\emptyset$, a conditional for W given V, a conditional for X given V, a conditional for Y given {W, X}, and a conditional for Z given Y. The VN representation of the DAG model is also shown in Figure 12. Suppose τ denotes the joint probability distribution. Then $\alpha = \tau(V)$, $\beta = \tau(W | V)$, $\gamma = \tau(X | V)$, $\delta = \tau(Y | W, X)$, and $\varepsilon = \tau(Z | Y)$.

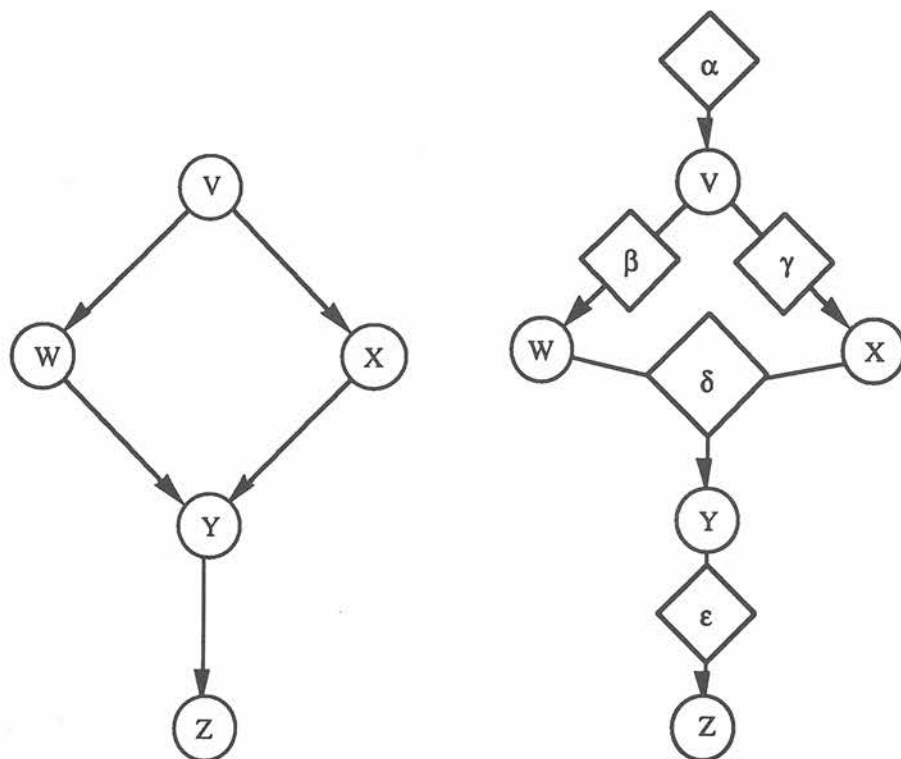


Figure 12. A DAG and a corresponding VN.

In the DAG of Figure 12, using Pearl's definition of d-separation, we cannot conclude, for example, that $W \perp_{\tau} X | \{V, Z\}$. However, we can conclude that $W \perp_{\tau} X | V$. We can draw the same conclusion using separation in VNs. If we fuse the VN with respect to Y, the resulting VN is shown in the left-hand side of Figure 13, and $\{V, Z\}$ is not a cut-set separating W and X. Therefore we cannot conclude that $W \perp_{\tau} X | \{V, Z\}$. If we further fuse the VN with respect to Z, the resulting VN is shown in the right-hand side of Figure 13, and V is a cut-set separating W and X. Therefore, $W \perp_{\tau} X | V$.

