Conditional Dependence in Probabilistic Networks

Remco R. Bouckaert

RUU-CS-92-34
1992

Utrecht University

Department of Computer Science

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

Conditional Dependence in Probabilistic Networks

Remco R. Bouckaert

Technical Report RUU-CS-92-34
1992

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

ISSN: 0024-3275

Conditional Dependence in Probabilistic Networks*

Remco R. Bouckaert

Department of Computer Science, Utrecht University,
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands,
E-Mail: remco@cs.ruu.nl

Abstract

In general, a probabilistic network is considered a representation of a set of conditional independency statements. However, probabilistic networks also represent dependencies. In this paper an axiomatic characterization of conditional dependence is given. Furthermore, a criterion is given to read conditional dependencies from a probabilistic network.

Keywords: probabilistic network, conditional independence, causal input list, minimal I-map.

1 Introduction

The graphical representation of probabilistic relationships between variables gets a lot of attention in different areas of research the last years. A probabilistic networks, also known by the name belief network [5], causal network [4], and influence diagram [9], is such a representation. In the field of artificial intelligence systems have been developed for efficient computation of inferences [4, 5, 9] using probabilistic networks.

Independencies between variables are practically indispensable when making inferences in large knowledge-based systems. Probabilistic networks are a powerful means for representing conditional independency statements on variables. A probabilistic network has associated a semantics that allows for reading independencies between variables [3, 5]. Independencies read from the network can be used to decide on relevance of variables to inference problems.

However, in a probabilistic network it is not generally true that if variables are not shown to be independent they are actually dependent. In other words, there may exist independencies that cannot be read from the network. Therefore, it is interesting to find out where these hidden independencies reside in the network. In order to do so, we consider variables that are definitely dependent in a given network.

In Section 2, we give an overview of conditional *in*dependence and its relation with probabilistic networks. In Section 3, we study properties of conditional dependence and in Section 4, we present a criterion for reading dependencies from a probabilistic network.

2 Conditional independence

We consider a joint probability distribution P over a set of variables U. In this paper we use capital letters to denote sets of variables and lower case letters to denote single variables. All variables or

*To appear in the Proceedings of the 4th International Workshop on AI and Statistics

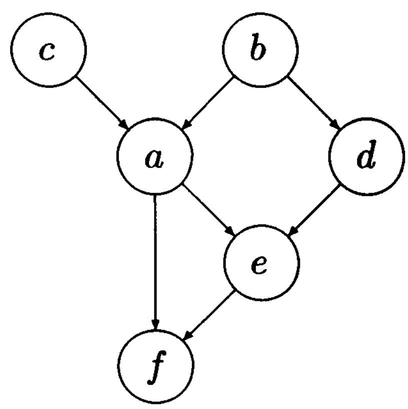


Figure 1: DAG in which $<c, \emptyset, d>$, $<c, b, d>$ and $<c, ab, d>$, but not $<c, a, d>$ or $<c, abf, d>$.

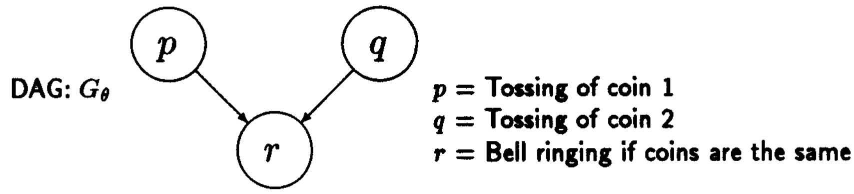
sets of variables mentioned are elements or subsets of U unless stated otherwise. We call X and Y *conditionally independent* given Z, written $I(X, Z, Y)$, if $P(XY|Z) = P(X|Z)P(Y|Z)$ for all values of the variables in XYZ (for sets we write XY to denote the union of X and Y); $I(X, Z, Y)$ is called an *independency statement.* By definition $I(X, Z, \emptyset)$ for any X and Z. An *independency model* over U is a set of independency statements. A *complete independency model* M_I of a distribution P over U is the set of all valid independency statements in P. For positive definite distributions, the following axioms called *independency axioms* apply [1, 7, 8].

symmetry	$I(X, Z, Y)$	$\Leftrightarrow I(Y, Z, X)$
decomposition	$I(X, Z, WY)$	$\Rightarrow I(X, Z, Y)$
weak union	$I(X, Z, WY)$	$\Rightarrow I(X, ZW, Y)$
contraction	$I(X, ZW, Y) \wedge I(X, Z, W)$	$\Rightarrow I(X, Z, WY)$
intersection	$I(X, ZW, Y) \wedge I(X, ZY, W)$	$\Rightarrow I(X, Z, WY)$

With these axioms independency statements can be derived from other independency statements. For instance, let M_I be an independency model for a given distribution P such that $I(a, b, c) \in M_I$. Then, by symmetry we have that $I(c, b, a)$ must also be in M_I. We sometimes omit braces to prevent an overflow of them. So, we write $I(a, b, c)$ for $I(\{a\}, \{b\}, \{c\})$.

