Probabilistic Relevance Relations

Dan Geiger and David Heckerman

***Abstract*— The intuition behind the construction of Bayesian networks and other graph-based representations of joint probability distributions from expert judgments is based on the assumed relationship between "connectedness" in the graphical model and "relatedness" among the variables involved. We show that several plausible definitions of relatedness do not adhere to such an equivalence. We then provide a definition of probabilistic relatedness that is closely related to connectedness in the graphical model and prove that the two concepts are equivalent whenever the model uses only propositional variables and assuming every combination of value assignment to these variables is feasible. We conjecture that the equivalence established holds also when these restrictions are lifted.**

***Index Terms*—Bayesian networks, conditional independence.**

I. Introduction

BAYESIAN networks are graph-based representations of joint probability distributions which have found a variety of applications for diagnosis, prediction, image recovery, and in many other domains [9]. There are three options for constructing Bayesian networks. The first is to build a network manually with the help of a domain expert. This approach is used quite often and is most useful for moderate-size models. A second approach is to construct a Bayesian network completely from data. This approach is most useful when a database of cases is available and when experts are too costly or unavailable. Finally, a hybrid method by which a rough model is build from expert's judgments and then tuned by data is perhaps the most promising approach. The analysis of this paper concentrates on issues arising from the construction of Bayesian networks from expert's judgments.

To be concrete, let us first consider a simple Bayesian network D_0 of the form $a \rightarrow c \leftarrow b$. This network represents a joint probability distribution $p(a, b, c)$ of three random variables a, b, and c, such that $p(a, b, c) = p(a)p(b)p(c|a, b)$. It is a minimal Bayesian network of p if none of it edges can be removed, that is, neither $p(a, b, c) = p(a)p(b)p(c|a)$ nor $p(a, b, c) = p(a)p(b)p(c|b)$ hold for all assignments for a, b, and c.

The intuition behind the construction of Bayesian networks from expert judgments is based on the assumed relationship between "connectedness" in the graphical model and "relatedness" between the variables involved. That is, for example, a and b are connected in D_0 because they are related to each other through c. However, one can easily construct examples of a distribution p such that D_0 is a minimal network of p yet a and b are marginally independent, and also conditionally independent given any specific value of c. Such an example seemingly contradicts the analogy between "connectedness" in the graphical model and "relatedness" in the joint distribution because a and b are seemingly unrelated in any context—when c is unknown and when c is known—yet they are connected in the graphical representation.

In this paper, we seek a definition of relatedness that fits the intuition that connected nodes in the graphical representation correspond to variables that are related probabilistically. We shall prove that our concept of relatedness is indeed equivalent to connectedness in a minimal Bayesian network under the assumption that all variables are propositional and that every combination of value assignment is feasible. We conjecture that the analogy established holds even when these restrictions are lifted and hope that this paper will stimulate the resolution of this conjecture

Apart from the epistemological reassurance given by our definition, our results also justify prevailing decomposition techniques that simplify the process of acquiring probabilistic knowledge from domain experts via models known as similarity networks [7]. A similarity network is a set of Bayesian networks, called the local networks, each constructed under a different set of hypotheses H_i. In each local network D_i, only those variables that "help to distinguish" between the hypotheses in H_i are depicted. The success of this model stems from the fact that only a small portion of variables helps to distinguish between the carefully chosen set of hypotheses H_i. Thus, the model usually includes several small networks instead of a single large Bayesian network. A plausible formal definition of what is meant by "help to distinguish" is provided herein, where x helps to distinguish values of h means that a and h are related probabilistically.

This paper is organized as follows. In Section II, we provide definitions of a Bayesian network and conditional independence and review some of their properties. In Section III, we associate connectedness with conditional independence. In Section IV, we develop a definition of probabilistic relatedness, and in Section V we prove the equivalence between relatedness and connectedness under some restrictions.

II. Background

Throughout the discussion we consider a finite set of variables $U = \{u_1, \cdots, u_n\}$ each with a finite domain $d(u_i)$ and a probability distribution $p(U)$ having the Cartesian

Manuscript received February 19, 1995; revised August 11, 1997. An earlier version of this paper was presented at [4].

D. Geiger is with Microsoft Research, Redmond, WA 98052-6399 USA, on leave from the Department of Computer Science, Technion—Israel Institute of Technology, Haifa, Israel, 32000 (e-mail: dang@cs.technion.ac.il).

D. Heckerman is with Microsoft Research, Redmond, WA 98052-6399 USA (e-mail: heckerma@microsoft.com).

Publisher Item Identifier S 1083-4427(98)00124-6.

product $\times_{u_i \in U} d(u_i)$ as its sample space. We use lowercase letters possibly subscripted (e.g., a, b, x, or u_i) to denote variables, and use uppercase letters (e.g., X, Y, or Z) to denote sets of variables. A bold lowercase or uppercase letter refers to a value (instance) of a variable or of a set of variables, respectively. A value $\mathbf{X}$ of a set of variables X is an element in the Cartesian product $\times_{x \in X} d(x)$, where $d(x)$ is the set of values of x. The notation $X = \mathbf{X}$ stands for $x_1 = \mathbf{x}_1, \cdots, x_n = \mathbf{x}_n$, where $X = \{x_1, \cdots, x_n\}$ and $\mathbf{x}_i$ is a value of x_i.

We let $X \perp Y | Z = \mathbf{Z}$ denote the statement that X and Y are conditionally independent given $Z = \mathbf{Z}$, namely, that $p(X, Y, Z = \mathbf{Z})p(Z = \mathbf{Z}) = p(X, Z = \mathbf{Z})p(Y, Z = \mathbf{Z})$ for every value of X and Y. We let $X \perp Y | Z$ denote the statement that X and Y are independent given every value for Z, namely, that $X \perp Y | Z = \mathbf{Z}$ holds for every $\mathbf{Z}$. Similarly, $X \perp Y$ denotes the statement X and Y are marginally independent which can also be thought of as a shorthand notation for $X \perp Y | \emptyset$.

A Bayesian network is a representation of independence statements as well as a representation of a joint probability distribution. Below we give a definition and some consequences. For a more comprehensive overview, consult [9].

Definition [9]: A directed acyclic graph D of a joint probability distribution $p(U)$ is a *Bayesian network of* p if D is constructed from p by the following steps: assign an arbitrary *construction order* $u_1, u_2, , \cdots, u_n$ to the variables in U, and designate a node u_i for each variable u_i.[1] For each u_i in U, identify a set $C_i \subseteq \{u_1, \cdots, u_{i-1}\}$ such that

$$\{u_i\} \perp \{u_1, \cdots, u_{i-1}\} \setminus C_i | C_i \tag{1}$$

holds wrt p (with respect to p). Assign a link from every node in C_i to u_i. Each node is associated with the conditional probability distribution $p(u_i | C_i)$. The resulting network is *minimal* if, for each $u_i \in U$, no proper subset of C_i satisfies (1).

By the chaining rule it follows that

$$p(u_1, \cdots, u_n) = \prod p(u_i | u_1, \cdots, u_{i-1})$$

and by the definition of $\{u_i\} \perp \{u_1, \cdots, u_{i-1}\} \setminus C_i | C_i$ we further obtain

$$p(u_1, \cdots, u_n) = \prod p(u_i | C_i). \tag{2}$$

Thus, the joint distribution is represented by the network and can be used for computing the posterior probability of every variable given a value to some other variables.

Note that the number of parameters that a Bayesian network requires and the complexity of its topology depend on the construction order, which is not dictated by its definition. There are many possibilities to choose a construction order. In practice, cause-and-effect and time-order relationships often suggest construction orders that yield simple networks.[2]

[1] We deliberately denote with u_i the node that corresponds to variable u_i. It will be clear from the context whether we talk about a node or a variable.

[2] Bayesian networks are often called *causal networks*.

