Bayesian Belief Networks and Conditional Independencies

Remco R. Bouckaert

RUU-CS-92-36
1992

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

Bayesian Belief Networks and Conditional Independencies

Remco R. Bouckaert

Technical Report RUU-CS-92-36
1992

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

ISSN: 0024-3275

Bayesian Belief Networks and Conditional Independencies

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 this paper, the preliminaries for the development of a theory for Bayesian Belief Networks (BBN) construction using conditional independencies are presented. This paper offers the following; an overview of the work done in constructing BBNs using conditional independencies, and a uniform formalization of this work including full proofs of claims made about this work.

1 Introduction

Bayesian belief networks (BBNs) offer a mathematically sound tool for reasoning with uncertainty in knowledge-based systems. By exploring the semantics of BBNs efficient algorithms have been developed for processing evidence and calculating probabilities of hypotheses. The theory of BBNs integrates two disciplines: probability and graph theory. So, the theory may draw from a rich theoretical background.

Compared to the work on reasoning with BBNs [9, 6, 10, 2], not much research is done on the construction of BBNs from statistical data. [9, 8]. The aim of this paper is to make a start with developing a theory for constructing BBNs. To this end, this report offers the following. Firstly, we give an extensive overview of the work already done on constructing BBNs using the independencies of variables in distributions [9, 4, 12]. Secondly, we present a uniform formalization of this work. Thirdly, we give full proofs of the mathematical properties involved; it is noted that proofs in literature are not so rigorous as one could wish for or are not available at all.

In Section 2, the basic notions and notational conventions used in this paper are presented. Amongst other things, we develop the notion of conditional independence to characterize a probability distribution. In Section 3, some properties of conditional independence are presented. In Section 4, various types of graphs and the probability distributions they can represent are presented. In the final section, we compare the different types of BBNs and consider practical aspects of the construction of the different types of graphs.

2 Preliminaries

In this section we present the nomenclature and our notation of this field of research. Since theory on BBNs relies heavily on probability and graph theory, two separate subsections have been addressed to them. The third subsection deals with the graph-theoretical notion of separation. In the last subsection terms are explained that do not fit with the other three subjects.

2.1 Probability theory

Definition 2.1 *A* **probability measure** *Q on a field F over a space Ω is a function $Q : F \to [0,1]$ satisfying*

- *for all $A \in F$, $0 \leq Q(A) \leq 1$*
- *$Q(\emptyset) = 0, Q(\Omega) = 1$*
- *for A_1, A_2 disjoint subsets of Ω $Q(A_1 \cup A_2) = Q(A_1) + Q(A_2)$.*

Definition 2.2 *A* **σ-field** *is a field that is closed under formation of countable unions.*

Definition 2.3 *A* **probability space** *is a triplet (Ω, F, Q) where Ω is an arbitrary space, F is a σ-field over Ω and Q is a probability measure on F.*

Definition 2.4 *A* **random variable** *u in a probability space (Ω, F, Q) is a function $u : \Omega \to \mathbb{R}$ such that*

- *$\forall_{x \in \mathbb{R}} A_x \in F$, and*
- *$Q(A_\infty) = 1$*

where $A_x = \{\omega | \omega \in \Omega, u(\omega) < x\}$. $u_i \in \mathbb{R}$ is a **value** *of a random variable u if $\exists_{\omega \in \Omega} u(\omega) = u_i \wedge Q(\omega) > 0$.*

Definition 2.5 *A* **probability distribution** *P over a random variable u in a probability space (Ω, F, Q) is a function defined by*

$$\forall_{B \subseteq \mathbb{R}} P(u(\omega) \in B) = Q(\{\omega | \omega \in \Omega, u(\omega) \in B\})$$