A directed acyclic graph (DAG) is a directed graph that does not contain paths starting and ending at the same node. A *trail* in a DAG is a path that does not consider the direction of the arcs. We denote a trail by the ordered sequence of nodes that are in the trail. For example, in the DAG in Figure 1 *cabd* is a trail. A *head-to-head node* in a trail is a triple of consecutive nodes x, y, z in the trail such that $x \rightarrow y \leftarrow z$ in the DAG. A *probabilistic network* is a pair (G, Γ) where G is a DAG and Γ is a set for every variable $u \in U$ of conditional probability tables $P(u|\pi_u)$ that enumerate the probabilities of all values of u given the values of its parents π_u in the DAG. The distribution represented by this network is $\prod_{u \in U} P(u|\pi_u)$ [5]. Independency statements that hold in the distribution represented by a probabilistic network can be read from the structure of the DAG using the notions of blocked trail and d-separation [2, 5].

Definition 2.1 *Let G be a DAG. A trail in G between two nodes x and y is* **blocked** *by a set of nodes Z if at least one of the following two conditions hold:*
- *the trail contains a head-to-head node e and $e \notin Z$ and every descendant of e in G is not in Z.*
- *there is a node e in the trail with $e \in Z$ and e is not a head-to-head node in the trail.*



Independency model: $M_I = \{I(p, \emptyset, q), I(p, \emptyset, r), I(q, \emptyset, r) + \text{sym.}\}$
Causal ordering: $\theta = \{p, q, r\}$
Causal input list: $L_\theta = \{I(p, \emptyset, \emptyset), I(p, \emptyset, q), I(r, pq, \emptyset)\}$
Associated model: $M_{I,\theta} = \{I(p, \emptyset, q), I(q, \emptyset, p)\}$

Figure 2: A DAG that is a minimal I-map of the coins and bell-example.

Definition 2.2 *In a DAG G, let X, Y and Z be sets of nodes. We say that X is* **d-separated** *from Y given Z, written* $<X, Z, Y>$, *if every trail between any node* $x \in X$ *and any node* $y \in Y$ *is blocked by Z.*

In Figure 1, an example DAG is depicted. $<c, \emptyset, d>$ and $<c, b, d>$ are valid separation statements since all trails between c and d, i.e., *cabd*, *caed* and *cafed*, are blocked. $<c, ab, d>$ is valid since the trail *cabd* is blocked by $\{a, b\}$ because it contains node b that is not a head-to-head node. The trail *caed* is blocked by $\{ab\}$ since it contains node a that is not a head-to-head node in the trail. However, $<c, a, d>$ is not valid since the trail *cabd* is not blocked by a because it contains node a which is a head-to-head node

A DAG G is an *I-map* of an independency model M_I if $<X, Z, Y>$ in G implies $I(X, Z, Y) \in M_I$; G is a *minimal I-map* of M_I if no arc can be removed from G without destroying its I-mappedness. G is a *D-map* of M_I if $I(X, Z, Y) \in M_I$ implies $<X, Z, Y>$ in G; G is a *perfect map* of M_I if it is both an I-map and a D-map of M_I.

It is not always possible to find a DAG that is a perfect map of a distribution. For example, consider the situation that a bell rings if the outcomes of two tossed coin is the same [5]. Let p and q represent the outcome of the coins and r the ringing of the bell. Just to keep the distribution positive definite, put the bell in a noisy factory such that it is not always clear if it rings or not. Then, all variables are pairwise independent: $M_I = \{I(p, \emptyset, q), I(p, \emptyset, r), I(q, \emptyset, r) + \text{sym.}\}$. But, given the third variable they are dependent. This nonmonotonic behavior cannot perfectly be represented by a DAG because composition ($<X, Z, Y> \wedge <X, Z, W> \Rightarrow <X, Z, WY>$) holds for d-separation [5]. So, if the DAG represents both $I(p, \emptyset, r)$ and $I(p, \emptyset, q)$ then it also represents $I(p, \emptyset, rq)$. The best we can do is represent the model by a minimal I-map as in Figure 2. However, this DAG does not represent the independence between p and r since $<p, \emptyset, r>$ does not hold in the DAG.

A minimal I-map can be constructed from an independency model M_I using the notion of causal input lists [6].

Definition 2.3 *Let θ be a total ordering on U. Let M_I be a complete independency model of a distribution P over U. A* **causal input list** *L_θ over M_I is a set of independency statements suck that for every $x \in U$, L_θ contains exactly one independency statement of the form:*

$$T = I(x, \pi_x, U_x \backslash \pi_x)$$

in which $U_x = \{y | y \in U,\ \theta(y) < \theta(x)\}$ and π_x is the smallest subset of U_x such that T holds in M_I. π_x is called the **parent set** *of x.*

For positive definite distributions, a causal input list can be constructed in $O(|U|^2)$ consultations of M_I: for $x \in U$, we have that a node $y \in U_x$ is in π_x if and only if $I(x, U_x \backslash y, y)$ is not in M_I.

Let $\theta = \{p, q, r\}$ be a causal ordering for the coins and bell example. Then the causal input list would be $L_\theta = \{I(p, \emptyset, \emptyset), I(p, \emptyset, q), I(r, pq, \emptyset)\}$. A DAG G_θ is associated with the ordering θ by letting G_θ be the DAG constructed in the following way: start with an arcless graph and place an arc for each node u from every node in its parent set to node u. For the coins and bell example, the nodes p and q don't have incoming arcs since their parent sets are empty. The parent set of r contains both p and q so node r will get incoming arcs from both these nodes. The result is the graph in Figure 2. Note that the parent set depend on the ordering θ. Let $\theta' = \{p, r, q\}$ be a causal ordering, then the causal input list $L_{\theta'}$ is $\{I(p, \emptyset, \emptyset), I(r, \emptyset, p), I(q, rp, \emptyset)\}$.