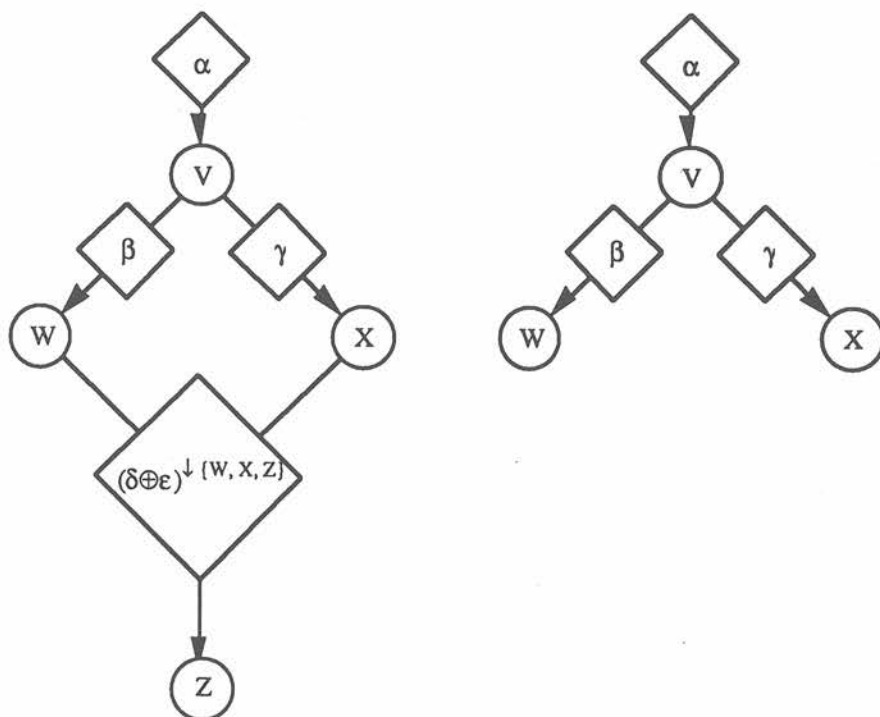


Figure 13. Left: After fusion with respect to Y (of the VN of Figure 12).
 Right: After further fusion with respect to Z .

The technique we have proposed for checking for conditional independence in VNs is an alternative to the d-separation method proposed by Pearl [39] for DAGs. Whether we have conditionals or not, checking a conditional independence statement in a VN is a matter of first fusing the VN to remove variables not in the conditional independence statement and then checking for separation in the fused VN. The information about conditionals is used in the fusion operation.

Lauritzen *et al.* [48] describe yet another method for checking for conditional independence in DAGs. Their method consists of converting a DAG to an equivalent UG and then checking for conditional independence in the UG using separation. In short, their method consists of examining a subgraph of the DAG (after eliminating the variables that succeed all variables in the conditional independence statement in an ordering consistent with the DAG), moralizing the graph, dropping directions, and then checking for separation.

Next, let us compare VNs and DBGs. DBGs are defined in [41]. A DBG includes a partition of the set of all variables. Each element of the partition is called a balloon. Non-singleton balloons are shown as ellipses encircling the corresponding variables. Each balloon has a set of variables as its parents. The parents of a balloon are shown by directed arcs pointing to the balloon. A DBG is acyclic in the same sense that DAGs are acyclic.

A DBG implies a probability model consisting of a conditional for each balloon given its parents. A DAG may be considered as a DBG in which each balloon is a singleton subset.

Figure 14 shows a DBG with 10 variables, $X_1, \dots, X_{10}$. There are two non-singleton balloons, $\{X_2, X_3\}$, and $\{X_5, X_6, X_7\}$. All other balloons are singleton subsets. The DBG of Figure 14 implies a conditional for X_1 given $\emptyset$, a conditional for $\{X_2, X_3\}$ given X_1 , a conditional for X_4 given X_2 , a conditional for $\{X_5, X_6, X_7\}$ given X_2 , a conditional for X_8 given X_3 , a conditional for X_9 given X_5 , and a conditional for X_{10} given $\{X_6, X_7\}$. The corresponding VN is also shown in Figure 14.