Likewise a probability distribution P over a set of random variables $U = \{u^1, \ldots, u^n\}$ in a probability space (Ω, F, Q) is defined by $\forall_{B_1,\ldots,B_n \subseteq \mathbb{R}} P(u^1(\omega) \in B_1 \wedge \ldots \wedge u^n(\omega) \in B_n) = Q(\{\omega | \omega \in \Omega, u^1(\omega) \in B_1, \ldots, u^n(\omega) \in B_n\})$. Since we are not interested in the argument ω we omit it and write $P(u \in B)$ instead of $P(u(\omega) \in B)$. In this paper we will use the notation $P(.)$ to denote a probability distribution. A random variable will be called a variable for short. Most of the time we are concerned with discrete variables, i.e. variables with only a finite number of values in the range with probability not equal to zero.

Definition 2.6 *A probability distribution P over a set of variables is called* **positive definite** *if for all combinations of values of the variables the probability is not equal to zero.*

We will adhere to the following conventions. We will use lower-case letters for single variables and upper-case letters for sets of variables. For single variables we use superscripts for distinction and subscript indices to denote the values of a variable. So, we write $P(a^1 = a^1_1)$ for the probability that variable a^1 will have the value a^1_1. We will write $P(a_i)$ for $P(a = a_i)$ and $P(a_i b_j)$ for the probability that $a = a_i$ and $b = b_j$. Any expression or derivation E that contains variable names or variable set names without having values assigned as arguments of a probability measure we mean E is true for any combination of values these variables can get. For example, let $a \in \{a_1, \ldots, a_k\}$ and $b \in \{b_1, \ldots, b_m\}$. We take $P(ab) = P(a)P(b)$ to mean $\forall_{i \in \{1,\ldots,k\} j \in \{1,\ldots,m\}} P(a_i b_j) = P(a_i)P(b_j)$.

In this paper we will use the letter U to denote the set of all variables concerned in a given context.

Definition 2.7 *Let $X, Y \subseteq U$ be sets of variables, such that $P(X) > 0$. Then, the* **conditional probability** *of Y given X is defined as*

$$P(Y|X) = \frac{P(XY)}{P(X)}$$

Definition 2.8 *Let $X, Y, Z \subseteq U$ be sets of variables. X is said to be* **conditionally independent** *of Y given Z, written $I(X, Z, Y)$, if*

$$P(XY|Z) = P(X|Z)P(Y|Z)$$

A statement $I(X, Z, Y)$ is called an **independency statement.**

Note that $I(X, \emptyset, Y)$ indicates unconditional independence, i.e. $P(XY) = P(X)P(Y)$.

Definition 2.9 *Let P be a probability distribution over a set of variables U. The* **independency model** *M of P is the set of all independency statements that hold in P.*

2.2 Graphs

We want to represent an independency model by a graph. To do so, several kinds of graphs are considered.

Definition 2.10 *An* **undirected graph (UG)** *G is an ordered pair $G = (V(G), E(G))$ where $V(G) = \{v^1, v^2, \ldots, v^n\}, n \geq 1$, is a finite set of nodes and $E(G)$ is a family of unordered pairs $(v^i, v^j) \in U(G)$ called edges.*

Definition 2.11 *Let $G = (V(G), E(G))$ be an undirected graph. A* **path** *$p(v^i, v^j)$ from node $v^i \in V(G)$ to node $v^j \in V(G)$ is an ordered set $\{p^0, p^1, \ldots, p^n\} \subseteq V(G)$ of nodes such that $p^0 = v^i$, $p^n = v^j$ and $\forall_{0 \leq k < n}(p^k, p^{k+1}) \in E(G)$. The* **length** *of a path $p(v^i, v^j)$ is the cardinality of $p(v^i, v^j)$ minus 1. A* **cycle** *is a path $p(v^i, v^i)$.*

Definition 2.12 *A* **chordal graph (CG)** *is a UG $G = (V(G), E(G))$ in which for every cycle $p(v^i, v^i) = \{v^i, v^j, \ldots, v^k, v^i\}$ ($v_. \in V(G)$) of length equal or greater than four, there exists a cycle $p(v^i, v^i) = \{v^i, v^j, v^k, v^i\}$ in G.*

