

Judea Pearl; Dan Geiger; Thomas S. Verma  
Conditional independence and its representations

*Kybernetika*, Vol. 25 (1989), No. Suppl, 33--44

Persistent URL: <http://dml.cz/dmlcz/125413>

## Terms of use:

© Institute of Information Theory and Automation AS CR, 1989

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library*  
<http://project.dml.cz>

# CONDITIONAL INDEPENDENCE AND ITS REPRESENTATIONS\*

JUDEA PEARL, DAN GEIGER, THOMAS VERMA

This paper summarizes recent investigations into the nature of informational dependencies and their representations. Axiomatic and graphical representations are presented which are both sound and complete for specialized types of independence statements.

## 1. INTRODUCTION

A central requirement for managing reasoning systems is to articulate the conditions under which one item of information is considered relevant to another, given what we already know, and to encode knowledge in structures that display these conditions vividly as the knowledge undergoes changes. Different formalisms give rise to different definitions of relevance. However, the essence of relevance can be captured by a structure common to all formalisms, which can be represented axiomatically or graphically.

A powerful formalism for informational relevance is provided by probability theory, where the notion of relevance is identified with *dependence* or, more specifically, *conditional independence*.

**Definition.** If  $X$ ,  $Y$ , and  $Z$  are three disjoint subsets of variables in a distribution  $P$ , then  $X$  and  $Y$  are said to be *conditionally independent* given  $Z$ , denoted  $I(X, Y, Z)_P$ , iff  $P(x, y | z) = P(x | z) P(y | z)$  for all possible assignments  $X = x$ ,  $Y = y$  and  $Z = z$  for which  $P(Z = z) > 0$ .  $I(X, Y, Z)_P$  is called a (*conditional independence*) *statement*. A conditional independence statement  $\sigma$  *logically follows* from a set  $\Sigma$  of such statements if  $\sigma$  holds in every distribution that obeys  $\Sigma$ . In such case we also say that  $\sigma$  is a *valid consequence* of  $\Sigma$ .

This paper addresses the problem of representing the sum total of all independence statements that logically follow from a given set of such statements. Since the size of this closure set is normally astronomical, it is important to find representations that are economical in storage and permit efficient procedures for verifying membership in the closure. Such procedures should enable us to determine, at any state of knowledge  $Z$ , what information is relevant to the task at hand and what can be ignored. Permission to ignore gives reasoning systems the power to act on the basis

\* This work was partially supported by the National Science Foundation Grant No. IRI-8610155, "Graphoids: A Computer Representation for Dependencies and Relevance in Automated Reasoning (Computer Information Science)". Sections of this paper are based on Pearl [17] and Pearl et al. [18].

of local information. We present axiomatic and graphical representations for independence statements of a specialized nature: causal statements, marginal statements and fixed-context statements. The results will be stated without proofs; these can be found in the references cited.

## 2. AXIOMATIC REPRESENTATION

It is easy to verify that the set of statements  $I(X, Z, Y)_P$  generated by any probability distribution  $P$  must obey the following set of axioms:

Symmetry

$$(1.a) \quad I(X, Z, Y) \Rightarrow I(Y, Z, X)$$

Decomposition

$$(1.b) \quad I(X, Z, Y \cup W) \Rightarrow I(X, Z, Y) \& I(X, Z, W)$$

Weak Union

$$(1.c) \quad I(X, Z, Y \cup W) \Rightarrow I(X, Z \cup W, Y)$$

Contraction

$$(1.d) \quad I(X, Z \cup Y, W) \& I(X, Z, Y) \Rightarrow I(X, Z, Y \cup W)$$

If  $P$  is strictly positive, than a fifth axiom holds:

Intersection

$$(1.e) \quad I(X, Z \cup W, Y) \& I(X, Z \cup Y, W) \Rightarrow I(X, Z, Y \cup W)$$

Axioms (1.a) through (1.d) form a system called *semi-graphoid* (Pearl and Verma [20]) and were first proposed as heuristic properties of conditional independence by Dawid [4]. Systems obeying all five axioms were called *graphoids* (Pearl and Paz [19]).