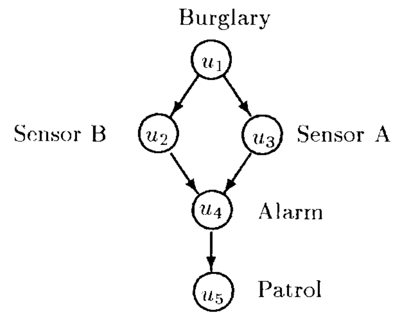


Fig. 1. An example of a Bayesian network.

For example, a Bayesian network can represent the following situation. Suppose an alarm system is installed in your house in order to detect burglaries; and suppose it can be activated by two separate sensors. Suppose also that, when the alarm sound is activated, there is a good chance that a police patrol will show up. We are interested in computing the probability of a burglary given a police car is near your house.

We consider five binary variables, *burglary* (u_1), *sensorA* (u_2), *sensorB* (u_3), *alarm* (u_4), and *patrol* (u_5), each having two values *yes* and *no*. We know that the outcome of the two sensors are conditionally independent given *burglary*, and that *alarm* is conditionally independent of *burglary* given the outcome of the sensors. We also know that *patrol* is conditionally independent of *burglary* given *alarm* (assuming that only the alarm prompts a police patrol). This qualitative information implies that the following three independence statements hold in any probability distribution that describes this story: $\{u_3\} \perp \{u_2\} | \{u_1\}$ $\{u_4\} \perp \{u_1\} | \{u_2, u_3\}$ $\{u_5\} \perp \{u_1, u_2, u_3\} | \{u_4\}$. Consequently, according to our definition, the graph shown in Fig. 1 is a Bayesian network of the burglary story.

In addition to the topology of the network, we need to specify the following conditional distributions: $p(u_1)$, $p(u_2 | u_1)$, $p(u_3 | u_1)$, $p(u_4 | u_2, u_3)$, and $p(u_5 | u_4)$. From these conditional distributions, we can compute via (2) any probability involving these variables. However, to do such computations efficiently we need to know additional independence statements which follow from the topology of the network but were not used to construct the Bayesian network (such as $\{$*patrol*$\} \perp \{$*burglary*$\} | \{$*sensorA*, *sensorB*$\}$). The criteria of d-separation, defined below, provides the most general mechanism to infer independence statements from the topology of the Bayesian network. Some terminology is first established.

A *trail* in a Bayesian network D is a path in D in which links are taken regardless of their direction. A node b is called a *head-to-head* node wrt a trail t if there are two consecutive links $a \rightarrow b$ and $b \leftarrow c$ on t. Two nodes are *connected* in a Bayesian network if there exists a trail connecting them. Otherwise they are *disconnected*. A *connected component* C of a Bayesian network D is a subgraph of D in which every two nodes are connected. A connected component is *maximal* if there exists no proper super-graph of C that is a connected component of D. If $x \rightarrow y$ is a link in a Bayesian network, then x is a *parent* of y and y is a *child* of x. If there is a directed path from x to y, then x is an *ancestor* of y and y is a *descendant* of x.

Definition [9]: A trail t is *active wrt* a set of nodes Z if 1) every head-to-head node wrt t either is in Z or has a descendant in Z and 2) every other node along t is outside Z. Otherwise, the trail is said to be *blocked* (or d*-separated*) by Z.

In Fig. 1, for example, both trails between $\{u_2\}$ and $\{u_3\}$ are d-separated by $Z = \{u_1\}$; the trail $u_2 \leftarrow u_1 \rightarrow u_3$ is d-separated by Z because node u_1, which is not a head-to-head node wrt this trail, is in Z. The trail $u_2 \rightarrow u_4 \leftarrow u_3$ is d-separated by Z, because node u_4 and its descendant u_5 are outside Z. In contrast, $u_2 \rightarrow u_4 \leftarrow u_3$ is not d-separated by $Z' = \{u_1, u_5\}$ because u_5 is in Z'.

The theorem below is the major building block for most of the developments presented in this article and is fundamental to the theory of Bayesian networks.

Theorem 1 [13]: Let D be a Bayesian network of a probability distribution $p(U)$ and let X, Y, and Z be three disjoint subsets of U. If all trails between a node in X and a node in Y are d-separated by Z, then $X \perp Y | Z$ holds wrt p.

For example, in the Bayesian network of Fig. 1, all trails between u_1 and u_5 are d-separated by $\{u_2, u_4\}$. Thus, Theorem 1 guarantees that $\{u_5\} \perp \{u_1\} | \{u_2, u_3\}$ holds wrt p. Geiger *et al.* [5] generalize Theorem 1 and show that no other graphical criteria reveals more independence statements of p than does d-separation. Lauritzen *et al.* [8] establish another graphical criteria and show that it is equivalent to d-separation.

One immediate consequence of Theorem 1 is that if two sets of nodes X and Y are disconnected in a Bayesian network of $p(u_1, \cdots, u_n)$, then $X \perp Y$ holds (wrt p) because there is no active trail between a node in X and a node in Y. Another well-known consequence is that if Z_a is the set of parents of a node a, and Y_a are the set of all nodes that are not descendants of a except a's parents. Then, $\{a\} \perp Y_a | Z_a$ holds (wrt p). The argument is simple. The set Z_a d-separates all trails between a node in Y_a and a because each such trail either passes through a parent of a and therefore is blocked by Z_a, or each such trail must reach a through one of a's children and thus must contain a head-to-head node w, where neither w nor its descendants are in Z_a.

In our proofs we will only use the following properties of conditional independence. A variant of these properties was introduced by Dawid [1] and Spohn [12] and further studied by Pearl [9] and Pearl and Paz [10].

Symmetry
$$X \perp Y | Z \Rightarrow Y \perp X | Z \tag{3}$$
Decomposition
$$X \perp (Y \cup W) | Z \Rightarrow X \perp Y | Z \tag{4}$$
Weak union
$$X \perp (Y \cup W) | Z \Rightarrow X \perp Y | (Z \cup W) \tag{5}$$
Contraction
$$X \perp Y | Z \;\&\; X \perp W | (Z \cup Y) \Rightarrow X \perp (Y \cup W | Z). \tag{6}$$

It is worth mentioning that the proof of Theorem 1 only uses these properties and therefore every trinary relation $\perp$ that satisfies these properties can be represented by a Bayesian network and the result of Theorem 1 applies. In particular, partial correlation and embedded multivalued dependencies (from relational database theory) satisfy these properties.

III. Connectedness in Terms of Independence

In this section, we show that if a and b are disconnected in one minimal network of p then a and b are disconnected in every minimal network of p. This result shows that the concept of connectivity can be phrased in terms of independence statements that hold in p. Indeed, we find that a and b are disconnected in a minimal Bayesian network of $p(U)$ if and only if there exists a partition U_a, U_b of U such that $U_a \perp U_b$, $a \in U_a$ and $b \in U_b$ (a partition of a set U is a pair of nonempty disjoint subsets of U whose union is U).

Lemma 2: Let D be a minimal Bayesian network of $p(U)$, and D_X be a connected component of D with a set of nodes X. Then, there exists no partition X_1, X_2 of X such that $X_1 \perp X_2$ holds wrt p.

Proof: Assume contrary to the lemma's claim that X_1, X_2 is a partition of X and that $X_1 \perp X_2$ holds. Since X_1 and X_2 are connected in D, there must exist a link between a node in X_1 and a node in X_2. Without loss of generality, assume it is directed from a node v in X_1 to a node u in X_2. Let Z_1, Z_2 be the parents of u in X_1 and X_2, respectively. Since $X_1 \perp X_2$ holds wrt p, by symmetry and decomposition, $\{u\} \cup Z_2 \perp Z_1$ holds too. By symmetry and weak union, $\{u\} \perp Z_1 | Z_2$ holds as well. Now since $Z_1 \cup Z_2$ are the parents of u, according to the comment that follows Theorem 1, $\{u\} \perp Y | Z_1 \cup Z_2$ holds, where Y is the set of u's nondescendants except its parents. Consequently, by the contraction property, $\{u\} \perp Z_1 \cup Y | Z_2$ holds. Since Z_2 is a proper subset of $Z_1 \cup Z_2$ (because Z_1 contains v), where $Z_1 \cup Z_2$ are the parents of u in D, D is not minimal. □

Theorem 3: If two nodes are disconnected in some minimal Bayesian network of $p(U)$, then they are disconnected in every minimal Bayesian network of $p(U)$.