Definition 2.13 *A* **multi undirected graph (MUG)** *is an ordered pair $MG = (V(MG), S(MG))$ where $V(MG) = \{v^1, v^2, \ldots, v^n\}$, $n \geq 1$, is a finite set of nodes and $S(MG)$ is a set of undirected graphs $G = (V(G), E(G))$ such that for every $G \in S(MG)$ $V(G) \subseteq V(MG)$.*

Definition 2.14 *A* **directed graph** *G is an ordered pair* $G = (V(G), A(G))$ *where* $V(G) = \{v^1, v^2, \ldots, v^n\}, n \geq 1$, *is a finite set of nodes and* $A(G)$ *is a family of ordered pairs* (v^i, v^j), $v^i, v^j \in V(G)$, *called arcs.*

We take the notion of a path, the length of a path and a cycle to apply to directed graphs by taking the direction of the arcs into consideration.

Definition 2.15 *Let* $G = (V(G), A(G))$ *be a directed graph. A* **trail** $t(v^i, v^j)$ *in G from node* $v^i \in V(G)$ *to node* $v^j \in V(G)$ *is an ordered set* $\{p^0, p^1, \ldots, p^n\} \subseteq V(G)$ *of nodes so that* $p^0 = v^i$, $p^n = v^j$ *and* $\forall_{0 \leq k < n} (p^k, p^{k+1}) \in A(G) \vee (p^{k+1}, p^k) \in A(G)$.

Definition 2.16 *A* **directed acyclic graph (DAG)** *is a directed graph without any cycles.*

Definition 2.17 *In a DAG* $G = (V(G), A(G))$ *a* **descendant** *of a node* $v \in V(G)$ *is a node* $w \in V(G)$ *such that there exists a path* $p(v, w)$ *in G.*

Definition 2.18 *A* **multi directed acyclic graph (MDAG)** *is an ordered pair* $MG = (V(MG), S(MG))$ *where* $V(MG) = \{v^1, v^2, \ldots, v^n\}$, $n \geq 1$, *is a finite set of nodes and* $S(MG)$ *is a set of directed acyclic graphs* $G = (V(G), E(G))$ *such that for every* $G \in S(MG)$, $V(G) \subseteq V(MG)$.

2.3 Separation

A graph represents an independency model of a probability distribution over U if every node in the graph represents a variable and vice versa. In addition, there is a criterion with which we can read the independency statements in the model from the graph [9, 5]. This criterion is called **separation** and for the various kinds of graphs there are various kinds of separation. First we will define separation for undirected graphs.

Definition 2.19 *Let* $G = (V(G), E(G))$ *be an undirected graph. Let* $X, Y, Z \subseteq V(G)$ *be sets of nodes. We say that Z* **separates** *X from Y, written* $<X, Z, Y>$, *if every path from any* $x \in X$ *to any* $y \in Y$ *contains at least one* $z \in Z$.

Separation in a UG is also called **vertex separation.** In the sequel, we write $<x, z, y>$ instead of $<\{x\}, \{z\}, \{y\}>$ for single nodes x, y and z and XY instead of $X \cup Y$. We write $X - Y$ or $X \backslash Y$ for the set X with those nodes that are in Y excluded.

Definition 2.20 *Let* $G = (V(G), E(G))$ *be a UG. Let* $a, b, c \in V(G)$ *be nodes. Then, a, b are* **adjacent in** *G, written* $a - b$, *if a and b cannot be separated by any subset in* $V(G) \backslash \{a, b\}$; *a, b are* **nonadjacent in** *G, written* $a \not- b$, *if a and b are not adjacent.*

The nodes a, b are **conditionally adjacent in** *G given c, written* $a - b | c$, *if a and b cannot be separated by any subset* $S \subseteq V(G) \backslash \{a, b\}$ *such that* $c \in S$; *a, b are* **conditionally nonadjacent in** *G given c, written* $a \not- b | c$, *if a and b are not conditionally adjacent given c.*

We take the notions of adjacency, nonadjacency, conditional adjacency and conditional nonadjacency to apply to DAGs, MUGs and MDAGs by inserting the proper type of graph in the definition.