Intuitively, the essence of these axioms lies in Eqs. (1.c) and (1.d) asserting that when we learn an irrelevant fact, relevance relationships among other variables in the system should remain unaltered; any information that was relevant remains relevant and that which was irrelevant remains irrelevant. These axioms, common to almost every formalization of informational dependencies, are very similar to those assembled by Dawid [4] for probabilistic conditional independence and those proposed by Smith [24] for *Generalized Conditional Independence*. The difference is only that Dawid and Smith lumped Eqs (1.b) through (1.d) into one, and added an axiom to handle some cases of overlapping sets  $X, Y, Z$ . We shall henceforth call axioms (1.a) through (1.d) *Dawid's axioms*, or semi-graphoid axioms, interchangeably.

The general problem of verifying whether a given conditional independence statement logically follows from an arbitrary set of such statements, may be undecidable. Its decidability would be resolved upon finding a complete set of axioms for conditional independence, i.e., axioms that are powerful enough to derive all valid consequences of an arbitrary input set. Until very recently, all indications were that

axioms (1.a) through (1.d) are complete, as conjectured in [19], but a new result of Studený [25] has refuted that hypothesis. Currently, it appears rather unlikely that there exists a finite set of axioms which is complete for conditional independence, thus, the general decidability problem remains unsettled. The completeness problem is treated in [10] and completeness results for specialized subsets of probabilistic dependencies are summarized in Section 7.

### 3. GRAPHICAL REPRESENTATIONS

Interestingly, both undirected graphs (UGs) and directed acyclic graphs (DAGs) conform to the graphoid axioms (hence the name) if we associate the statement  $I(X, Z, Y)_G$  with the graphical condition “every path from  $X$  to  $Y$  is blocked by the set of nodes corresponding to  $Z$ ”. In UGs, blocking corresponds to ordinary interception. In DAGs, it is defined by a criterion called  $d$ -separation.

**Definition ( $d$ -separation).** If  $X$ ,  $Y$ , and  $Z$  are three disjoint subsets of nodes in a DAG  $D$ , then  $Z$  is said to  $d$ -separate  $X$  from  $Y$ , denoted  $I(X, Z, Y)_D$ , iff there is no path\* from a node in  $X$  to a node in  $Y$  along which the following two conditions hold: (1) every node with converging arrows either is or has a descendant in  $Z$ , and (2) every other node is outside  $Z$ . A path satisfying the conditions above is said to be *active*, otherwise it is said to be *blocked* (by  $Z$ ). Whenever a statement  $I(X, Z, Y)_D$  holds in a DAG  $D$ , the predicate  $I(X, Z, Y)$  is said to be *graphically-verified* (or an *independency*), otherwise it is *graphically-unverified* by  $D$  (or a *dependency*).

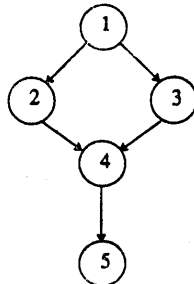


Fig. 1.

In Figure 1, for example,  $X = \{2\}$  and  $Y = \{3\}$  are  $d$ -separated by  $Z = \{1\}$ ; the path  $2 \leftarrow 1 \rightarrow 3$  is blocked by  $1 \in Z$  while the path  $2 \rightarrow 4 \leftarrow 3$  is blocked because 4 and all its descendants are outside  $Z$ . Thus  $I(2, 1, 3)$  is graphically-verified by  $D$ . However,  $X$  and  $Y$  are not  $d$ -separated by  $Z' = \{1, 5\}$  because the path  $2 \rightarrow 4 \leftarrow 3$  is rendered active. Consequently,  $I(2, \{1, 5\}, 3)$  is graphically-unverified by  $D$ ;

\* By *path* we mean a sequence of edges in the underlying undirected graph, i.e., ignoring the directionality of the arcs.

by virtue of 5, a descendent of 4, being in  $Z$ . Learning the value of the consequence 5, renders its causes 2 and 3 dependent, like opening a pathway along the converging arrows at 4.