An independency model $M_{I,\theta}$ can be associated with a DAG G_θ constructed from a causal input list L_θ by letting $I(X, Z, Y) \in M_{I,\theta}$ if and only if $<X, Z, Y>$ holds in G_θ. Now $M_{I,\theta}$ is the closure of L_θ under the independency axioms (follows from [10]). Furthermore, it is known $M_{I,\theta} \subseteq M_I$ for any θ [6].

3 Conditional dependencies

As we have seen in the previous section, a probabilistic network can be used to represent independencies. However, it is not always possible to find a perfect map for a given distribution. So, $I(X, Z, Y)$ does not always imply $<X, Z, Y>$.

We call X and Y *conditionally dependent* given Z, written $D(X, Z, Y)$, if not $I(X, Z, Y)$; $D(X, Z, Y)$ is called a *dependency statement*. A *dependency model* over U is a set of dependency statements $D(X, Z, Y)$ with X, Z, Y disjoint subsets of U.The *complete dependency model* M_D of a distribution over U is a dependency model containing all dependency statements that hold in the distribution. We define the following *dependency axioms*:

symmetry	$D(X, Z, Y)$	$\Leftrightarrow D(Y, Z, X)$
composition	$D(X, Z, Y)$	$\Rightarrow D(X, Z, WY)$
weak reunion	$D(X, ZW, Y)$	$\Rightarrow D(X, Z, WY)$
extraction	$D(X, Z, WY) \wedge I(X, Z, W)$	$\Rightarrow D(X, ZW, Y)$
extraction+	$D(X, Z, WY) \wedge I(X, ZY, W)$	$\Rightarrow D(X, Z, Y)$
intersection	$D(X, Z, WY) \wedge I(X, ZY, W)$	$\Rightarrow D(X, ZW, Y)$

Theorem 3.1 *For any probability distribution all dependency axioms but intersection hold. Furthermore, for positive definite distributions also the intersection axiom holds.*

Proof: We will only proof that symmetry holds for any probability distribution P. Let $D(X, Z, Y)$ be a valid dependency statement for P. By definition we have $D(X, Z, Y) \Leftrightarrow$ *not* $I(X, Z, Y)$. Now, assume that $I(Y, Z, X)$ holds. Then, by symmetry for independency statements it follows that $I(X, Z, Y)$. This last independency statement is false, however. So, we have *not* $I(Y, Z, X)$ from which it follows that $D(Y, Z, X)$. The proofs for the other axioms are analogues. □

These dependency axioms can be used to deduce new dependency statements from given statements. For example, let M_D be a dependency model for a given distribution P and $D(a, b, c)$ is in M_D. Then, by symmetry we have that $D(c, b, a)$ must also be in M_D.

Lemma 3.1 *For any positive definite distribution P $D(X, Z, Y) \in M_D$ if and only if two nodes $x \in X, y \in Y$ exist such that $D(x, XYZ \backslash \{x, y\}, y) \in M_D$.*

A proof can be found in the appendix. From the lemma it follows that for $D(X,Z,Y)$ for a positive definite distributions, it is sufficient to show that two variables $x \in X$ and $y \in Y$ exist such that $D(x, XYZ\backslash\{x,y\}, y)$ is in the complete dependency model.

Let L_θ be a causal input list over an independency model of a positive definite distribution over U. Then, the *dependency base* associated with L_θ is the set of dependency statements $\Sigma_\theta = \{D(x, \pi_x\backslash\{y\}, y) | x \in U, y \in \pi_x\}$. We define the dependency model associated with θ, denoted as $M_{D,\theta}$, as the closure of the dependency base Σ_0 under the dependency axioms. It can be shown that the closure of Σ_θ under symmetry, weak reunion, composition, extraction and intersection is equal to the closure of Σ_θ under all dependency axioms (see Theorem 4.3).

Theorem 3.2 *Let M_I be a complete independency model of a positive definite distribution over U and M_D be its complete dependency model. Let L_θ be a causal input list over M_I. Let $M_{D,\theta}$ be the dependency model associated with L_θ. Then, $M_{D,\theta} \subseteq M_D$.*

Proof: The property stated in the theorem will be proved by contradiction. Assume for a statement $D(x, \pi_x\backslash y, y) \in \Sigma_\theta$ that it is not in M_D. Then, by definition $I(x, \pi_x\backslash y, y) \in M_I$. $I(x, \pi_x\backslash y, y)$ and $I(x, \pi_x, U_x\backslash\pi_x)$ imply $I(x, \pi_x\backslash y, U_x\backslash(\pi_x\backslash y))$ using contraction. But this is not a valid statement since it implies that π_x was not the smallest subset of U_x such that $I(x, \pi_x, U_x\backslash\pi_x)$. So, all statements in Σ_θ are in M_D. Since the dependency axioms are sound for positive definite distributions, the theorem follows from the definition of $M_{D,\theta}$. □