Definition 2.21 *Let $MG = (V(MG), S(MG))$ be a multi undirected graph. Let $X, Y, Z \subseteq V(MG)$ be sets of nodes. We say that Z* **separates** *X from Y in MG if there exists a UG $G = (V(G), E(G)) \in S(MG)$ such that*

- *Z separates X from Y in G.*
- $\forall_{a \in X \cup Y \cup Z} \; a \in V(G)$.

For DAGs the separation criterion is more sophisticated and we need a couple of new notions to define it.

When a directed graph G contains an arc (v^i, v^j) we say v^j has an **incoming arrow** from v^i, written $v^i \rightarrow v^j$. If a trail $t(v^i, v^j)$ in a directed graph G contains three consecutive nodes v^{k-1}, v^k and v^{k+1} for which $v^{k-1} \rightarrow v^k$ and $v^{k+1} \rightarrow v^k$ we say the trail contains a **head-to-head node**, written $\rightarrow v^k \leftarrow$.

Definition 2.22 *Let $G = (V(G), A(G))$ be a DAG. Let $x, y \in V(G)$ be nodes and $Z \in V(G)$ be a set of nodes. A trail $t(x, y)$ in G is* **blocked** *by Z, written $x \not\leftrightsquigarrow y|Z$, if at least one of the following conditions hold:*

- *the trail contains a head-to-head node e such that $e \notin Z$ and none of the descendants of e is in Z.*
- *there is a node e in the trail with $e \in Z$ that is not a head-to-head node in the trail.*

A trail $t(x, y)$ in G is **active** given Z, written $x \leftrightsquigarrow y|Z$, if the trail is not blocked by Z. Sometimes there are head-to-head nodes in a DAG where the corresponding variable has a fixed value, the so-called **fixed value nodes**. When we want to determine whether a trail is blocked or not by a set Z we first have to add the fixed value nodes to Z and than apply the definition.

Now we are able to define separation for DAGs.

Definition 2.23 *Let $G = (V(G), A(G))$ be a DAG. Let $X, Y, Z \subseteq V(G)$ be sets of nodes. We say that Z* **separates** *X from Y if every trail between any $x \in X$ to any $y \in Y$ is blocked by Z.*

Separation in DAGs is also called **d-separation**.

Definition 2.24 *Let $MG = (V(MG), S(MG))$ be a multi directed acyclic graph. Let $X, Y, Z \subseteq V(MG)$ be sets of nodes. We say that Z* **separates** *X from Y in MG if there exists a DAG $G = (V(G), A(G)) \in S(MG)$ for which*

- *Z d-separates X from Y in G.*
- $\forall_{a \in X \cup Y \cup Z} \; a \in V(G)$.

2.4 Auxiliary

To represent an independency model with a graph, independency statements and separation have to be related. With every variable $u^i \in U$ in the probability distribution a node $u^i \in V(G)$ in the graph is associated. Similarly, we associate with every set of variables $X \subseteq U$ a set of nodes $X \subseteq V(G)$. From now on we will not make a distinction between U and $V(G)$.