Proof: It suffices to show that any two minimal Bayesian networks of p share the same maximal connected components. Let D_{A} and D_{B} be two minimal Bayesian networks of p. Let C_{A} and C_{B} be maximal connected components of D_{A} and D_{B}, respectively. Let A and B be the nodes of C_{A} and C_{B}, respectively. We show that either $A = B$ or $A \cap B = \emptyset$. This demonstration will complete the proof, because for an arbitrary maximal connected component C_{A} in D_{A} there must exist a maximal connected component in D_{B} that shares at least one node with C_{A}. Thus, by the above claim, it must have exactly the same nodes as C_{A}. Therefore, each maximal connected component of D_{A} shares the same nodes with exactly one maximal connected component of D_{B}. Hence, D_{A} and D_{B} share the same maximal connected components.

Since D_{A} is a minimal Bayesian network of p and C_{A} is a maximal connected component of D_{A}, there is no trail between A and $U \setminus A$ and so, by Theorem 1, $A \perp U \setminus A$ holds. Using symmetry and decomposition, $(A \cap B) \perp (B \setminus A)$ holds too. Thus, by Lemma 2, for C_{B} to be a maximal connected component, either $A \cap B$ or $B \setminus A$ must be empty, lest D_{B}

would not be minimal. Similarly, for C_A to be a maximal connected component, $A \cap B$ or $A \setminus B$ must be empty. Thus, either $A = B$ or $A \cap B = \emptyset$. □

Theorem 4: Two variables x and y are disconnected in every minimal network of $p(U)$ iff there exists a partition U_x, U_y of U such that $U_x \perp U_y$ and $x \in U_x$ and $y \in U_y$.

Proof: Suppose x and y are disconnected in some minimal network D of $p(U)$. Let U_x be the variables connected to x in D, and U_y be the rest of the variables in U. Thus there is no trail between U_x and U_y and so, by Theorem 1, $U_x \perp U_y$ holds (wrt p).

Suppose there exists a partition U_x, U_y of U such that $x \in U_x$, $y \in U_y$, and $U_x \perp U_y$ holds (wrt p). We show that in every minimal Bayesian network D of p, nodes x and y do not reside in the same connected component. Assume, to the contrary, that x and y reside in the same maximal connected component of some minimal Bayesian network D of p, and that C are the nodes of that component. Now, $(U_x \cap C) \perp (U_y \cap C)$ holds (wrt p) because it follows from $U_x \perp U_y$ by the symmetry and decomposition properties. Moreover, $U_x \cap C$ and $U_y \cap C$ are not empty, because they include x and y, respectively. Since U_x and U_y are disjoint, the two sets $U_x \cap C$, $U_y \cap C$ partition C. Therefore, by Lemma 2, D cannot be minimal, contrary to our assumption. □

IV. Alternative Definitions for Relatedness

In this section, we discuss several possibilities for defining probabilistic relatedness (or unrelatedness as a complementary notion), indicate the pitfalls of the proposed definitions and conclude with a definition that bypasses these pitfalls. The common ground of the proposed definitions is the idea that two variables are unrelated iff they are independent given an appropriately large set of contexts.

As a first alternative, we could define a and b to be *unrelated* [wrt $p(u_1, \cdots, u_n)$] if and only if $\{a\} \perp \{b\} | V = \mathbf{V}$ where V is a subset of variables of $U \setminus \{a, b\}$ and $\mathbf{V}$ is a specific assignment to each variable in V. In other words, a context consists of a set of assignments to a subset of variables and a and b are unrelated iff they are independent given any such context. For example, if $U = \{a, b, c\}$ and if both $a \perp b$ and $a \perp b | c$ hold, then a and b are said to be unrelated. The following well-known property of conditional independence [1], [12], which we call property B,

$$\{a\} \perp \{b\} \text{ and } \{a\} \perp \{b\} | \{c\} \Rightarrow \{a, c\} \perp \{b\} \text{ or } \{a\} \perp \{b, c\} \tag{7}$$

holds whenever c is a binary variable. The converse of (7) follows immediately from (3)–(6). Thus, due to Theorem 4, we conclude that in any minimal network of $p(a, b, c)$, if c is a binary variable, then a and b will reside on two distinct components iff they are unrelated (wrt p).

The technical problem with this definition lies in the fact that if c is not a binary variable, then property B does not hold anymore. The difficulty can be traced to the fact that if we conceive a and b to be unrelated (and therefore expect a and b to be disconnected), we indeed mean to say that a and b are marginally independent and conditionally independent given any possible context. One particular context not considered in our first attempt is the situation when c is equal to either c_i or c_j but we do not know to which value. Of course, if c is a binary variable, then saying that c gets one of its values is a tautology that adds nothing to our knowledge but if c is not a binary variable, then restricting the domain of c is a new context and so if a and b are to be considered unrelated, then they should also be independent conditioned on $c = c_i \vee c_j$.

Indeed the following theorem shows one way to extend property B to nonbinary variables. This theorem justifies a second definition of relatedness for the simple case of three variables.

Theorem 5: Let $p(a, b, c)$ be a joint probability distribution of three random variables a, b, and c. If $\{a\} \perp \{b\}$ and $\{a\} \perp \{b\} | c = c_i$ for $i = 1, \cdots, k$ and if $\{a\} \perp \{b\} | c = c_i \vee c_j$ for every i and j, $1 \leq i < j \leq k$ (i.e., a and b are unrelated), then either $\{a, c\} \perp \{b\}$ or $\{a\} \perp \{b, c\}$.

Proof: Let c_i be a value of c. We say that c_i has a Type I factorization if $p(a, b, c_i) = p(a, c_i)p(b)$ and it has a Type II factorization if $p(a, b, c_i) = p(a)p(b, c_i)$. We shall now prove that every pair of values c_{i_1} and c_{i_2} of c has a common factorization, namely, either both values have a Type I factorization or both have a Type II factorization. This observation completes the proof because it implies that all values of c have a common factorization either of Type I or of Type II. Rearrange the values of c such that c_{i_1} is one value and the disjunction of all other values is considered to be the single second value. According to property B, the value c_{i_1} either has a Type I factorization or a Type II factorization. So is the case with c_{i_2} as well. If both values have the same factorization, then the proof is completed. Suppose, without loss of generality, that c_{i_1} has a Type I factorization and that c_{i_2} has a Type II factorization. Now rearrange the values of c such that $c_{i_1} \vee c_{i_2}$ is one value and the disjunction of the other values is the second value. According to property B, the value $c_{i_1} \vee c_{i_2}$ has a Type I factorization or a Type II factorization. If it has a Type I factorization, and since c_{i_1} has a Type I factorization, it follows, by subtracting the corresponding equations, that also c_{i_2} must have a Type II factorization. Similarly, if $c_{i_1} \vee c_{i_2}$ has a Type II factorization then c_{i_1} must have a Type II factorization as well. □

A straightforward generalization of Theorem 5 to n variables can be phrased as follows. For a pair of variables a and b in U, we define a new variable c whose domain is the Cartesian product of the domains of the variables in $U \setminus \{a, b\}$. Now we say that a and b are unrelated wrt $p(U)$ if $\{a\} \perp \{b\}$, $\{a\} \perp \{b\} | \{c\}$ and $\{a\} \perp \{b\} | c = c_i \vee c_j$ for every two values of the combined variable c. The difficulty with this definition is that it is too strong. Due to Theorem 5, if a and b are unrelated wrt $p(U)$ according to this definition, then either a is independent of $U \setminus \{a\}$ or b is independent of $U \setminus \{b\}$. Consequently, either a or b is unrelated to all other variables and this claim is stronger than saying that a and b are unrelated merely between themselves.