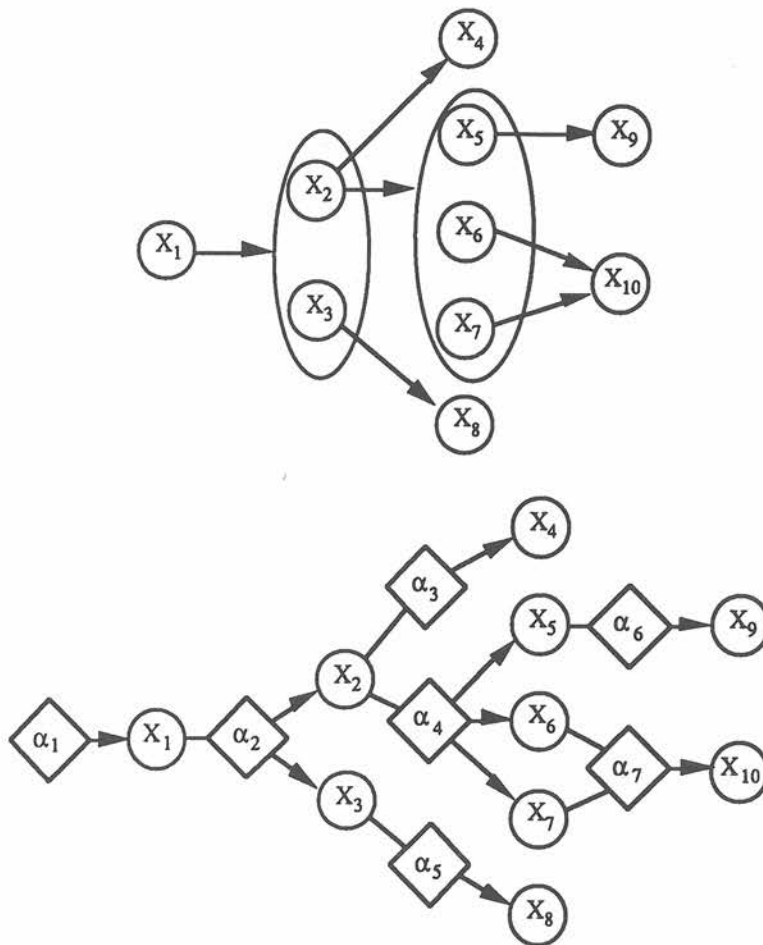


Figure 14. A DBG and a corresponding VN.

The conditional independence theory of DBGs is described in [42], and is analogous to the conditional independence theory of DAGs. In the DBG and VN of Figure 14, we have, for example, $\{X_5, X_6, X_7\} \perp \{X_1, X_3, X_4\} \mid \{X_2\}$.

Finally, we compare VNs to RCGs. RCGs are defined in [31]. A RCG consists of two kinds of vertices (variables)—exogenous and endogenous, and two kinds of edges—undirected and directed. An undirected edge always connects two exogenous variables, and a directed edge always points to an endogenous variable. RCGs generalize DAGs in the sense that a DAG is a RCG with at most one exogenous variable.

Figure 15 shows a RCG with five variables, V, W, X, Y, Z . Variables V, W , and X are exogenous, and variables Y and Z are endogenous. The cliques $\{V, W\}$ and $\{V, X\}$ imply valuations for $\{V, W\}$ and $\{V, X\}$ respectively. The directed edges pointing to Y imply a conditional for Y given $\{W, X\}$, and the directed edge pointing to Z implies a conditional for Z given X . The corresponding VN is also shown in Figure 15.

Conditional independence properties of RCG are given in [31]. Briefly, if we look at the subgraph of a RCG restricted to the exogenous variables, the subgraph is an UG and its conditional independence properties are the same as those given by the UG models. On the other hand, the conditional independence relation in the complete RCG is given by the d-separation relation of DAGs. Since the basis of the conditional independence relations in RCGs is the underlying factorization and the additional information about conditionals, and since this information is encoded in VNs, a corresponding VN encodes the same conditional independence relation as a RCG. For example, in the RCG and VN of Figure 15, we have $\{W\} \perp \{X\} \mid \{V\}$, and $\{Z\} \perp \{V, W, Y\} \mid \{X\}$.

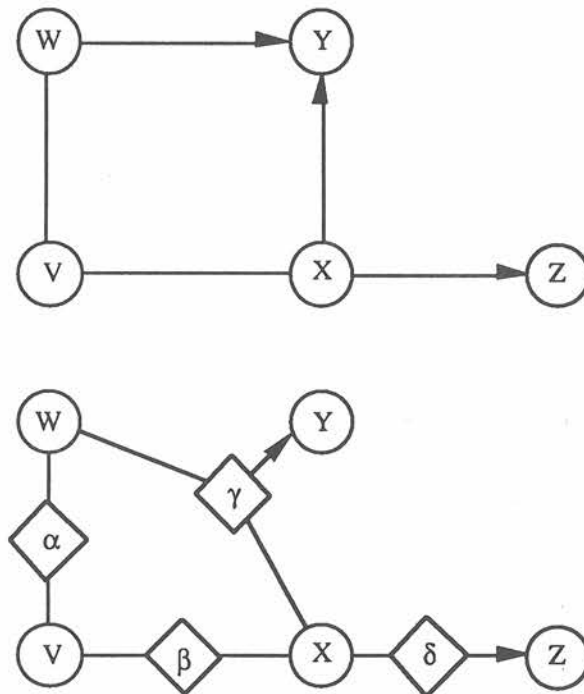


Figure 15. A RCG and a corresponding VN.

5. Conclusion

We have described valuation networks and how they encode conditional independence. Given a valuation network, $r \perp_{\tau} s \mid v$ if v is a cut-set separating r from s in the marginalized valuation network for $r \cup s \cup v$. We have compared valuation networks to undirected graphs, directed acyclic graphs, directed balloon graphs, and recursive causal graphs. All probability models encoded by one of these graphs can be represented by corresponding valuation networks.