Definition 2.25 *Let $G = (V(G), V(G))$ be a UG. Let M be an independency model over a set of variables $V(G)$.*

- *G is called an* **independency map** *or* **I-map** *of M iff for all $X, Y, Z \subseteq V(G)$ we have $<X, Z, Y> \Rightarrow I(X, Z, Y)$.*
- *G is called a* **dependency map** *or* **D-map** *of M iff for all $X, Y, Z \subseteq V(G)$ we have $I(X, Z, Y) \Rightarrow <X, Z, Y>$.*
- *G is called a* **perfect map** *of M iff for all X, Y and Z we have $<X, Z, Y> \Leftrightarrow I(X, Z, Y)$.*

A fully connected UG is a trivial I-map and an edgeless graph is a trivial D-map. If conditional independence implies separation and vice versa then the graph is called a perfect map. Unfortunately, not every independency model has a perfect map that is a UG or a DAG.

We take the notions of independency map, dependency map and perfect map to apply to DAGs, MUGs and MDAGs by inserting the proper type of graph in the definition.

Definition 2.26 *An independency model M is called* **graph-isomorph** *if there exists a graph G that is a perfect map of M.*

We speak about UG-isomorphism if there exists a UG that is a perfect map of the model and similar of DAG-isomorphism, MUG-isomorphism and MDAG-isomorphism.

3 Some properties of conditional independence

In this section we discuss some properties of the conditional independence relation which can be found in [1, 7, 9]. Throughout this section P is a probability distribution over a set of variables U and I is the independency relation in P.

Lemma 3.1 *Let $X, Y, Z \subseteq U$ be sets of variables, then*

$$I(X,Z,Y) \Leftrightarrow P(XY|Z) = P(X|Z)P(Y|Z) \quad (1)$$

$$\Leftrightarrow P(XYZ) = P(X|Z)P(YZ) \quad (2)$$

$$\Leftrightarrow P(X|ZY) = P(X|Z) \quad (3)$$

$$\Leftrightarrow \exists_{f,g} P(XYZ) = f(X,Z)g(Y,Z) \quad (4)$$

Proof: In the lemma, (1) is by definition; (2) follows from $P(XZY) = P(XY|Z)P(Z) = P(X|Z)P(Y|Z)P(Z) = P(X|Z)P(YZ)$. Equation (3) follows from $P(X|ZY) = P(XYZ)/P(YZ) = P(X|Z)P(YZ)/P(YZ) = P(X|Z)$. Equation (4) follows from the axioms of probability theory and is left to the reader. □

Note that $I(X, Z, Y)$ implies $P(Z) > 0$ by definition of conditional independence.

Definition 3.1 *Let $X, Y, Z, W \subseteq U$ be sets of variables and $a, b, c, d \in U$ be variables. Then we define the following properties:*

reflexivity	$X \subseteq Z$	$\Rightarrow$	$I(X,Z,Y)$
relative disjunction	$I(X,Z,Y)$	$\Rightarrow$	$X \cap Y \subseteq Z$
symmetry	$I(X,Z,Y)$	$\Leftrightarrow$	$I(Y,Z,X)$
decomposition	$I(X,Z,WY)$	$\Rightarrow$	$I(X,Z,W) \wedge I(X,Z,Y)$
weak union	$I(X,Z,WY)$	$\Rightarrow$	$I(X,ZW,Y)$
contraction	$I(X,Z,Y) \wedge I(X,ZY,W)$	$\Rightarrow$	$I(X,Z,WY)$
intersection	$I(X,ZW,Y) \wedge I(X,ZY,W)$	$\Rightarrow$	$I(X,Z,YW)$
strong union	$I(X,Z,Y)$	$\Rightarrow$	$I(X,ZW,Y)$
transitivity	$I(X,Z,Y)$	$\Rightarrow$	$I(X,Z,c) \vee I(c,Z,Y)$
composition	$I(X,Z,W) \wedge I(X,Z,Y)$	$\Rightarrow$	$I(X,Z,WY)$
weak transitivity	$I(X,Z,Y) \wedge I(X,Z \cup \{c\},Y)$	$\Rightarrow$	$I(X,Z,c) \vee I(c,Z,Y)$
chordality	$I(a,cd,b) \wedge I(c,ab,d)$	$\Rightarrow$	$I(a,c,b) \vee I(a,d,b)$
d − transitivity			