So we conclude with the following definition.

Definition: Let $p(u_1, \cdots, u_n)$ be a probability distribution. Variables u_i and u_j are *unrelated* if $\{u_i\} \perp \{u_j\}$ and $\{u_i\} \perp \{u_j\} | \{v_1 = V_1, \cdots, v_m = V_m\}$ for every disjunction of values $V_1, \cdots, V_m$ for $v_1, \cdots, v_m$, respectively, where $\{v_1, \cdots, v_m\}$ is an arbitrary subset of $\{u_1, \cdots, u_n\} \setminus \{u_i, u_j\}$.

For example, suppose that $U = \{a, b, c, d\}$ and that the domains of c and d are $\{c_1, c_2\}$, and $\{d_1, d_2, d_3\}$, respectively. Then, according to our definition, a and b are unrelated (wrt p) iff $\{a\}\perp\{b\}$, $\{a\}\perp\{b\}|c = c_i (i = 1, 2)$, $\{a\}\perp\{b\}|d = d_j (j = 1, \cdots, 3)$, $\{a\}\perp\{b\}|d = d_{j_1} \vee d_{j_2} (1 \leq j_1 < j_2 \leq 3)$, $\{a\}\perp\{b\}|\{c = c_i, d = d_j\} (i = 1, 2, \; j = 1, \cdots, 3)$, and $\{a\}\perp\{b\}|\{c = c_i, d = d_{j_1} \vee d_{j_2}\} (i = 1, 2, \; 1 \leq j_1 < j_2 \leq 3)$.

Note that when $u_1, \cdots, u_n$ are all binary variables, then the sets V_i are singletons, namely, a single specific assignment for v_i. Therefore, for binary variables, our final definition coincides with our first one.

V. Proof of Equivalence

We now show that relatedness and connectedness are equivalent when all variables are binary and when the distribution p is strictly positive.

Definition: A *strictly positive binary* distribution $P(u_1, \cdots, u_n)$ is a probability distribution where every variable has a domain of two values—say, *true* and *false*—and every combination of the variables' values has a probability greater than zero.

First we must generalize property B.

Theorem 6: Let $p(u_1, \cdots, u_n, e)$ be a strictly positive binary distribution. Let $\{X_1, X_2\}$, $\{Y_1, Y_2\}$, and $\{Z_1, Z_2\}$ be three partitions of $U = \{u_1, \cdots, u_n\}$. Let R_1 be $X_1 \cap Y_1 \cap Z_1$, and R_2 be $X_2 \cap Y_2 \cap Z_2$. Then

$$\begin{aligned} &X_1 \perp X_2 \text{ and } Y_1 \perp Y_2 | e = \mathbf{e}' \text{ and } Z_1 \perp Z_2 | e = \mathbf{e}'' \\ &\quad \Rightarrow R_1 \perp (\{e\} \cup U \setminus R_1) \text{ or } R_2 \perp (\{e\} \cup U \setminus R_2) \end{aligned} \tag{8}$$

where $\mathbf{e}'$ and $\mathbf{e}''$ are two distinct values of e.

When all three partitions are identical, the above theorem can be phrased as follows. If two sets of variables A and B are marginally independent, and if $A \perp B | e$ holds as well, then either $A \perp (\{e\} \cup B)$ or $B \perp (\{e\} \cup A)$. This special case is precisely property B. The proof of Theorem 6 is given in the Appendix.

Theorem 7: Let $p(u_1, \cdots, u_n, u_{n+1})$ be a strictly positive binary distribution. Suppose x and y are in $\{u_1, \cdots, u_{n+1}\}$. Then, x and y are unrelated (wrt p) if and only if x and y are disconnected in some minimal Bayesian network of p.

Proof: If x and y are disconnected in some minimal network, then according to Theorem 4 there exists a partition U_x, U_y of U such that $x \in U_x$ and $y \in U_y$ and $U_x \perp U_y$. Thus, using symmetry (3), decomposition (4) and weak union (5) it follows that $\{x\}\perp\{y\}|Z$ where Z is an arbitrary subset of $U \setminus \{x, y\}$. Thus, x and y are unrelated wrt (p).

The converse is proven by induction on n. If $n = 1$, and x and y are unrelated, then $\{x\}\perp\{y\}$ holds wrt p. Consequently, there exists a partition $U_x = \{x\}$, $U_y = \{y\}$ of U such that $U_x \perp U_y$ and so x and y are disconnected in any minimal network of p. Otherwise, assume without loss of generality that x is u_1 and y is u_2, and denote u_{n+1} by e. Since x and y are unrelated wrt $p(u_1, \cdots, u_{n+1})$, x and y are also unrelated wrt $p(u_1, \cdots, u_n)$, $p(u_1, \cdots, u_n | e = \mathbf{e}')$, and $p(u_1, \cdots, u_n | e = \mathbf{e}'')$, where $\mathbf{e}'$ and $\mathbf{e}''$ are two distinct values of u_{n+1}. Thus, by applying the induction hypothesis three times, we conclude that there are three partitions $\{X_1, X_2\}$, $\{Y_1, Y_2\}$, and $\{Z_1, Z_2\}$ of $U = \{u_1, \cdots, u_n\}$ such that x is in X_1, Y_1, and Z_1, and y is in X_2, Y_2, and Z_2. Hence, the antecedents of (8) are satisfied. Consequently, $\{u_1, \cdots, u_{n+1}\}$ can be partitioned into two marginally independent sets: either R_1 and $U \setminus R_1$, or R_2 and $U \setminus R_2$, where R_1 is $X_1 \cap Y_1 \cap Z_1$ and R_2 is $X_2 \cap Y_2 \cap Z_2$. Because, in both cases, one set contains x and the other contains y, it follows that x and y are disconnected. □

VI. Summary

This paper shows that for strictly positive binary distributions the notion of probabilistic relatedness as defined herein is equivalent to the notion of connectedness in minimal Bayesian networks. We conjecture that the equivalence established holds also when these restrictions are lifted.

Appendix

Below, we prove Theorem 6. First, we phrase the theorem differently.

Theorem 8: Strictly positive binary distributions satisfy the following property:[3]

$$\begin{aligned} &(A_1A_2A_3A_4 \perp B_1B_2B_3B_4 | \emptyset) \\ &\quad \&(A_1A_2B_3B_4 \perp B_1B_2A_3A_4 | e = \mathbf{e}') \\ &\quad \&(A_1A_3B_2B_4 \perp B_1B_3A_2A_4 | e = \mathbf{e}'') \\ &\quad \Rightarrow (A_1 \perp eA_2A_3A_4B_1B_2B_3B_4 | \emptyset) \\ &\quad \vee (B_1 \perp eA_1A_2A_3A_4B_2B_3B_4 | \emptyset) \end{aligned} \tag{9}$$

where all sets mentioned are pairwise disjoint and do not contain e, and $\mathbf{e}'$ and $\mathbf{e}''$ are distinct values of e.

To obtain the original theorem, we set $A_1A_2A_3A_4$, $B_1B_2B_3B_4$, $A_1A_2B_3B_4$, $B_1B_2A_3A_4$, $A_1A_3B_2B_4$, and $B_1B_3A_2A_4$ to be equal to X_1, X_2, Y_1, Y_2, Z_1, and Z_2 of the original theorem, respectively.

Denote the three antecedents of (9) by I_1, I_2, and I_3. We need the following two Lemmas.