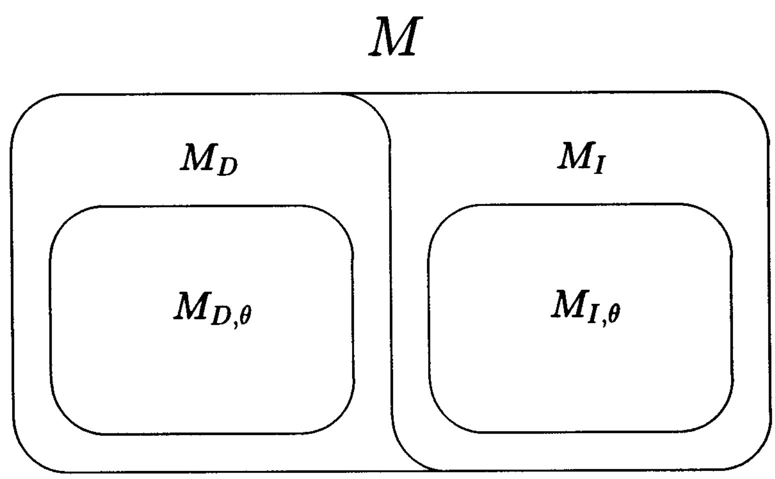


Figure 3: Division of the dependency pool

The *dependency pool* M over U is the set of all triples (X,Z,Y) with X,Y and Z disjoint subsets of U. For a given causal input list L_θ over U, we can divide the pool M into three disjoint sets: $M_{I,\theta}$, $M_{D,\theta}$ and $M\backslash(M_{I,\theta} \cup M_{D,\theta})$ as depicted in Figure 3. Note that from a given causal input list of an independency model, not all independency and dependency statements may be known. So, there are statements for which it cannot be decided from the structure of the graph alone whether it is a independency statement or a dependency statement.

For the coins and bell example, $M_{I,\theta} = \{I(p, \emptyset, q) + \text{sym.}\}$ as shown in Figure 2. The associated dependency model $M_{D,\theta}$ is $\{I(r,q,p), I(r,p,q), I(r,\emptyset,pq), I(p,\emptyset,rq), I(q,\emptyset,rp), + \text{sym.}\}$ which leaves the statements $\{I(p,\emptyset,r), I(q,\emptyset,r), + \text{sym.}\}$ to be unknown.

4 Graphical criterion for conditional dependencies

In Section 2 we have argued that all independency statements in the independency model $M_{I,\theta}$ associated with the causal input list L_θ can be read from G_θ constructed from L_θ using the d-separation criterion. It would be useful to have a similar graphical criterion for reading dependency statements from the DAG G_θ. In this section we investigate the properties of such a criterion.

Consider a DAG G_θ constructed from a causal input list L_θ over an independency model M_I. Let $D(X,Z,Y)$ be a dependency statements in $M_{D,\theta}$. Then, a derivation exists starting with $D(x,\pi_x\backslash y,y)$ and ending with $D(X,Z,Y)$. By structural induction over the steps in the derivation (that use symmetry, composition, weak reunion, extraction and intersection and even extraction+) it can be shown that in every step the following properties are preserved: Let $D(X',Z',Y')$ be the result of a step in the derivation then two nodes $x \in X'$ and $y \in Y'$ or $x \in Y'$ and $y \in X'$ exist such that $y \rightarrow x$ and $\pi_x \subset XYZ$. So, any graphical criterion for reading dependency statements $D(X,Z,Y)$ from G_θ must satisfy these conditions. So, we have the following lemma.

Lemma 4.1 *In a DAG G_θ that is a minimal I-map of an independency model M_I any graphical criterion that holds for a triple (X,Z,Y) if and only if $D(X,Z,Y) \in M_{D,\theta}$ must satisfy the following conditions: two nodes $a \in X \wedge b \in Y$ or $a \in Y \wedge b \in X$ exist such that*
- $b \rightarrow a$.
- $\pi_a \subset XYZ$.

See the appendix for a proof. Some conditional dependency statements can be read from the graph using the following criterion:

Definition 4.1 *In a DAG G, we say that X and Y are* **coupled** *given Z, written $>X,Z,Y<$, if nodes $x \in X and y \in Y$ or $x \in Y and y \in X$ exist such that all following conditions hold:*
- *$y \rightarrow x$ is an arc in G.*
- *$\pi_x \subset XYZ$.*
- *a set Q exists such that $Z \subseteq Q \subseteq XYZ\backslash\{x,y\}$ and $<x,Q,y>$ in G when the arc $y \rightarrow x$ is removed.*

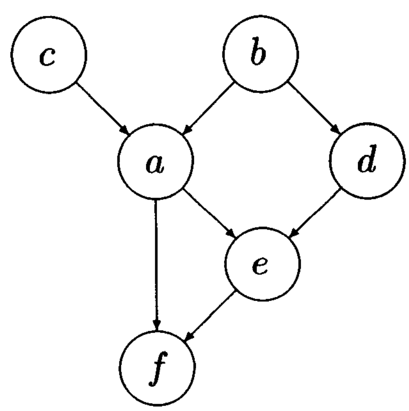


Figure 4: Minimal I-map for which $>a,c,b<$, $>a,cd,b<$ and $>ac,e,db<$, but not $>a,ce,b<$ or $>a,de,b<$.