$$I(X,Z,Y) \Leftrightarrow \forall_{x \in X, y \in Y \text{ or } y \in X, x \in Y}$$

$$\{x \in Z \vee y \in Z\}$$

$$or$$

$$\left\{ \begin{array}{c} \{x \neq y\} \\ and \\ \left\{ \forall_{a,b \in U} I(x,Zy,a) \vee I(b,Zx,y) \vee a - b \vee \exists_{c \in Z} \; a \not- b|c \vee Z = \emptyset \right\} \\ and \\ \left\{ \forall_{a \in U} \; I(a,Zx,y) \vee x \not- a \vee (\exists_{w \in U} \; x \not- w \wedge x - w|a) \right\} \end{array} \right\}$$

In table 1 a list of properties for conditional independency statements is given together with the distributions for which they hold. Most other known rules of inference, such as the chaining rule ($I(X,Y,Z) \wedge I(XY,Z,W) \Rightarrow I(X,Y,W)$ [9, 1]) and mixing rule ($I(X,Z,WY) \wedge I(Y,Z,W) \Rightarrow I(XW,Z,Y)$ [9]), can be derived from a subset of these statements and Bayes rule.

In the last column 'any' means that the axiom holds for any probability distribution. The term 'positive definite' refers to probability distributions that are positive definite. The '−−' signs mean that the corresponding axiom does not hold for any distribution in general; this does not mean, however, that there are no distributions satisfying these axioms.

We begin by proving the validity of the properties reflexivity, relative disjunction, symmetry, decomposition, weak union and contraction for any probability distribution.

Lemma 3.2 *Reflexivity holds for any probability distribution.*

name	*distribution*
reflexivity	any
relative disjunction	positive definite
symmetry	any
decomposition	any
weak union	any
contraction	any
intersection	positive definite
strong union	--
transitivity	--
composition	--
weak transitivity	normal
chordality	--
d − transitivity	--

Table 1: List of axioms for conditional independence.

Proof: Let $X, Y, Z \subset U$ be sets of variables. If $X \subseteq Z$ either the values of the variables in X are the same as those of Z and $P(XY|Z) = P(Y)$, or there is a variable with a different value and $P(XY|Z) = 0$. In both cases $P(XY|Z) = P(X|Z)P(Y|Z)$ thus $I(X, Z, Y)$ □

Lemma 3.3 *Relative disjunction holds for positive definite probability distribution.*

Proof: Let $X, Y, Z \subseteq U$ be sets of variables. Suppose $I(X, Z, Y)$ and $X \cap Y = Z' \not\subseteq Z$ and let $X' = X \backslash Z'$ and $Y' = Y \backslash Z'$. Then $I(X, Z, Y) = I(X'Z', Z, Y'Z')$ implies $I(Z', Z, Z')$ using the decomposition axiom (which is shown to hold for any probability distribution in lemma 3.5). So $P(Z'|Z) = P(Z'Z'|Z) = P(Z'|Z)P(Z'|Z)$ which implies $P(Z'|Z) = 0$ or 1. This is only true if $Z' \subset Z$ (or if Z' is determined by Z). □

Lemma 3.4 *Symmetry holds for any probability distribution.*

Proof: Let $X, Y, Z \subseteq U$ be sets of variables. We have to show $I(X, Z, Y) \Leftrightarrow I(Y, Z, X)$. From $I(X, Z, Y)$ we have

$$\begin{aligned} P(XY|Z) &= P(X|Z)P(Y|Z) && \{\text{Def. } I(X,Z,Y)\} \\ &= P(Y|Z)P(X|Z) && \{\text{comm. of mult.}\} \\ &= P(YX|Z) && \{\text{Def.}\} \end{aligned}$$

And therefore, $I(Y, Z, X)$. Since the proof consists of equalities only, we also have $I(Y, Z, X)$ implies $I(Y, Z, X)$. □