Lemma 9: Let X and Y be two disjoint sets of variables, and let $\mathbf{e}$ be an instance of a single binary variable e not in $X \cup Y$. Let p be a joint probability distribution of the variables $X \cup Y \cup \{e\}$. If $(X \perp Y | e = \mathbf{e})$ holds for p, then for every pair of instances $\mathbf{X}'$, $\mathbf{X}''$ of X and $\mathbf{Y}'$, $\mathbf{Y}''$ of Y, the following equation must hold:

$$\frac{p(\mathbf{e}|\mathbf{X}'\mathbf{Y}')p(\mathbf{X}'\mathbf{Y}')}{p(\mathbf{e}|\mathbf{X}''\mathbf{Y}')p(\mathbf{X}''\mathbf{Y}')} = \frac{p(\mathbf{e}|\mathbf{X}'\mathbf{Y}'')p(\mathbf{X}'\mathbf{Y}'')}{p(\mathbf{e}|\mathbf{X}''\mathbf{Y}'')p(\mathbf{X}''\mathbf{Y}'')}.$$

Proof: Bayes' theorem states that

$$p(\mathbf{X}'|\mathbf{e}\mathbf{Y}') = \frac{p(\mathbf{e}|\mathbf{X}'\mathbf{Y}')p(\mathbf{X}'\mathbf{Y}')}{p(\mathbf{e}\mathbf{Y}')}.$$

[3] In complicated expressions, A_1A_2 is used as a shorthand notation for $A_1 \cup A_2$ and eA_1 denotes $\{e\} \cup A_1$.

Thus,

$$\frac{p(\mathbf{e}|\mathbf{X}'\mathbf{Y}')p(\mathbf{X}'\mathbf{Y}')}{p(\mathbf{e}|\mathbf{X}''\mathbf{Y}')p(\mathbf{X}''\mathbf{Y}')} = \frac{p(\mathbf{X}'|\mathbf{e},\mathbf{Y}')}{p(\mathbf{X}''|\mathbf{e},\mathbf{Y}')} = \frac{p(\mathbf{X}'|\mathbf{e},\mathbf{Y}'')}{p(\mathbf{X}''|\mathbf{e},\mathbf{Y}'')} = \frac{p(\mathbf{e}|\mathbf{X}'\mathbf{Y}'')p(\mathbf{X}'\mathbf{Y}'')}{p(\mathbf{e}|\mathbf{X}''\mathbf{Y}'')p(\mathbf{X}''\mathbf{Y}'')}.$$

The middle equality follows from the fact that $(X \perp Y | e = \mathbf{e})$ holds for p. □

Lemma 10: Let A_1, A_2, A_3, A_4, B_1, B_2, B_3, and B_4 be disjoint sets of variables, and e be a single binary variable not contained in any of these sets. Let p be a joint probability distribution of the union of these variables. If the antecedents I_1, I_2, and I_3 of (9) hold for p, then the following conditions must also hold:

$$(A_1 \perp e | \mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') \Rightarrow (A_1 \perp e | \mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4' B_1 B_2 \mathbf{B}_3'\mathbf{B}_4') \tag{10}$$

$$(B_1 \perp e | \mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') \Rightarrow (B_1 \perp e | A_1 A_2 \mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') \tag{11}$$

$$(A_1 \perp e | \mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') \Rightarrow (A_1 \perp e | \mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4' B_1 \mathbf{B}_2' B_3 \mathbf{B}_4') \tag{12}$$

$$(B_1 \perp e | \mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') \Rightarrow (B_1 \perp e | A_1 \mathbf{A}_2' A_3 \mathbf{A}_4'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') \tag{13}$$

$$(A_1 \perp e | \mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') \Rightarrow (A_1 \perp e | \mathbf{A}_2'\mathbf{A}_3' A_4 B_1 \mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') \tag{14}$$

$$(B_1 \perp e | \mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') \Rightarrow (B_1 \perp e | A_1 \mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_2'\mathbf{B}_3' B_4) \tag{15}$$

where each $\mathbf{A}_i'$ and $\mathbf{B}_i'$ denote a specific value for A_i and B_i, respectively. [In words, (10) states that if A_1 and e are conditionally independent for one specific value $\mathbf{B}_1'$ of B_1 and $\mathbf{B}_2'$ of B_2, then they are conditionally independent given every value of B_1 and B_2, provided the values of the other variables remain unaltered. The other five equations have a similar interpretation.]

Proof: First, we prove (10). Then we show that the proofs of (11)–(13) are symmetric. Finally, we will prove (14) and (15). Let $X = A_1 A_2 B_3 B_4$ and $Y = B_1 B_2 A_3 A_4$. Then, Lemma 9 and I_2 yield (15a), shown at the bottom of the page, where $\mathbf{A}_1^*$, $\mathbf{B}_1^*$, and $\mathbf{B}_2^*$ are arbitrary instances of A_1, B_1, and B_2, respectively. Applying I_1 and cancelling equal terms yields

$$\frac{p(\mathbf{e}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')}{p(\mathbf{e}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')} = \frac{p(\mathbf{e}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}_3'\mathbf{B}_4')}{p(\mathbf{e}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}_3'\mathbf{B}_4')}. \tag{16}$$

Furthermore, $(A_1 \perp e | \mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')$ [the antecedent of (10)] implies that

$$p(\mathbf{e}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') = p(\mathbf{e}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4').$$

Thus, from (16), it follows that

$$p(\mathbf{e}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}_3'\mathbf{B}_4') = p(\mathbf{e}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}_3'\mathbf{B}_4'). \tag{17}$$

Subtracting each side of (17) from 1 yields

$$p(\overline{\mathbf{e}}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}_3'\mathbf{B}_4') = p(\overline{\mathbf{e}}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}_3'\mathbf{B}_4'). \tag{18}$$

Thus, $(A_1 \perp e | \mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}_3'\mathbf{B}_4')$ holds for p. Because $\mathbf{B}_1^*$ and $\mathbf{B}_2^*$ are arbitrary instances, $(A_1 \perp e | \mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4' B_1 B_2 \mathbf{B}_3'\mathbf{B}_4')$ also holds for p. Thus, (10) is proved.

Equation (11) is symmetric with respect to (10) by switching the role of A_1 with that of B_1 and the role of A_2 with that of B_2. Equation (12) is symmetric with respect to (10) by switching the roles of B_2 and B_3. Equation (13) is symmetric with respect to (11) by switching the roles of A_2 and A_3.

Now we prove (14). Equation (15) is symmetric with respect to (14) by switching the role of A_1 with that of B_1 and the role of A_4 with that of B_4.

Let $X = A_1 A_2 B_3 B_4$ and $Y = B_1 B_2 A_3 A_4$. Applying Lemma 9 and I_2 and then using I_1 to cancel equal terms, yields the following equation:

$$\frac{p(\mathbf{e}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')p(\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4')}{p(\mathbf{e}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')p(\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4')} = \frac{p(\mathbf{e}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*\mathbf{B}_1^*\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')p(\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*)}{p(\mathbf{e}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*\mathbf{B}_1^*\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')p(\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*)} \tag{19}$$

where $\mathbf{A}_1^*$, $\mathbf{B}_1^*$, and $\mathbf{A}_4^*$ are arbitrary instances of A_1, B_1, and A_4, respectively. Similarly, let $X = A_1 A_3 B_2 B_4$ and $Y = B_1 B_3 A_2 A_4$. Then, applying Lemma 9 and I_3 and using I_1 to cancel equal terms, yields the following equation:

$$\frac{p(\overline{\mathbf{e}}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')p(\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4')}{p(\overline{\mathbf{e}}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')p(\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4')} = \frac{p(\overline{\mathbf{e}}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*\mathbf{B}_1^*\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')p(\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*)}{p(\overline{\mathbf{e}}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*\mathbf{B}_1^*\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')p(\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*)}. \tag{20}$$

Now $(A_1 \perp e | \mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')$ implies the following two conditions:

$$p(\mathbf{e}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') = p(\mathbf{e}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') \tag{21}$$

$$p(\overline{\mathbf{e}}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') = p(\overline{\mathbf{e}}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4'). \tag{22}$$

$$\frac{p(\mathbf{e}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')p(\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')}{p(\mathbf{e}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')p(\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')} = \frac{p(\mathbf{e}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}_3'\mathbf{B}_4')p(\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}_3'\mathbf{B}_4')}{p(\mathbf{e}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}_3'\mathbf{B}_4')p(\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_1^*\mathbf{B}_2^*\mathbf{B}_3'\mathbf{B}_4')} \tag{15a}$$

After using (21) to cancel equal terms in (19) and using (22) to cancel equal terms in (20), we compare (19) and (20) and obtain

$$\frac{p(\mathbf{e}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*\mathbf{B}_1^*\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')}{p(\mathbf{e}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*\mathbf{B}_1^*\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')} = \frac{p(\overline{\mathbf{e}}|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*\mathbf{B}_1^*\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')}{p(\overline{\mathbf{e}}|\mathbf{A}_1^*\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*\mathbf{B}_1^*\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')}. \tag{23}$$

Equation (23) has the form

$$\frac{x}{y} = \frac{1-x}{1-y}$$

which yields $x = y$.