Ideally, to employ a graph  $G$  as a representation for dependencies of some distribution  $P$  we would like to require that for every three disjoint sets of variables in  $P$  (and nodes in  $G$ ) the following equivalence should hold

$$(2) \quad I(X, Z, Y)_G \text{ iff } I(X, Z, Y)_P$$

This would provide a clear graphical representation of all variables that are conditionally independent. When Eq. (2) holds,  $G$  is said to be a *perfect map* of  $P$ . Unfortunately, this requirement is often too strong because there are many distributions that have no perfect map in UGs or in DAGs. The spectrum of probabilistic dependencies is in fact so rich that it cannot be cast into any representation scheme that uses polynomial amount of storage (cf. [27]). Geiger [7] provides a graphical representation based on a collection of graphs (Multi-DAGs) that is powerful enough to perfectly represent an arbitrary distribution, however, as shown by Verma, it requires, on the average, an exponential number of DAGs.

Being unable to provide perfect maps at a reasonable cost, we compromise the requirement that the graph represents each and every dependency of  $P$ , and allow some independencies to escape representation. We will require though that the set of undisplayed independencies be minimal. A treatment of minimal representations using undirected graphs is given in [17] and [19], here we focus on DAG representations.

**Definition.** A DAG  $D$  is said to be an *I-map* of  $P$  if for every three disjoint subsets  $X$ ,  $Y$  and  $Z$  of variables the following holds:

$$(3) \quad I(X, Z, Y)_D \Rightarrow I(X, Z, Y)_P$$

$D$  is said to be a *minimal I-map* of  $P$  if no edge can be deleted from  $D$  without destroying its *I-mapness*.

The task of finding a DAG which is a minimal *I-map* of a given distribution  $P$  can be accomplished by the traditional chain-rule decomposition of probability distributions. The procedure consists of the following steps: assign a total ordering  $d$  to the variables of  $P$ . For each variable  $i$  of  $P$ , identify a minimal set of predecessors  $S_i$  that renders  $i$  independent of all its other predecessors (in the ordering of the first step). Assign a direct link from every variable in  $S_i$  to  $i$ . The analysis of [26] ensures that the resulting DAG is a minimal *I-map* of  $P$ . The information required for this construction consists of  $n$  conditional independence statements, one for each variable, all of the form  $I(i, S_i, U_{(i)} - S_i)$  where  $U_{(i)}$  is the set of predecessors of  $i$  and  $S_i$  is a subset of  $U_{(i)}$  that renders  $i$  conditionally independent of all its other predecessors. This set of conditional independence statements is called a *causal* input list  $L$  and is said to *define* the DAG  $D$ . The term “causal” input list stems from the following

analogy: Suppose we order the variables chronologically, such that a cause always precedes its effect. Then, from all potential causes of an effect  $i$ , a causal input list selects a minimal subset that is sufficient to explain  $i$ , thus rendering all other preceding events superfluous. This selected subset of variables are considered *direct causes* of  $i$  and therefore each is connected to it by a direct link.

#### 4. THE MAIN RESULTS

Clearly, the constructed DAG represents more independencies than those listed in the input  $L$ , namely, all those that are graphically verified by the  $d$ -separation criterion. The results reported in the preceding subsection guarantee that all graphically-verified statements are indeed valid in  $P$ , i.e., the DAG is an  $I$ -map of  $P$ . It turns out that the constructed DAG has another useful property; it graphically-verifies *every* conditional independence statement that logically follows from  $L$  (i.e. holds in every distribution that obeys  $L$ ). Hence, we cannot hope to improve the  $d$ -separation criterion to detect more independencies, because all valid consequences of  $L$  (which defines  $D$ ) are already captured by  $d$ -separation.

The three theorems below formalize these results. Proofs can be found in the references cited.

**Theorem 1 (Soundness)** [26]. Let  $D$  be a DAG defined by a causal input list  $L$  of any dependency model obeying axioms (1.a) through (1.d) (e.g., probabilistic dependence). Then, every graphically-verified statement is a valid consequence of  $L$ .