In Figure 4 some examples of coupling statements are given. We have that $>a,c,b<$ since $b \rightarrow a$ is an arc in the graph, $\pi_a = \{b,c\} \subset \{a,b,c\}$ and all trails between a and b not containing $b \rightarrow a$, i.e., *aedb* and *afedb*, contain a head-to-head node and therefore are blocked. Also we have that $>ac,e,bd<$ since still $b \rightarrow a$ is an arc in G, $\pi_a \subset abcde$ and a set $Q = \{d,e\}$ exists such that all trails are blocked. The statement $>a, ce, b<$ does not hold since the trail *aedb* is not blocked and no set Q that is a subset of *ce* can be found that does so. The statement $>a, de, b<$ does not hold since π_a is not subset of $\{a, d, e, b\}$.

Theorem 4.1 *In a DAG G that is a minimal I-map of an independency model M_I, the following property holds:*

$$>X, Z, Y< \Rightarrow D(X, Z, Y)$$

A proof can be found in the appendix. As a consequence we have that if the first two conditions for coupled hold for a triple of sets (X, Z, Y) and the set Z does not contain descendants of x where x corresponds to the node in the definition then $D(X, Z, Y)$.

The theorem implies that the definition of coupled gives sufficient conditions for reading dependency statements from a minimal I-map. However, it does not give necessary conditions to do this. For example, if the DAG in Figure 4 is a minimal I-map of a model then we know from the dependency base that $D(a, c, b)$, so by symmetry $D(b, c, a)$ and using composition $D(b, c, aef)$. From the DAG we have $<b, ace, f>$ implying $I(b, ace, f)$. And, $D(b, c, aef)$ and $I(b, ace, f)$ imply $D(b, cf, ae)$ using intersection. However, $>b, cf, ae<$ does not hold in the DAG: the trail *aedb* will always be unblocked by a set containing f and not d.

The question that arises is: does a mathematical aesthetic and not too complex criterion exists such that this criterion can be used to read dependency statements from the graph. Such a criterion must contain the following property. As we saw in the previous example, $D(b, cf, ae)$ in the DAG of Figure 4. However, $D(be, cf, a)$ can not be derived using the dependency axioms and the independency statements in the DAG. This example shows that the derivation depends on the place of the variable e in the statement. A graphical criterion must reckon with this possibility.

For dependency statements $D(X, Z, Y)$ where X and Y are single variables this problem does not arise. We have the following result.

Theorem 4.2 *In a DAG G_θ constructed from a causal input list L_θ of a complete independency model M_I, the following property holds:*

$$>x, Z, y< \Leftrightarrow D(x, Z, y) \in M_{D,\theta}$$

For a proof we refer to the appendix. This theorem says that at least for all dependency statements in $M_{D,\theta}$ concerning single nodes can be read from the graph. With the help of these theorems we can show the following theorem.

Theorem 4.3 *The closure of Σ_0 under the dependency axioms is equal to the closure of Σ_0 under all dependency axioms but extraction+.*

For a proof one may consult the appendix. A consequence of this lemma is that no dependency statements would be underivable if $M_{D,\theta}$ if it was closed under de dependency axioms without extraction+.

5 Conclusions

Not all independency statements can always be represented by a probabilistic network. Therefore, if the d-separation criterion does not hold for three sets of variables, it does not necessarily mean the sets are dependent. In this paper we gave an axiomatic characterization of conditional dependencies which is sound. Furthermore, a graphical criterion is presented to read most of the dependency statements from a DAG that is a minimal I-map. It is shown that for some statements it cannot be deduced from the structure of the probabilistic network whether it is an independency or a dependency statement.

Acknowledgement

I thank Linda van der Gaag for helpful comments on earlier drafts resulting in a better presentation of the paper.

Appendix

Lemma 3.1 *For any positive definite distribution P $D(X,Z,Y) \in M_D$ if and only if two nodes $x \in X, y \in Y$ exist such that $D(x, XYZ\backslash\{x,y\}, y) \in M_D$.*

Proof: First we show the $\Leftarrow$ part. Assume that two nodes $x \in X$ and $y \in Y$ exist such that $D(x, XYZ\backslash\{x,y\}, y)$. Then,

$$
\begin{array}{l}
D(x, XYZ\backslash\{x,y\}, y) \\
\Rightarrow \ \{ \text{ Weak reunion } \} \\
D(x, XZ\backslash\{x\}, Y) \\
\Rightarrow \ \{ \text{ Symmetry } \} \\
D(Y, ZX\backslash\{x\}, x) \\
\Rightarrow \ \{ \text{ Weak reunion } \} \\
D(Y, Z, X) \\
\Rightarrow \ \{ \text{ Symmetry } \} \\
D(X, Z, Y)
\end{array}
$$

Now the $\Rightarrow$ part. We assume $D(X,Z,Y)$ that holds, Let $X = \{x_1, ..., x_n\}$, $Y = \{y_1, ..., y_m\}$. Now suppose that for all $x \in X$, $y \in Y$, the statement $I(x, XYZ\backslash\{x,y\}, y)$ is valid. Then,

$$
\begin{array}{l}
\forall_{x \in X} I(x, XYZ\backslash\{x, y_1\}, y_1) \\
\Rightarrow \ \{ \text{ Intersection with } I(x, XYZ\backslash\{x, y_2\}, y_2) \} \\
\forall_{x \in X} I(x, XYZ\backslash\{x, y_1, y_2\}, y_1 y_2) \\
\Rightarrow \ \{ \text{ Intersection with } I(x, XYZ\backslash\{x, y_3\}, y_3) \} \\
\forall_{x \in X} I(x, XYZ\backslash\{x, y_1, y_2, y_3\}, y_1 y_2 y_3) \\
\qquad\qquad \cdots \\
\forall_{x \in X} I(x, XYZ\backslash\{x, y_1, y_2, .., y_{m-1}\}, y_1 y_2 .. y_{m-1}) \\
\Rightarrow \ \{ \text{ Intersection with } I(x, XYZ\backslash\{x, y_m\}, y_m) \} \\
\forall_{x \in X} I(x, XZ\backslash\{x\}, Y) \\
\Rightarrow \ \{ \text{ Symmetry } \} \\
\forall_{x \in X} I(Y, XZ\backslash\{x\}, x)
\end{array}
$$