Consequently, we obtain $(A_1 \perp e|\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4^*\mathbf{B}_1^*\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')$. Furthermore, because $\mathbf{A}_4^*$ and $\mathbf{B}_1^*$ are arbitrary instances, $(A_1 \perp e|\mathbf{A}_2'\mathbf{A}_3'A_4B_1\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4')$ holds for p. □

Next, we prove Theorem 6. Let $C = A_2A_3A_4$ and $D = B_2B_3B_4$. We will see that I_1, I_2, and I_3 imply the following four properties:

$$(A_1 \perp e|CDB_1) \text{ or } (B_1 \perp e|CDA_1) \tag{24}$$

$$(A_1 \perp e|CDB_1) \Rightarrow (A_1 \perp A_4|A_2A_3) \tag{25}$$

$$(A_1 \perp e|CDB_1) \;\&\; (A_1 \perp A_4|A_2A_3) \Rightarrow (A_1 \perp A_3|A_2) \tag{26}$$

$$(A_1 \perp e|CDB_1) \;\&\; (A_1 \perp A_4|A_2A_3) \;\&\; (A_1 \perp A_3|A_2) \Rightarrow (A_1 \perp A_2|\emptyset). \tag{27}$$

First, we prove (9), using these four properties. Then, we will show that these properties are valid. From (24), there are two symmetric cases to consider. Without loss of generality, assume $(A_1 \perp e|CDB_1)$ holds. [Otherwise, we switch the roles of subscripted A's with subscripted B's in (25)–(27).] By a single application of each of (25)–(27), the following independence statements are proved to hold for p:

$$(A_1 \perp A_2|\emptyset), (A_1 \perp A_3|A_2), (A_1 \perp A_4|A_2A_3).$$

These three statements yield $(A_1 \perp A_2A_3A_4|\emptyset)$ ($\equiv I_4$) by two applications of contraction. Consider (9). The statement $(A_1A_2A_3A_4 \perp B_1B_2B_3B_4|\emptyset)$ (i.e., I_1) implies $(A_1 \perp B_1B_2B_3B_4|A_2A_3A_4)$ using weak union, which together with I_4 imply using contraction $(A_1 \perp A_2A_3A_4B_1B_2B_3B_4|\emptyset)$. This statement together with the statement $(A_1 \perp e|CB_1D)$ imply, using contraction, the statement $(A_1 \perp eA_2A_3A_4B_1B_2B_3B_4|\emptyset)$, thus completing the proof.

It remains to prove (24)–(27). First, we prove (24). Let $\mathbf{A}'$, $\mathbf{A}''$, $\mathbf{B}'$, $\mathbf{B}''$, $\mathbf{C}^*$, and $\mathbf{D}^*$ be arbitrary instances of A_1, B_1, C, and D, respectively. Let $X = AC$ and $Y = BD$. Then, Lemma 9 and I_2 yield the following equation:

$$\frac{p(\mathbf{e}|\mathbf{A}'\mathbf{C}^*\mathbf{D}^*\mathbf{B}')p(\mathbf{A}'\mathbf{C}^*\mathbf{D}^*\mathbf{B}')}{p(\mathbf{e}|\mathbf{A}''\mathbf{C}^*\mathbf{D}^*\mathbf{B}')p(\mathbf{A}''\mathbf{C}^*\mathbf{D}^*\mathbf{B}')} = \frac{p(\mathbf{e}|\mathbf{A}'\mathbf{C}^*\mathbf{D}^*\mathbf{B}'')p(\mathbf{A}'\mathbf{C}^*\mathbf{D}^*\mathbf{B}'')}{p(\mathbf{e}|\mathbf{A}''\mathbf{C}^*\mathbf{D}^*\mathbf{B}'')p(\mathbf{A}''\mathbf{C}^*\mathbf{D}^*\mathbf{B}'')}. \tag{28}$$

From I_1, we obtain $p(ACDB) = p(AC)p(DB)$. Consequently, (28) yields

$$\frac{p(\mathbf{e}|\mathbf{A}'\mathbf{B}'\mathbf{C}^*\mathbf{D}^*)}{p(\mathbf{e}|\mathbf{A}''\mathbf{B}'\mathbf{C}^*\mathbf{D}^*)} = \frac{p(\mathbf{e}|\mathbf{A}'\mathbf{B}''\mathbf{C}^*\mathbf{D}^*)}{p(\mathbf{e}|\mathbf{A}''\mathbf{B}''\mathbf{C}^*\mathbf{D}^*)}. \tag{29}$$

Equation (29) has the following algebraic form, where subscripted Xs replace the corresponding terms

$$\frac{X_{\mathbf{A}'\mathbf{B}'}}{X_{\mathbf{A}''\mathbf{B}'}} = \frac{X_{\mathbf{A}'\mathbf{B}''}}{X_{\mathbf{A}''\mathbf{B}''}}. \tag{30}$$

Using Lemma 9 and I_3, we obtain a relationship similar to (29), where the only change is that $\mathbf{e}$ is replaced with $\overline{\mathbf{e}}$

$$\frac{p(\overline{\mathbf{e}}|\mathbf{A}'\mathbf{B}'\mathbf{C}^*\mathbf{D}^*)}{p(\overline{\mathbf{e}}|\mathbf{A}''\mathbf{B}'\mathbf{C}^*\mathbf{D}^*)} = \frac{p(\overline{\mathbf{e}}|\mathbf{A}'\mathbf{B}''\mathbf{C}^*\mathbf{D}^*)}{p(\overline{\mathbf{e}}|\mathbf{A}''\mathbf{B}''\mathbf{C}^*\mathbf{D}^*)}. \tag{31}$$

We rewrite (31) in terms of Xs, and then use (30) to obtain

$$\frac{1 - X_{\mathbf{A}'\mathbf{B}'}}{1 - X_{\mathbf{A}''\mathbf{B}'}} = \frac{1 - kX_{\mathbf{A}'\mathbf{B}'}}{1 - kX_{\mathbf{A}''\mathbf{B}'}} \tag{32}$$

where $k = (X_{\mathbf{A}'\mathbf{B}''}/X_{\mathbf{A}'\mathbf{B}'})$. Equation (32) implies that either $X_{\mathbf{A}'\mathbf{B}''} = X_{\mathbf{A}'\mathbf{B}'}$ (i.e., $k = 1$) or $X_{\mathbf{A}'\mathbf{B}'} = X_{\mathbf{A}''\mathbf{B}'}$. Because the choice of instances for A_1 and B_1 is arbitrary, at least one of the following two sequences of equalities must hold

- for every instance $\mathbf{B}$ of B_1, $X_{\mathbf{A}^1\mathbf{B}} = X_{\mathbf{A}^2\mathbf{B}} = \cdots = X_{\mathbf{A}^m\mathbf{B}}$;
- for every instance $\mathbf{A}$ of A_1, $X_{\mathbf{A}\mathbf{B}^1} = X_{\mathbf{A}\mathbf{B}^2} = \cdots = X_{\mathbf{A}\mathbf{B}^n}$;

where $\mathbf{A}^1, \cdots, A^m$ are the instances of A_1 and $\mathbf{B}^1, \cdots, \mathbf{B}^n$ are the instances of B_1.