**Theorem 2 (Closure)** [26]. Let  $D$  be a DAG defined by a causal input list  $L$ . Then, the set of graphically-verified statements is exactly the closure of  $L$  under axioms (1.a) through (1.d).

**Theorem 3 (Completeness)** [9]. Let  $D$  be a DAG defined by a causal input list  $L$ . Then, every valid consequence of  $L$  is graphically-verified by  $D$  (equivalently, every graphically-unverified statement in  $D$  is not a valid consequence of  $L$ ).

Theorem 1 guarantees that the DAG displays only valid statements. Theorem 2 guarantees that the DAG displays all statements that are derivable from  $L$  via axioms (1.a) through (1.d). Theorem 3 assures that the DAG displays all statements that logically follow from  $L$ , i.e., the semi-graphoid axioms are complete, capable of deriving all valid consequences of a causal input list. Moreover, since a statement in a DAG can be verified in polynomial time, Theorems 1 through 3 provide a complete polynomial inference mechanism for deriving all independency statements that are implied by a causal input list.

The first two theorems are more general than the third in the sense that they hold for every dependence relationship that obeys axioms (1.a) through (1.d), not necessarily those based on probabilistic conditional independence (proofs can be found

in [26] and [29]). Among these dependence relationships are partial correlations ([3], [19]) and qualitative dependencies ([5], [23]) which can readily be shown to be semi-graphoids. The completeness of  $d$ -separation (Theorem 3) relative partial correlations has been established in [10] while completeness relative to qualitative dependencies has not been examined yet.

## 5. DETERMINISTIC NODES AND $D$ -SEPARATION

Theorems 1 through 3 assume that  $L$  contains only statements of the form  $I(i, S_i, U_{(i)} - S_i)$ . Occasionally, however, we are in possession of stronger forms of independence relationships, in which case additional statements should be read off the DAG. A common example is the case of a *deterministic variable*, cf. [22], i.e., a variable that is functionally dependent on its corresponding parents in the DAG. The existence of each such variable  $i$  could be encoded in  $L$  by a statement of *global independence*  $I(i, S_i, U - S_i - i)$  asserting that conditioned on  $S_i$ ,  $i$  is independent of *all* other variables, not merely its predecessors. The independencies that are implied by the modified input list can be read from the DAG using an enhanced version of  $d$ -separation, named  $D$ -separation.

**Definition.** If  $X$ ,  $Y$ , and  $Z$  are three disjoint subsets of nodes in a DAG  $D$ , then  $Z$  is said to  $D$ -separate  $X$  from  $Y$ , iff there is no path from a node in  $X$  to a node in  $Y$  along which the following three conditions hold: (1) every node with converging arrows either is or has a descendant in  $Z$ ; (2) every other node is outside  $Z$ , and (3) no node is functionally determined by  $Z$ .

The new criterion certifies all independencies that are revealed by  $d$ -separation plus additional ones due to the enhancement of the input list. For example, if the arc  $2 \rightarrow 4$  in Figure 1 were deterministic, then 2  $D$ -separates 5 from 3 and, indeed, 5 and 3 must be conditionally independent given 2 because 2 determines the value of 4. The soundness and completeness of  $D$ -separation is stated in Theorem 4. In fact, the link  $3 \rightarrow 4$  is redundant because  $I(4, 2, 135)$  implies  $I(4, 23, 1)$ . It is for this reason that deterministic nodes can be presumed to have deterministic incoming arrows.

**Theorem 4** [11]. Let  $D$  be a DAG defined by causal input list  $L$ , possibly containing functional dependencies. Then a statement is a valid consequence of  $L$  if and only if it is graphically-verified in  $D$  by the  $D$ -separation criterion.