Lemma 3.5 *Decomposition holds for any probability distribution.*

Proof: Let $W, X, Y, Z \subseteq U$ be sets of variables. We have to show $I(X, Z, WY) \Rightarrow I(X, Z, W) \wedge I(X, Z, Y)$. From $I(X, Z, WY)$ we find

$$\begin{aligned} P(XY|Z) &= \textstyle\sum_W P(XWY|Z) \\ &= \textstyle\sum_W P(X|Z)P(WY|Z) && \{\text{Def. } I(X,Z,WY)\} \\ &= P(X|Z)P(Y|Z) \end{aligned}$$

From which we have $I(X, Z, Y)$. The proof for $I(X, Z, WY) \Rightarrow I(X, Z, W)$ is analogous. □

Lemma 3.6 *Weak union holds for any probability distribution.*

Proof: Let $W, X, Y, Z \subseteq U$ be sets of variables. We have to show $I(X, Z, WY) \Rightarrow I(X, ZW, Y)$. Using $I(X, Z, WY)$ we find

$$\begin{aligned} P(XY|ZW) &= \frac{P(XYZW)}{P(ZW)} \\ &= \frac{P(XZ)}{P(Z)}\frac{P(YZW)}{P(ZW)} & \{\text{def. } I(X,Z,WY)\} \\ &= P(X|Z)P(Y|WZ) \\ &= P(X|WZ)P(Y|WZ) \end{aligned}$$

From which we have $I(X, ZW, Y)$. The last step follows from $I(X, Z, WY) \Rightarrow I(X, Z, W)$ by decomposition and from $P(X|ZW) = P(X|Z)$ since (3). □

Lemma 3.7 *Contraction holds for any probability distribution.*

Proof: Let $W, X, Y, Z \subseteq U$ be sets of variables. We have to show $I(X, Z, Y) \wedge I(X, ZY, W) \Rightarrow I(X, Z, WY)$. Using $I(X, Z, Y)$ and $I(X, ZY, W)$ we find

$$\begin{aligned} P(WXY|Z) &= P(WX|YZ)P(Y) \\ &= P(X|YZ)P(W|YZ)P(Y) \\ &= P(X|Z)P(WY|Z) \end{aligned}$$

From which we have $I(X, Z, WY)$. □

Now we prove the validity of intersection for positive definite distributions and the invalidity of weak transitivity for binary variables.

Lemma 3.8 *Intersection holds for positive definite probability distributions.*

Proof: Let $W, X, Y, Z \subseteq U$ be sets of variables. We have to show $I(X, ZW, Y) \wedge I(X, ZY, W) \Rightarrow I(X, Z, YW)$. Using $I(X, ZW, Y)$ and $I(X, ZY, W)$ we find

$$\begin{aligned} P(WXYZ) &= P(X|WZ)P(WYZ) \\ &= P(X|YZ)P(WYZ) \end{aligned}$$

Therefore, $P(X|WZ) = P(X|YZ)$ for every value of Y and W under the condition $P(WYZ) > 0$. But, $P(X|WZ) = P(X|YZ)$ implies $P(X|WZ)$ is constant, i.e. not depending on the value of W. In other words $I(X, W, Z)$. Now, $I(X, ZW, Y)$ and $I(X, W, Z)$ imply $I(X, Z, WY)$ by contraction. □

Intersection does not hold for probability distributions that are not positive definite. Consider the following counterexample: a represents 'the water boils', b represents the temperature of the water in degrees Celsius and c stands for the temperature of the water in degrees Fahrenheit. Observe that b and c are determined by each other. Now we know that the water boils if we know the temperature in degrees Celsius without needing to know the temperature in Fahrenheit and vica versa. So, $I(a, b, c) \wedge I(a, c, b)$ but not $I(a, \emptyset, bc)$.

Contrary to suggestions in [9] (pp. 130) we show that weak transitivity does not hold for binary variables.