Thus, by definition of the Xs, we obtain

$$\forall\, \mathbf{C}^*\mathbf{D}^* \text{ instances of } CD\ [(e \perp A_1|\mathbf{C}^*\mathbf{D}^*B_1) \text{ or } (e \perp B_1|\mathbf{C}^*\mathbf{D}^*A_1)]. \tag{33}$$

On the other hand, (24), which we are now proving, states

$$[\forall \mathbf{C}^*\mathbf{D}^*(e \perp A_1|\mathbf{C}^*\mathbf{D}^*B_1)] \text{ or } [\forall \mathbf{C}^*\mathbf{D}^*(e \perp B_1|\mathbf{C}^*\mathbf{D}^*A_1)] \tag{34}$$

which is stronger than (33). Equation (24) can also be written as follows:

$$\neg(B_1 \perp e|CDA_1) \Rightarrow (A_1 \perp e|CDB_1). \tag{35}$$

We prove (35). The statement $\neg(B_1 \perp e|CDA_1)$ implies that there exists instances $\mathbf{A}_1'$, $\mathbf{A}_2'$, $\mathbf{A}_3'$, $\mathbf{A}_4'$, $\mathbf{B}_1'$, $\mathbf{B}_2'$, $\mathbf{B}_3'$, $\mathbf{B}_4'$, and $\mathbf{e}'$ of A_1, A_2, A_3, A_4, B_1, B_2, B_2, B_3, B_4, and e, respectively, such that

$$\neg(\mathbf{B}_1' \perp \mathbf{e}'|\mathbf{A}_1'\mathbf{A}_2'\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4'). \tag{36}$$

Hence,

$$\neg(B_1 \perp e|A_1A_2\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4'). \tag{37}$$

From Lemma 10 [contrapositive form of (11)], (37) implies

$$\neg(B_1 \perp e|\mathbf{A}_1^*\mathbf{A}_2^*\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') \tag{38}$$

where $\mathbf{A}_1^*$ and $\mathbf{A}_2^*$ are arbitrary instances of A_1 and A_2, respectively. Hence, in particular, if $\mathbf{A}_1^* = \mathbf{A}_1'$, we have

$$\neg(B_1 \perp e|\mathbf{A}_1'\mathbf{A}_2^*\mathbf{A}_3'\mathbf{A}_4'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4'). \tag{39}$$

Similarly, from Lemma 10 (13), (39) implies

$$\neg(B_1 \perp e|\mathbf{A}_1^*\mathbf{A}_2^*\mathbf{A}_3^*\mathbf{A}_4'\mathbf{B}_2'\mathbf{B}_3'\mathbf{B}_4') \tag{40}$$

where $\mathbf{A}_3^*$ is an arbitrary instance of A_3. Also, from Lemma 10 (15), (40) implies

$$\neg(B_1 \perp e | \mathbf{A}_1^* \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}_4' \mathbf{B}_2' \mathbf{B}_3' \mathbf{B}_4^*) \tag{41}$$

where $\mathbf{B}_4^*$ is an arbitrary instance of B_4. Examine (33). Equation (41) states that the second disjunct cannot be true for every instance of $A_1 A_2 A_3 B_1 B_4$ and e. Hence, for each of these instances the other disjunct must hold. That is,

$$\forall\, \mathbf{A}_1^* \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{B}_1^* \mathbf{B}_4^* \mathbf{e}^*\ (\mathbf{A}_1^* \perp \mathbf{e}^* | \mathbf{B}_1^* \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}_4' \mathbf{B}_2' \mathbf{B}_3' \mathbf{B}_4^*) \tag{42}$$

or, equivalently,

$$(A_1 \perp e | B_1 A_2 A_3 \mathbf{A}_4' \mathbf{B}_2' \mathbf{B}_3' B_4). \tag{43}$$

Applying (43) to (10) yields,

$$(A_1 \perp e | B_1 A_2 A_3 \mathbf{A}_4' B_2 \mathbf{B}_3' B_4). \tag{44}$$

Similarly, applying (44) to (12) and (14) yields the statement

$$(A_1 \perp e | A_2 A_3 A_4 B_1 B_2 B_3 B_4) \tag{45}$$

which is the desired consequence of (35). Thus, we have proved (24).

Next, we show that (25) must hold. Lemma 9 and I_2 yield (46), shown at the bottom of the page. Incorporating I_1 and $(e \perp A_1 | A_2 A_3 A_4 B_1 B_2 B_3 B_4)$ (45), and cancelling some equal terms yields

$$\frac{p(\mathbf{A}_4' | \mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3^*) p(\mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3^*) p(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{p(\mathbf{A}_4' | \mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3^*) p(\mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3^*) p(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} = \frac{p(\mathbf{A}_4'' | \mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3^*) p(\mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3^*) p(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{p(\mathbf{A}_4'' | \mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3^*) p(\mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3^*) p(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}.$$

Further cancellation of equal terms yields

$$\frac{p(\mathbf{A}_4' | \mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3^*)}{p(\mathbf{A}_4'' | \mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3^*)} = \frac{p(\mathbf{A}_4' | \mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3^*)}{p(\mathbf{A}_4'' | \mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3^*)}.$$

Thus, $p(\mathbf{A}_4' | \mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3^*) = p(\mathbf{A}_4' | \mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3^*)$ for every instance $\mathbf{A}_1'$, $\mathbf{A}_1''$, and $\mathbf{A}_4'$. That is, $(A_4 \perp A_1 | \mathbf{A}_2^* \mathbf{A}_3^*)$ holds. Because $\mathbf{A}_2^*$ and $\mathbf{A}_3^*$ are arbitrary instances, $(A_4 \perp A_1 | A_2 A_3)$ follows.

Next, we show that (26) must hold. Lemma 9 and I_2 yield (47), shown at the bottom of the page. Incorporating I_1, $(A_1 \perp A_4 | A_2 A_3)$, and $(e \perp A_1 | A_2 A_3 A_4 B_1 B_2 B_3 B_4)$ and cancelling some equal terms yields (48), shown at the bottom of the page. Further cancellation of equal terms yields

$$\frac{p(\mathbf{A}_3' | \mathbf{A}_1' \mathbf{A}_2^*)}{p(\mathbf{A}_3'' | \mathbf{A}_1' \mathbf{A}_2^*)} = \frac{p(\mathbf{A}_3' | \mathbf{A}_1'' \mathbf{A}_2^*)}{p(\mathbf{A}_3'' | \mathbf{A}_1'' \mathbf{A}_2^*)}.$$

Thus, $p(\mathbf{A}_3' | \mathbf{A}_1' \mathbf{A}_2^*) = p(\mathbf{A}_3' | \mathbf{A}_1'' \mathbf{A}_2^*)$ for every instance $\mathbf{A}_1'$, $\mathbf{A}_1''$, and $\mathbf{A}_3'$. That is, $(A_3 \perp A_1 | \mathbf{A}_2^*)$ holds. Because $\mathbf{A}_2^*$ is an arbitrary instance, $(A_3 \perp A_1 | A_2)$ follows.