These graphical criteria provide easy means of recognizing conditional independence in Bayesian networks (see Section 8) as well as identifying the set of parameters needed for any given computation. We now show how these theorems can be employed as an inference mechanism. Assume an expert has identified the following conditional independencies between variables denoted 1 through 5:

$$L = \{I(2, 1, \emptyset), I(3, 1, 2), I(4, 23, 1), I(5, 4, 123)\}$$

(the first statement in  $L$  is trivial). We address two questions. First, what is the set of all valid consequences of  $L$ ? Second, in particular, is  $I(3, 124, 5)$  a valid consequence of  $L$ ? For general input lists the answer for such questions may be undecidable but, since  $L$  is a causal list, it defines a DAG that graphically verifies each and every valid consequences of  $L$ . The DAG  $D$  is the one shown in Figure 1, which constitutes a compact representation of all valid consequences of  $L$ . To answer the second question, we simply observe that  $I(3, 124, 5)$  is graphically-verified in  $D$ . A graph-based algorithm for another subclass of statements, called *fixed context* statements, is given in Section 7 (see Theorem 8).

## 6. STRONG COMPLETENESS

Theorem 3 can be restated to assert that for every DAG  $D$  and any statement  $\sigma$  graphically unverified by  $D$  there exists a probability distribution  $P_\sigma$  that embodies  $D$ 's causal input set  $L$  and the dependency  $\sigma$ . By Theorem 2,  $P_\sigma$  must embody all graphically-verified statements as well because they are all derivable from  $L$  by Dawid's axioms. Thus, Theorems 2 and 3 guarantee the existence of a distribution  $P_\sigma$  that satisfies all graphically verified statements and a single, arbitrarily-chosen, graphically unverified statement (i.e., a dependency). The question answered by the next theorem is the existence of a distribution  $P$  that embodies all independencies of  $D$  and *all* its dependencies, not merely a single dependency. A set of axioms that guarantees the existence of such a distribution is said to be strongly complete [2].

**Theorem 5 (Strong Completeness)** [9]. For every DAG  $D$  there exists a distribution  $P$  such that for every three disjoint sets of variables  $X$ ,  $Y$  and  $Z$  the following holds;

$$I(X, Z, Y)_D \text{ iff } I(X, Z, Y)_P$$

Theorem 5 legitimizes the use of DAGs as a representation scheme for probabilistic dependencies; a model builder who uses the language of DAGs to express dependencies is guarded from inconsistencies.

## 7. OTHER COMPLETENESS RESULTS

This section presents a summary of completeness results for two specialized subsets of independence statements. Proofs can be found in [8] and [10]. The first result establishes an axiomatic characterization of *marginal statements*, i.e., statements of the form  $I(X, Z_0, Y)$  where the middle argument  $Z_0$  is fixed. The second result provides a complete axiomatic characterization of fixed-context statements, i.e., statements of the form  $I(X, Z, Y)$  where  $X \cup Z \cup Y$  sum to a fixed set of variables  $U$ .

**Theorem 6 (Completeness for Marginal Independence).** Let  $\Sigma$  be a set of marginal



statements closed under the following axioms:

$$\begin{array}{ll}
 \textit{Symmetry} & I(X, Z_0, Y) \rightarrow I(Y, Z_0, X) \\
 \textit{Decomposition} & I(X, Z_0, YW) \rightarrow I(X, Z_0, Y) \\
 \textit{Mixing} & I(X, Z_0, Y) \& I(XY, Z_0, W) \rightarrow I(X, Z_0, YW).
 \end{array}$$

There exists a probability model  $P$  that obeys all statements in  $\Sigma$  and none other.

**Theorem 7 (Completeness for Fixed-Context).** Let  $\Sigma$  be a set of fixed-context statements closed under the axioms:

$$\begin{array}{ll}
 \textit{symmetry} & I(X, Z, Y) \rightarrow I(Y, Z, X) \\
 \textit{weak union} & I(X, Z, YW) \rightarrow I(X, ZY, W) \\
 \textit{weak contraction} & I(XY, Z, W) \& I(X, ZW, Y) \rightarrow I(X, Z, YW).
 \end{array}$$

There exists a probability model  $P$  that obeys all statements in  $\Sigma$  and none other. Moreover, if  $\Sigma$  is closed also under

$$\textit{intersection} \quad I(X, ZW, Y) \& I(X, ZY, W) \rightarrow I(X, Z, YW),$$

then  $P$  can be selected to be strictly positive.