Factorization is fundamental to conditional independence. The power of the valuation network representation arises from the fact that it represents factorization explicitly. Also notice that valuation networks encode conditional independence not only in probabilistic models, but also in all uncertainty theories that fit in the VBS framework. This includes Dempster-Shafer's belief-function theory, Spohn's epistemic belief theory, and Zadeh's possibility theory [25].

We have not compared valuation networks to chain graphs [30, 33], decision trees [44], and similarity networks [45]. This remains to be done.

Acknowledgments

This work is based upon work supported in part by the National Science Foundation under Grant No. SES-9213558, and in part by the General Research Fund of the University of Kansas under Grant No. 3605-XX-0038. This paper has benefited from comments by two anonymous referees.

References

1. Shenoy, P. P., "A valuation-based language for expert systems," *International Journal of Approximate Reasoning*, 3(5), 1989, 383-411.
2. Shenoy, P. P., "Valuation-based systems: A framework for managing uncertainty in expert systems," in L. A. Zadeh and J. Kacprzyk (eds.), *Fuzzy Logic for the Management of Uncertainty*, 1992, 83-104, John Wiley & Sons, New York, NY.
3. Shenoy, P. P., "Using Dempster-Shafer's belief-function theory in expert systems," in R. R. Yager, M. Fedrizzi and J. Kacprzyk (eds.), *Advances in the Dempster-Shafer Theory of Evidence*, 1994, 395-414, John Wiley and Sons, New York, NY.
4. Shenoy, P. P., "On Spohn's rule for revision of beliefs," *International Journal of Approximate Reasoning*, 5(2), 1991, 149-181.
5. Shenoy, P. P., "Using possibility theory in expert systems," *Fuzzy Sets and Systems*, 52(2), 1992, 129-142.
6. Shenoy, P. P., "Valuation-based systems for propositional logic," in Z. W. Ras, M. Zemankova and M. L. Emrich (eds.), *Methodologies for Intelligent Systems*, 5, 1990, 305-312, North-Holland, Amsterdam.
7. Shenoy, P. P., "Consistency in valuation-based systems," *ORSA Journal on Computing*, 6(3), 1994, in press.
8. Shenoy, P. P., "Valuation-based systems for discrete optimization," in P. P. Bonissone, M. Henrion, L. N. Kanal and J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*, 6, 1991, 385-400, North-Holland, Amsterdam.

9. Shenoy, P. P., "Valuation-based systems for Bayesian decision analysis," *Operations Research*, 40(3), 1992, 463-484.
10. Shenoy, P. P., "A new method for representing and solving Bayesian decision problems," in D. J. Hand (ed.), *Artificial Intelligence Frontiers in Statistics: AI and Statistics III*, 1993, 119-138, Chapman & Hall, London.
11. Shenoy, P. P. and G. Shafer, "Constraint propagation," Working Paper No. 208, 1988, School of Business, University of Kansas, Lawrence, KS.
12. Shenoy, P. P., "A comparison of graphical techniques for decision analysis," *European Journal of Operational Research*, 1994, in press.
13. Dawid, A. P., "Conditional independence in statistical theory (with discussion)," *Journal of the Royal Statistical Society, Series B*, 41(1), 1979, 1-31.
14. Spohn, W., "Stochastic independence, causal independence, and shieldability," *Journal of Philosophical Logic*, 9, 1980, 73-99.
15. Lauritzen, S. L., *Lectures on Contingency Tables*, 3rd edition, 1989, University of Aalborg Press, Aalborg, Denmark.
16. Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, 1988, Morgan Kaufmann, San Mateo, CA.
17. Smith, J. Q., "Influence diagrams for statistical modeling," *Annals of Statistics*, 17(2), 1989, 654-672.
18. Whittaker, J., *Graphical Models in Applied Multivariate Statistics*, 1990, John Wiley & Sons, Chichester, UK.
19. Pearl, J. and A. Paz, "Graphoids: A graph based logic for reasoning about relevance relations," in B. D. Boulay et al. (eds.), *Advances in Artificial Intelligence*, 2, 1987, 357-363, North-Holland, Amsterdam.
20. Verma, T. and J. Pearl, "Causal networks: Semantics and expressiveness," in R. D. Shachter, T. S. Levitt, L. N. Kanal and J. F. Lemmer (eds.), *Uncertainty in Artificial Intelligence*, 4, 1990, 69-76, North-Holland, Amsterdam.
21. Fagin, R., "Multivalued dependencies and a new form for relational databases," *ACM Transactions on Database Systems*, 2(3), 1977, 262-278.
22. Spohn, W., "Ordinal conditional functions: A dynamic theory of epistemic states," in W. L. Harper and B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, 2, 1988, 105-134, D. Reidel, Dordrecht, Holland.
23. Hunter, D., "Graphoids and natural conditional functions," *International Journal of Approximate Reasoning*, 5(6), 1991, 489-504.
24. Shafer, G., P. P. Shenoy, and K. Mellouli, "Propagating belief functions in qualitative Markov trees," *International Journal of Approximate Reasoning*, 1(4), 1987, 349-400.
25. Shenoy, P. P., "Conditional independence in valuation-based systems," *International Journal of Approximate Reasoning*, 10(3), 1994, in press.
26. Wright, S., "The method of path coefficients," *Annals of Mathematical Statistics*, 5, 1934, 161-215.
27. Good, I. J., "A causal calculus," *Philosophy of Science*, 11, 1961, 305-318.
28. Darroch, J. N., S. L. Lauritzen, and T. P. Speed, "Markov fields and log-linear interaction models for contingency tables," *The Annals of Statistics*, 8(3), 1980, 522-539.
29. Lauritzen, S. L., "Mixed graphical association models," *Scandinavian Journal of Statistics*, 16(4), 1989, 273-306.