A similar observation holds for $x_1, \ldots, x_n$. So, we can derive $I(X,Z,Y)$ which contradicts the assumption $D(X,Z,Y)$. Therefore, the assumption that for all $x \in X, y \in Y$ the statements $I(x, XYZ\backslash\{x,y\}, y)$ holds is false. □

Lemma 4.1: *In a DAG G_θ that is a minimal I-map of an independency model M_I any graphical criterion that holds for a triple (X,Z,Y) if and only if $D(X,Z,Y) \in M_{D,\theta}$ must satisfy the following conditions: two nodes $x \in X \wedge y \in Y$ or $x \in Y \wedge y \in X$ exist such that*

- *$y \rightarrow x$ is an arc in G_θ.*
- $\pi_x \subset XYZ$.

Proof: $D(X,Z,Y) \in M_{D,\theta}$ implies by definition that a derivation $\sigma_0 \overset{\gamma_0}{\Rightarrow} \sigma_1 \overset{\gamma_1}{\Rightarrow} \dots \overset{\gamma_n}{\Rightarrow} \gamma_n$ exists where σ_i is a clause of the form $D(A,B,C)$ or $D(A,B,C), I(E,F,G)$ and γ_i is one of the dependency axioms. Furthermore, σ_0 contains the statement $D(a, \pi_a \backslash b, b)$ and $\sigma_n = D(X,Z,Y)$.

We proof the lemma by structural induction over the steps in the derivation. If the derivation is of length 0 ($n = 0$) it is obviously true. If the derivation is of length $n >= 1$ assume the lemma is true for derivations of length $n-1$. We make distinction between the axiom used for γ_n.

- (symmetry) $D(X,Z,Y) \Leftrightarrow D(Y,Z,X)$. By induction hypothesis in σ_{n-1} there are $x \in X, y \in Y$ or $x \in Y, y \in X$ such that the two conditions hold. Therefore, the conditions also hold for $x \in Y, y \in X$ or $x \in X, y \in Y$.

- (composition) $D(X,Z,Y) \Rightarrow D(X,Z,WY)$. Likewise.

- (weak reunion) $D(X,ZW,Y) \Rightarrow D(X,Z,WY)$. Likewise.

- (extraction+) $D(X,Z,WY) \wedge I(X,ZY,W) \Rightarrow D(X,Z,Y)$. Let $x \in X, y \in WY$ such that the two conditions hold. If $y \in W$ then $<X,ZY,W>$ would not hold because by condition 1 ($y \rightarrow x$) $x \in X$ and $y \in W$ are adjacent. So $y \in Y$ and condition 1 still holds in σ_n containing $D(X,Z,Y)$. Let $\pi_x \subset WXYZ$ and let there be a node $w \in W$ such that $w \in \pi_x$. Then $<X,ZY,W>$ would not hold since $x \in X$ and $w \in W$ are adjacent. So no node $w \in W$ is in π_x. This implies $\pi_x \subset XYZ$ so the second condition also holds for σ_n.

  Let $x \in WY, y \in X$ such that the two conditions hold. If $x \in W$ then $<X,ZY,W>$ would not hold because by condition 1 ($y \rightarrow x$) $x \in W$ and $y \in X$ are adjacent. So $x \in Y$ and condition 1 still holds in σ_n containing $D(X,Z,Y)$. Let $\pi_x \subset WXYZ$ and let there be a node $w \in W$ such that $w \in \pi_x$. Then $<X,ZY,W>$ would not hold since there would be a trail $y \rightarrow x \leftarrow w$ and $y \in X, x \in Y, w \in W$ so this trail is blocked. This implies $\pi_x \subset XYZ$ so the second condition also holds for σ_n.

- (extraction) $D(X,Z,WY) \wedge I(X,Z,W) \Rightarrow D(X,ZW,Y)$. Like for extraction assuming x or y in W makes $I(X,Z,W)$ impossible. Therefore, $x \in X$ and $y \in Y$ or $y \in X$ and $x \in Y$ and the first condition is shown. The second condition ($\pi_x \subset WXYZ$) still holds also.

- (intersection) $D(X,Z,WY) \wedge I(X,ZW,Y) \Rightarrow D(X,ZY,W)$. Let $x \in X, y \in WY$ then $y \in Y$ implies $<X,ZW,Y>$ would not hold since $y \rightarrow x$. So $y \in W$ which shows the first condition. The second condition $\pi_x \subset WXYZ$ still holds since σ_n contains all sets W, X, Y and Z

  Let $x \in WY, y \in X$ then $x \in Y$ implies $<X,ZW,Y>$ would not hold since $y \rightarrow x$. So $x \in W$ which shows the first condition. The second condition $\pi_x \subset WXYZ$ still holds since σ_n contains all sets W, X, Y and Z.