The membership question, whether a given marginal statement follows from an arbitrary set of such statements, can be answered in linear time (cf. [8]). For fixed-context statements, the membership question can be answered in quadratic time (cf. [2] and [28]). Inferences involving fixed-context statements relative to strictly positive distributions are governed, in fact, by much stronger results, equivalent to Theorems 1 through 3, based on undirected-graph representation.

**Definition.** An undirected graph  $G$  is said to be the *skeleton* of a list  $L$  of statements if a pair  $(\alpha, \beta)$  of vertices is adjacent in  $G$  iff there is no statement  $I(X, Z, Y)$  in  $L$  such that  $\alpha \in X$  and  $\beta \in Y$ .

**Theorem 8.** Let  $L$  be a list of fixed-context independence statements relative to a positive distribution  $P$ , and let  $G$  be the skeleton of  $L$ . Then, the following hold:

1. *Soundness.* Every graphically-verified statement in  $G$  is a valid consequence of  $L$ .
2. *Closure.* The set of graphically-verified statements is exactly the closure of  $L$  under axioms (1.a) through (1.e).
3. *Completeness.* Every valid consequence of  $L$  is graphically-verified by  $G$  (equivalently, every graphically-unverified statement in  $G$  is not a valid consequence of  $L$ ).

Theorem 8 strengthens the soundness property known to hold for Markov fields (cf. [13], [14]).

## 8. COROLLARIES

Theorem 1 leads to four corollaries which are the key to the construction of minimal DAG representations for a given distribution  $P$ . Such representations were called *Bayesian networks* in [17], and *influence diagrams* in [18].

**Definition.** Given a probability distribution  $P$  on a set  $U$  of variables,  $U = \{X_1, X_2, \dots, X_n\}$ , a DAG  $D = (U, E)$  is called a *Bayesian network* of  $P$  iff  $D$  is a minimal  $I$ -map of  $P$ .

**Corollary 1.** Given a probability distribution  $P(x_1, x_2, \dots, x_n)$  and any ordering  $d$  of the variables. The DAG created by designating as parents of  $X_i$  any minimal set  $S_i$  of predecessors satisfying

$$(4) \quad I(X_i, S_i, U_{(i)} - S_i), \quad U_{(i)} = \{X_1, X_2, \dots, X_{i-1}\}$$

is a Bayesian network of  $P$ . Conversely, every Bayesian networks of  $P$  can be constructed by identifying the parent sets  $S_i$  defined in (4) along some ordering  $d$ . If  $P$  is strictly positive, then all the parent sets are unique (cf. [19]) and the Bayesian network is unique as well (given  $d$ ).

Since all conditional independencies portrayed in the network (via  $d$ -separation) are valid in  $P$ , they are order independent. This yields an order-independent test for minimal  $I$ -mapness.

**Corollary 2.** Given a DAG  $D$  and a probability distribution  $P$ , a necessary and sufficient condition for  $D$  to be a minimal  $I$ -map (hence a Bayesian network) of  $P$  is that each variable  $X_i$  be conditionally independent of all its non-descendants, given its parents  $S_i$ , and no proper subset of  $S_i$  satisfies this condition.

The necessary part follows from the fact that every parent-set  $S_i$   $d$ -separates  $X_i$  from all its non-descendants. The sufficient part holds because  $X_i$ 's independence of all its nondescendants entails  $X_i$ 's independence of its predecessors in a particular ordering  $d$  (as required by Corollary 1).

**Corollary 3.** If a Bayesian network  $D$  is constructed from  $P$  (by the method of Corollary 1) in some ordering  $d$ , then any ordering  $d'$  consistent with the direction of arrows in  $D$  would give rise to an identical network.

The validity of Corollary 3 follows from that of Corollary 2, which ensures that the set  $S_i$  will satisfy Eq. (4) in any new ordering, as long as the new set of  $X_i$ 's predecessors does not contain any of  $X_i$ 's old descendants. Thus, once the network is constructed, the original order can be forgotten; only the partial order displayed in the network matters.