30. Wermuth, N. and S. L. Lauritzen, "Graphical and recursive models for contingency tables," *Biometrika*, 70(3), 1983, 537-552.
31. Kiiveri, H., T. P. Speed, and J. B. Carlin, "Recursive causal models," *Journal of the Australian Mathematics Society, Series A*, 36, 1984, 30-52.
32. Pearl, J., D. Geiger, and T. Verma, "The logic of influence diagrams," in R. M. Oliver and J. Q. Smith (eds.), *Influence Diagrams, Belief Nets and Decision Analysis*, 1990, 67-87, John Wiley & Sons, Chichester, England.
33. Lauritzen, S. L. and N. Wermuth, "Graphical models for associations between variables, some of which are qualitative and some quantitative," *The Annals of Statistics*, 17(1), 1989, 31-57.
34. Frydenberg, M., "The chain graph Markov property," *Scandinavian Journal of Statistics*, 17, 1990, 333-353.
35. Wermuth, N. and S. L. Lauritzen, "On substantive research hypothesis, conditional independence graphs and graphical chain models (with discussion)," *Journal of the Royal Statistical Society, Series B*, 52(1), 1990, 21-72.
36. Geiger, D. and J. Pearl, "Logical and algorithmic properties of conditional independence and graphical models," CIS Report No. 9108, 1991, Technion - Israel Institute of Technology, Computer Science Department, Haifa, Israel.
37. Paz, A., "A full characterization of pseudographoids in terms of families of undirected graphs," Technical Report No. R-95, 1987, UCLA Cognitive Systems Laboratory, Los Angeles, CA.
38. Verma, T. S., "Some mathematical properties of dependency models," Technical Report No. R-103, 1987, UCLA Cognitive Systems Laboratory, Los Angeles, CA.
39. Pearl, J., "Fusion, propagation and structuring in belief networks," *Artificial Intelligence*, 29, 1986, 241-288.
40. Geiger, D., "Towards the formalization of informational dependencies," Technical Report No. R-102, 1987, UCLA Cognitive Systems Laboratory, Los Angeles, CA.
41. Shafer, G., *Probabilistic Expert Systems*, 1994, Society for Industrial and Applied Mathematics, Philadelphia, PA, to appear.
42. Shafer, G., "Probabilistic causation," Unpublished manuscript, 1993, Graduate School of Management, Rutgers University, Newark, NJ.
43. Ur, S. and A. Paz, "The representation power of probabilistic knowledge by undirected graphs and directed acyclic graphs: A comparison," in *Proceedings of the Workshop on Uncertainty Processing in Expert Systems*, 1991, Czechoslovak Academy of Sciences, Prague, Czech Republic.
44. Jirousek, R., "Decision trees and their power to represent probability distributions," in *Proceedings of the Workshop on Uncertainty Processing in Expert Systems (WUPES)*, 1991, Czechoslovak Academy of Sciences, Prague, Czech Republic.
45. Heckerman, D., "Probabilistic similarity networks," *Networks*, 20(5), 1990, 607-636.
46. Cano, J., M. Delgado and S. Moral, "An axiomatic framework for propagating uncertainty in directed acyclic networks," *International Journal of Approximate Reasoning*, 8(4), 1993, 253-280.
47. Petrich, M., *Introduction to Semigroups*, 1973, Charles E. Merrill Publishing Co., Columbus, OH.
48. Lauritzen, S. L., A. P. Dawid, B. N. Larsen, and H.-G. Leimer, "Independence properties of directed Markov fields," *Networks*, 20(5), 1990, 491-505.