So the two conditions still hold for derivations with n clauses. □

For the proof of Theorem 4.1 we use the following lemma.

Lemma A1 *Let $G = (U, A(G))$ be a DAG. Let $x, y \in U$ be two nodes such that $y \rightarrow x$ in G. Let $Z \subseteq U \backslash \{x,y\}$ be a set of nodes such that $\pi_x \subset Z \cup \{y\}$. Let $c \in U \backslash Z$ be a node that is a descendant of node x and has no descendants in Z. Furthermore, let any trail between x and y not containing $y \rightarrow x$ be blocked by Zc. Then, either $<c, Zx, y>$ or $<c, Zy, x>$.*

Proof: Assume two trails $t(x,c)$ and $t(c,y)$ exist such that both $t(x,c)$ is not blocked given Zy and $t(c,y)$ is not blocked given Zx. From the conditions in the lemma then also $t(x,c)$ is not blocked given Z and $t(c,y)$ is not blocked given Z. Since c does not contain descendants in Z both trails must have an incoming arrow into c.

So, it follows that the trail $t(x,y)$ that arises when $t(x,c)$ and $t(c,y)$ are concatenated is not blocked by Zc (the part between x and c is not blocked by Z, c is a head-to-head node and the part between y and c is also not blocked by Z) and the condition would not be fulfilled. Therefore, either all trails $t(x,c)$ are blocked by Zy or all trails $t(y,c)$ are blocked by Zx. □

Theorem 4.1 *In a DAG G that is a minimal I-map of an independency model M_I, the following property holds:*

$$>X,Z,Y< \Rightarrow D(X,Z,Y)$$

Proof: We assume that $>X,Z,Y<$ holds in G. Without loss of generality we take $x \in X$ and $y \in Y$ to be the nodes such $y \rightarrow x$ is an arc in G, $\pi_x \subset XYZ$ and let Q be the set such that $Z \subseteq Q \subseteq XYZ \backslash \{x,y\}$ and $<x,Q,y>$ in G when the arc $y \rightarrow x$ is removed.

Let $Q_u \subseteq Q$ be the set of nodes that are not descendants of x and let $Q_d = Q \backslash Q_u$. Since G is a minimal I-map we have that,

$$\begin{aligned}
& D(x,\pi_x \backslash y, y) \\
\Rightarrow & \ \{ \text{ Composition } \} \\
& D(x,\pi_x \backslash y, yQ_u) \\
\Rightarrow & \ \{ \text{ Intersection with } I(x,\pi_x,Q_u) \ \} \\
& D(x,Q_u\pi_x \backslash y, y)
\end{aligned}$$

Now we will proof that $D(x,Q\pi_x \backslash y, y)$ also holds by adding nodes in Q that are descendants of x one by one using Lemma A1. If all trails between x and y are blocked by $Q_u \backslash y$ then they are also blocked by $Q_u\pi_x \backslash y$. Let $q \in Q_d$ be the node that is lowest in the ordering θ then Lemma A1 implies either $I(q, xQ_u\pi_x \backslash y, y)$ or $I(q, Q_u\pi_x, x)$. So,

$$\begin{aligned}
& D(x,Q_u\pi_x \backslash y, y) \\
\Rightarrow & \ \{ \text{ Intersection with } I(q, xQ_u\pi_x \backslash y, y) \text{ or } I(q,Q_u\pi_x,x) \ \} \\
& D(x,qQ_u\pi_x \backslash y, y)
\end{aligned}$$

Repeat this derivation step with $Q_u := Q_u \cup q$ and $Q_d = Q_d \backslash q$ until $Q_d = \emptyset$. Observe that if all trails between a and b not containing $y \rightarrow x$ are blocked by a set D then they are also blocked by a set $D \backslash p$ where p is a descendant of a that has no descendants in D. Therefore, Lemma A1 applies every time the derivation step is repeated. This derivation results in $D(x,Q\pi_x \backslash y, y)$. So,

$$\begin{aligned}
& D(x,Q\pi_x \backslash y, y) \\
\Rightarrow & \ \{ \text{ Composition and symmetry } \} \\
& D(x \cup (X \backslash Q\pi_x), Q\pi_x \backslash y, y \cup (Y \backslash Q\pi_x)) \\
\Rightarrow & \ \{ \text{ Weak reunion and symmetry } \} \\
& D(X,Z,Y)
\end{aligned}$$

So, $D(X,Z,Y) \in M_{D,\theta}$ and therefore by Theorem 3.2 $D(X,Z,Y) \in M_D$. □

Theorem 4.2 *In a DAG G_θ constructed from a causal input list L_θ of a complete independency model M_I, the following property holds:*

$$>x,Z,y< \Leftrightarrow D(x,Z,y) \in M_{D,\theta}$$

Proof: The $\Rightarrow$ part follows from Theorem 4.1. For the $\Leftarrow$ part we have to show that for all derivations resulting in a dependency statement of the form $D(x, Z, y)$ $>x, Z, y<$ holds.