Another interesting corollary of Theorem 1 is a generalization of the celebrated *Markov-chain* property which is used extensively in the probabilistic analysis of

random walks, time-series data and other stochastic processes (cf. [6], [15]). The property states that if in a sequence of  $n$  trials  $X_1, X_2, \dots, X_n$  the outcome of any trial  $X_k$ ,  $k = 2, 3, \dots, n$  depends only on the outcome of its directly preceding trial  $X_{k-1}$  then, given the entire past and future trials  $X_1, X_2, \dots, X_{k-1}, X_{k+1}, \dots, X_n$ , the outcome of  $X_k$  depends only on its two nearest neighbors  $X_{k-1}$  and  $X_{k+1}$ . Formally:

$$\text{If } I(X_k, X_{k-1}, X_{k-2} \dots X_1), 2 \leq k \leq n, \text{ then}$$

$$I(X_k, X_{k-1} X_{k+1}, X_n \dots X_{k+2} X_{k-2} \dots X_1), 2 \leq k \leq n - 1.$$

(The converse holds only for full graphoids, e.g., strictly positive distributions.) Theorem 1 generalizes the Markov-chain property to dependencies other than probabilistic and to structures other than chains. The  $d$ -separation criterion uniquely determines a *Markov blanket* for any given node  $X_i$  in a Bayesian network, namely, a set  $BL(X_i)$  of variables that renders  $X_i$  independent of all variables not in  $BL(X_i)$ , i.e.,  $I(X_i, BL(X_i), U - BL(X_i) - X_i)$ .

**Corollary 4.** In any Bayesian network, the union of the following three types of neighbors is sufficient for forming a Markov blanket of a node  $X_i$ : the direct parents of  $X_i$ , the direct successors of  $X_i$  and all direct parents of the latter.

Thus, if the network consists of a single path (i.e., a Markov chain), the Markov blanket of any non-terminal node consists of its two immediate neighbors, as expected. In trees, the Markov blanket consists of the (unique) parent and the immediate successors. In Figure 1, however, the Markov blanket of node 2 is  $\{1, 4, 3\}$ . Note that in general, these Markov blankets are not minimal; alternative orderings might display  $X_i$  with a smaller set of neighbors.

The necessary part of Corollary 2 was stated without proof in [12] and was later used in the derivations of [16] and [21]. Corollaries 2 and 3 are proven in Smith [24] using the axioms of Eq. (1). Since Theorem 1 establishes  $d$ -separation a sound procedure relative to Dawid's axioms, the validity of such corollaries can now be verified by purely graphical means. The use of  $d$ -separation in proving the validity of graphical transformations on influence diagrams is discussed in [17].

## ACKNOWLEDGEMENT

We thank N. Dalkey, R. Fagin, S. Lauritzen and A. Paz for many valuable discussions.

## REFERENCES

- 
- [1] C. Beeri: On the membership problem for functional and multivalued dependencies in relational database. *ACM Trans. Database Systems* 5 (1980), 3, 241–249.
  - [2] C. Beeri, R. Fagin and R. A. Howard: A complete axiomatization of functional dependencies and multi-valued dependencies in database relations. In: *Proceedings, 1977 ACM SIGMOD Internat. Conference on Management of Data, Toronto, Canada 1977*, pp. 47–61.