Finally, we must show that (27) holds. Lemma 9 and I_3 yield (49), shown at the bottom of the page. Incorporating $(e \perp A_1 | A_2 A_3 A_4 B_1 B_2 B_3 B_4)$, I_1, I_3, $(A_1 \perp A_4 | A_2 A_3)$, and $(A_1 \perp A_3 | A_2)$ and cancelling some equal terms yields

$$\frac{p(\mathbf{e} | \mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}_4' \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) p(\mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}_4' \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{p(\mathbf{e} | \mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}_4' \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) p(\mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}_4' \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} = \frac{p(\mathbf{e} | \mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}_4'' \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) p(\mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}_4'' \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{p(\mathbf{e} | \mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}_4'' \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) p(\mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3^* \mathbf{A}_4'' \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} \tag{46}$$

$$\frac{p(\mathbf{e} | \mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3' \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) p(\mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3' \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{p(\mathbf{e} | \mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3' \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) p(\mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3' \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} = \frac{p(\mathbf{e} | \mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3'' \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) p(\mathbf{A}_1' \mathbf{A}_2^* \mathbf{A}_3'' \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{p(\mathbf{e} | \mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3'' \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) p(\mathbf{A}_1'' \mathbf{A}_2^* \mathbf{A}_3'' \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} \tag{47}$$

$$\frac{p(\mathbf{A}_4^* | \mathbf{A}_2^* \mathbf{A}_3') p(\mathbf{A}_3' | \mathbf{A}_1' \mathbf{A}_2^*) p(\mathbf{A}_1' \mathbf{A}_2^*) p(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{p(\mathbf{A}_4^* | \mathbf{A}_2^* \mathbf{A}_3') p(\mathbf{A}_3' | \mathbf{A}_1'' \mathbf{A}_2^*) p(\mathbf{A}_1'' \mathbf{A}_2^*) p(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} = \frac{p(\mathbf{A}_4^* | \mathbf{A}_2^* \mathbf{A}_3'') p(\mathbf{A}_3'' | \mathbf{A}_1' \mathbf{A}_2^*) p(\mathbf{A}_1' \mathbf{A}_2^*) p(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{p(\mathbf{A}_4^* | \mathbf{A}_2^* \mathbf{A}_3'') p(\mathbf{A}_3'' | \mathbf{A}_1'' \mathbf{A}_2^*) p(\mathbf{A}_1'' \mathbf{A}_2^*) p(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} \tag{48}$$

$$\frac{p(\bar{\mathbf{e}} | \mathbf{A}_1' \mathbf{A}_2' \mathbf{A}_3^* \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) p(\mathbf{A}_1' \mathbf{A}_2' \mathbf{A}_3^* \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{p(\bar{\mathbf{e}} | \mathbf{A}_1'' \mathbf{A}_2' \mathbf{A}_3^* \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) p(\mathbf{A}_1'' \mathbf{A}_2' \mathbf{A}_3^* \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} = \frac{p(\bar{\mathbf{e}} | \mathbf{A}_1' \mathbf{A}_2'' \mathbf{A}_3^* \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) p(\mathbf{A}_1' \mathbf{A}_2'' \mathbf{A}_3^* \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{p(\bar{\mathbf{e}} | \mathbf{A}_1'' \mathbf{A}_2'' \mathbf{A}_3^* \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*) p(\mathbf{A}_1'' \mathbf{A}_2'' \mathbf{A}_3^* \mathbf{A}_4^* \mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} \tag{49}$$

$$\frac{p(\mathbf{A}_4^* | \mathbf{A}_2' * \mathbf{A}_3^*) p(\mathbf{A}_3^* | \mathbf{A}_1' \mathbf{A}_2') p(\mathbf{A}_2' | \mathbf{A}_1') p(\mathbf{A}_1') p(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{p(\mathbf{A}_4^* | \mathbf{A}_2' * \mathbf{A}_3^*) p(\mathbf{A}_3^* | \mathbf{A}_1'' \mathbf{A}_2') p(\mathbf{A}_2' | \mathbf{A}_1'') p(\mathbf{A}_1'') p(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} = \frac{p(\mathbf{A}_4^* | \mathbf{A}_2'' * \mathbf{A}_3^*) p(\mathbf{A}_3^* | \mathbf{A}_1' \mathbf{A}_2'') p(\mathbf{A}_2'' | \mathbf{A}_1') p(\mathbf{A}_1') p(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)}{p(\mathbf{A}_4^* | \mathbf{A}_2'' * \mathbf{A}_3^*) p(\mathbf{A}_3^* | \mathbf{A}_1'' \mathbf{A}_2'') p(\mathbf{A}_2'' | \mathbf{A}_1'') p(\mathbf{A}_1'') p(\mathbf{B}_1^* \mathbf{B}_2^* \mathbf{B}_3^* \mathbf{B}_4^*)} \tag{50}$$

(50), shown at the bottom of the preceding page. Further cancellation of equal terms yields

$$\frac{p(\mathbf{A}_2'|\mathbf{A}_1')}{p(\mathbf{A}_2''|\mathbf{A}_1')} = \frac{p(\mathbf{A}_2'|\mathbf{A}_1'')}{p(\mathbf{A}_2''|\mathbf{A}_1'')}.$$

Thus, $p(\mathbf{A}_2'|\mathbf{A}_1') = p(\mathbf{A}_2'|\mathbf{A}_1'')$ for every instance $\mathbf{A}_1'$, $\mathbf{A}_1''$, and $\mathbf{A}_2'$. That is, $(A_2 \perp A_1 | \emptyset)$ holds. □

References

[1] A. P. Dawid, "Conditional independence in statistical theory," *J. Roy. Stat. Soc. B,* vol. 41, pp. 1–31, 1979.

[2] D. Geiger, "Graphoids: A qualitative framework for probabilistic inference," Ph.D. dissertation, Dept. Comput. Sci., Univ. Calif., Los Angeles, 1990.

[3] D. Geiger and D. Heckerman, "Knowledge representation and inference in similarity networks and Bayesian multinets," *Artif. Intell.,* vol. 82, pp. 45–74, 1996.

[4] ——, "Separable and transitive graphoids," in *Proc. 6th Conf. Uncertainty Artificial Intelligence D,* 1990.

[5] D. Geiger, T. S. Verma, and J. Pearl, "Identifying independence in Bayesian networks," *Networks,* vol. 20, pp. 507–534, 1990.

[6] D. Geiger and J. Pearl, "On the logic of causal models," in *Uncertainty in Artificial Intelligence 4,* R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, Eds. Amsterdam, The Netherlands, Elsevier, 1990, pp. 3–12.

[7] D. Heckerman, *Probabilistic Similarity Networks.* Cambridge, MA: MIT Press, 1991.

[8] S. L. Lauritzen, A. P. Dawid, B. N. Larsen, and H. G. Leimer, "Independence properties of directed Markov fields," *Networks,* vol. 20, pp. 491–550, 1990.

[9] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, CA: Morgan Kaufmann, 1988.

[10] J. Pearl and A. Paz, "Graphoids: A graph-based logic for reasoning about relevance relations," in *Advances in Artificial Intelligence-II,* B. Du Boulay *et al.,* Eds. Amsterdam, The Netherlands: North-Holland, 1990, pp. 357–363.

[11] R. Shachter, "An ordered examination of influence diagrams," *Networks,* vol. 20, pp. 535–564, 1990.

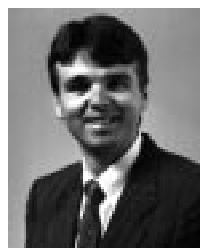
[12] W. Spohn, "Stochastic independence, causal independence, and shieldability," *J. Phil. Logic,* vol. 9, pp. 73–99, 1980.

[13] T. Verma and J. Pearl, "Causal networks: Semantics and expressiveness," in *Proc. 4th Workshop Uncertainty Artificial Intelligence,* Minneapolis, MN, 1988, pp. 352–359; also, T. Verma, Tech. Rep. R-65, Cognitive Systems Lab., Univ. Calif., Los Angeles, 1986.

Dan Geiger was born in Haifa, Israel, in 1959. He received the B.Sc. degree in computer engineering and the B.A. degree in mathematics from the Technion—Israel Institute of Technology, Haifa, in 1985, and the M.Sc. and Ph.D. degrees from the University of California, Los Angeles, in 1987 and 1990, respectively.

His research is focused on the study of conditional independence, learning Bayesian networks, discrete optimization problems in probability and statistics, and genetic linkage analysis.

Dr. Geiger received the Technion's 1997 Taub Award for excellence in research and the 1997 AAAI Best Paper Award for "A practical algorithm for finding optimal triangulations."



David Heckerman received the Ph.D. degree in medical information sciences from Stanford University, Stanford, CA, in 1990.

Since 1992, he has been a Senior Researcher at Microsoft Research, Redmond, WA, where he has been developing methods for learning from data using graphical models.

Dr. Heckerman received the outstanding computer-science dissertation award from the ACM in 1990 for his paper "Probabilistic similarity networks."