Assume $D(x, Z, y)$ but not $>x, Z, y<$. By Lemma 4.1 we know that two nodes $a \in \{x\}$, $b \in \{y\}$ or $a \in \{x\}$, $b \in \{y\}$, must exist such that $b \rightarrow a$ and $\pi_a \subset XYZ$. Without loss of generality we assume that $a = x$ and $b = y$. Since no $Q \subseteq XYZ \backslash \{x, y\} = Z$ exists such that all trails between x and y are blocked (otherwise we have $>x, Z, y<$), a trail not containing $y \rightarrow x$ must exist between x and y that is unblocked. Since, $\pi_x \subset Zy$ the trail cannot contain an arc from a parent of x to x. So the trail contains a descendant of x (thus also of y) that forms a head-to-head node. For the trail to be blocked a node $w \in Z \cap desc(x)$ must exists such that w is a head-to-head node or descendant of a head-to-head node in a trail between x and y.

If $D(x, Z, y)$ can be derived from $D(a, \pi_a \backslash b, b)$ using the dependency axioms then $I(a, \pi_a \backslash b, b)$ can be derived from $D(x, Z, y)$ using the independency axioms. Since we know $(\pi_a \backslash b) \subseteq Z$ we have to remove the nodes $Z \backslash (\pi_a \backslash b)$ from Z. How can w be removed from Z in the derivation? This can only be done when nodes Q are introduced such that the trail between x and y via w is blocked using contraction. However, introducing such a node cannot be done since the Q must fulfill $<Q, Zx, y>$ or $<Q, Zy, x>$. These statements will hold for no such Q since the trail between x and y via w is not blocked. So, $D(x, Z, y)$ could not be derived. □

Theorem 4.3: *The closure of Σ_0 under the dependency axioms is equal to the closure of Σ_0 under all dependency axioms but extraction+.*

Proof: We call the dependency axioms $A1$ and define $A2$ as $A1$ without extraction+. So, we have to show that the closure of Σ_0 under $A1$ is equal to the closure of Σ_0 under $A2$. Since $A2 \subset A1$, the closure of Σ_0 under $A2$ is a subset of the closure of Σ_0 under $A1$. Leaves to be shown that the closure of Σ_0 under $A1$ is a subset of the closure of Σ_0 under $A2$.

Assume $D(X, Z, Y)$ is in the closure of Σ_0 under $A1$ but not in the closure of Σ_0 under $A2$. Observe $>X, Z, Y<$ cannot hold since in the proof that $>X, Z, Y< \Rightarrow D(X, Z, Y)$ (Theorem 4.1) we only used the axioms in $A2$. By Lemma 4.1 two nodes x and y must exist such that $x \in X$ and $y \in Y$ or $x \in Y$ and $y \in X$ $y \rightarrow x$ and $\pi_x \subset XYZ$ (Note that Lemma 4.1 also uses extraction+). Since no $Q \subseteq XYZ \backslash \{x, y\}$ exists such that all trails between x and y not containing $y \rightarrow x$ are blocked, a node $w \in Z \cap desc(x)$ must exists such that w is in a trail between x and y.

Observe a derivation $D(a, \pi_a \backslash b, b) \stackrel{*}{\Rightarrow} D(X, Z, Y)$ is isomorph with the derivation $I(X, Z, Y) \stackrel{*}{\Rightarrow} I(a, \pi_a \backslash b, b)$ where the axioms for independency statements are used and both extraction and extraction+ are replaced by contraction. Furthermore, observe that the set $A2$ contains axioms that do not remove nodes from the dependency statements while $A1$ does (via extraction+). Likewise, the independency axioms used in the derivation isomorph with a derivation using $A2$ do not introduce new nodes while the derivation isomorph with $A1$ does.

Assume no derivation using the isomorph axioms in $A2$ can be made, i.e. such that no new nodes are introduced. Then one may ask which nodes are helpful to introduce. The only ones that are helpful are nodes that block a trail like the one containing w. For introducing such a node q $<q, Z', x>$ or $<q, Z', y>$ must hold. However, this will never be the case since the trail not blocked by w is still not blocked. So, such a dependency statement $D(X, Z, Y)$ does not exist. □

References

[1] A.P. Dawid. Conditional independence in statistical theory. *J.R. Statistical Society B*, pages 1–31, 1979.

[2] D. Geiger, T. Verma, and J. Pearl. Identifying independence in bayesian networks. *Networks, Vol. 20*, pages 507–534, 1990.

[3] S.L. Lauritzen, A.P. Dawid, A.P. Larsen, and H.G. Leimer. Independence properties of directed markov fields. *Networks,* Vol 20, pages 491–505, 1990.

[4] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their applications to expert systems (with discussion). *J.R. Stat. Soc. (Series B),*Vol. 50, pages 157–224, 1988.

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

[6] J. Pearl, D. Geiger, and T. Verma. The logic of influence diagrams. In R.M. Oliver and J.Q. Smith, editors, *Influence Diagrams, Belief Nets and Decision Analysis*, pages 67–87. John Wiley & Sons Ltd., 1990.

[7] J. Pearl and A. Paz. Graphoids: a graph based logic for reasoning about relevance relations. In *Proceedings ECAI*, 1986.

[8] J. Pearl and T. Verma. The logic of representing dependencies by directed acyclic graphs. In *Proceedings AAAI*, pages 374–79, 1987.

[9] Ross D. Shachter. Evaluating influence diagrams. *Operations Research*, 34:871–882, 1986.

[10] T. Verma and J. Pearl. Causal networks: Semantics and expressiveness. In *Proceedings of the 4th Workshop on uncertainty in AI*, pages 352–359, 1988.