- [3] H. Cramér: *Mathematical Methods of Statistics*. Princeton Univ. Press, Princeton, N. J. 1946.
- [4] A. P. Dawid: Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B* 41 (1979), 1, 1—31.
- [5] R. Fagin: Multivalued dependencies and a new form for relational databases. *ACM Trans. Database Systems* 2 (1977), 3, 262—278.
- [6] W. Feller: *An Introduction to Probability Theory and Its Applications* (Chapter XV). Third edition. John Wiley and Sons, New York 1968.
- [7] D. Geiger: Towards the Formalization of Informational Dependencies. Technical Report R-102, UCLA Cognitive Systems Laboratory, December 1987.
- [8] D. Geiger, A. Paz and J. Pearl: Axioms and Algorithms for Inferences Involving Probabilistic Independence. Technical Report R-119, January 1989.
- [9] D. Geiger and J. Pearl: On the logic of causal models. In: *Proc. of the 4th Workshop on Uncertainty in AI*, St. Paul, Minn., August 1988, pp. 136—147.
- [10] D. Geiger and J. Pearl: Logical and Algorithmic Properties of Conditional Independence. Technical Report 870066 (R-97), UCLA Cognitive Systems Laboratory, February 1988. In: *Proc. of Workshop on Statistics and AI*, Ft. Lauderdale, Fl., January 1989, 19-1—19-10.
- [11] D. Geiger, T. Verma and J. Pearl: Identifying Independence in Bayesian Networks. Technical Report R-116, UCLA Cognitive Systems Laboratory.
- [12] R. A. Howard and J. E. Matheson: *Influence Diagrams*. In: *Principles and Applications of Decision Analysis*, Strategic Decisions Group, Menlo Park, CA 1981.
- [13] V. Isham: An introduction to spatial point processes and Markov random fields. *Internationat. Statist. Rev.* 49 (1981), 21—43.
- [14] S. L. Lauritzen: *Lectures on Contingency Tables*. Second edition. University of Aalborg Press, Aalborg, Denmark 1982.
- [15] J. S. Meditch: *Stochastic Optimal Linear Estimation and Control*. McGraw-Hill, New York 1969.
- [16] S. M. Olmsted: On Representing and Solving Decision Problems. Ph. D. Thesis, EES Dept., Stanford University, Stanford, CA 1983.
- [17] J. Pearl: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA 1988.
- [18] J. Pearl, D. Geiger and T. Verma: The logic of influence diagrams. In: *Proceedings of the Conference on Influence Diagrams*, Berkeley, CA, May 1988.
- [19] J. Pearl and A. Paz: GRAPHOIDS: a Graph-based Logic for Reasoning about Relevance Relations. Technical Report 850038 (R-52), UCLA Computer Science Department, October 1985; also *Proceedings, ECAI-86*, Brighton, U. K. June 1986.
- [20] J. Pearl and T. Verma: The logic of representing dependencies by directed acyclic graphs. In: *Proceedings AAA-I*, Seattle, Washington, July 1987, pp. 374—379.
- [21] R. D. Shachter: Evaluating influence diagrams. *Oper. Res.* 34 (1986), 6, 871—882.
- [22] R. D. Shachter: Probabilistic inference and influence diagrams. *Oper. Res.* 36 (1988), 4, 589—604.
- [23] G. Shafer, P. P. Shenoy and K. Mellouli: Propagating belief functions in qualitative Markov trees. *Internat. J. of Approximate Reasoning* 1 (1988), 4, 349—400.
- [24] J. Q. Smith: *Influence Diagrams for Statistical Modeling*. Technical Report No. 117, Department of Statistics, University of Warwick, Coventry, England, June 1987.
- [25] M. Studený: Attempts at axiomatic description of conditional independence. In: *Proceedings of the Workshop on Uncertainty Processing in Expert Systems*. *Kybernetika* 25 (1989), Supplement to No. 1—3, pp. 65—72.
- [26] T. S. Verma: *Causal Networks: Semantics and Expressiveness*. Technical Report R-65,

- UCLA Cognitive Systems Laboratory, 1986; also in Proceedings of the 4th Workshop on Uncertainty in AI, St. Paul, Minn., August 1988, pp. 352—359.
- [27] T. S. Verma: Some Mathematical Properties of Dependency Models. Technical Report R-103, UCLA Cognitive Systems Laboratory, 1987.
- [28] T. S. Verma: On the Membership Problem in Semi-Graphoids. Technical Report R-115, UCLA Cognitive Systems Laboratory, 1988.
- [29] T. Verma and J. Pearl: Causal networks: semantics and expressiveness. In: Proceedings of the 4th Workshop on Uncertainty in AI, St. Paul, Min. August 1988, pp. 352—359.

*Dr. Judea Pearl, Dr. Dan Geiger, Dr. Thomas Verma, Cognitive Systems Laboratory, Computer Science Department, University of California Los Angeles, CA 90024, U.S.A.*