Consider the following example. Let $U = \{p, q, r, s\}$ with $p \in \{p_0, p_1\}$, $q \in \{q_0, q_1\}$, $r \in \{r_0, r_1\}$ and $s \in \{s_0, s_1\}$. Let

$$P(pqrs) = \begin{cases} P(p)P(q|p)P(r|q)P(s|qr) \text{ if } q = q_0 \\ P(p)P(q|p)P(r|q)P(s|qp) \text{ if } q = q_1 \end{cases}$$

then it can be verified that $I(p, q, r)$ and $I(p, qs, r)$ hold since $P(pqrs)$ can be written as $f(pqs)g(rqs)$ and $P(pqr)$ can be written as $f(pq)g(qr)$. But, neither $I(p, q, s)$ nor $I(s, q, r)$ hold (assuming properly chosen parameters). For example let
$P(p = p_1) = \frac{1}{2}$
$P(q = q_1|p = p_0) = \frac{3}{4}$ and $P(q = q_1|p = p_1) = \frac{1}{4}$
$P(r = r_1|q = q_0) = \frac{2}{3}$ and $P(r = r_1|q = q_1) = \frac{1}{3}$
$P(s = s_1|q = q_0 r = r_0) = \frac{1}{2}$ and $P(s = s_1|q = q_0 r = r_1) = \frac{3}{4}$
$P(s = s_1|q = q_1 p = p_0) = \frac{1}{2}$ and $P(s = s_1|q = q_1 p = p_1) = \frac{3}{4}$

$pqrs$	$P(pqsr).96$	$P(pr\|qs)$	$P(p\|qs)P(r\|qs)$	$P(pr\|q)$	$P(p\|q)P(r\|q)$
$p_0q_0r_0s_0$	2	$\frac{1}{8}$	$\frac{1}{4}\cdot\frac{1}{2}$	$\frac{1}{12}$	$\frac{1}{4}\cdot\frac{1}{3}$
$p_0q_0r_0s_1$	2	$\frac{1}{16}$	$\frac{1}{4}\cdot\frac{1}{4}$		
$p_0q_0r_1s_0$	2	$\frac{1}{8}$	$\frac{1}{4}\cdot\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{4}\cdot\frac{2}{3}$
$p_0q_0r_1s_1$	6	$\frac{3}{16}$	$\frac{1}{4}\cdot\frac{3}{4}$		
$p_0q_1r_0s_0$	12	$\frac{4}{7}$	$\frac{6}{7}\cdot\frac{2}{3}$	$\frac{1}{2}$	$\frac{3}{4}\cdot\frac{2}{3}$
$p_0q_1r_0s_1$	12	$\frac{4}{9}$	$\frac{2}{3}\cdot\frac{2}{3}$		
$p_0q_1r_1s_0$	6	$\frac{2}{7}$	$\frac{6}{7}\cdot\frac{1}{3}$	$\frac{1}{4}$	$\frac{3}{4}\cdot\frac{1}{3}$
$p_0q_1r_1s_1$	6	$\frac{2}{9}$	$\frac{2}{3}\cdot\frac{1}{3}$		
$p_1q_0r_0s_0$	6	$\frac{3}{8}$	$\frac{3}{4}\cdot\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{4}\cdot\frac{1}{3}$
$p_1q_0r_0s_1$	6	$\frac{3}{16}$	$\frac{3}{4}\cdot\frac{1}{4}$		
$p_1q_0r_1s_0$	6	$\frac{3}{8}$	$\frac{3}{4}\cdot\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{4}\cdot\frac{2}{3}$
$p_1q_0r_1s_1$	18	$\frac{9}{16}$	$\frac{3}{4}\cdot\frac{3}{4}$		
$p_1q_1r_0s_0$	2	$\frac{2}{21}$	$\frac{1}{7}\cdot\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{4}\cdot\frac{2}{3}$
$p_1q_1r_0s_1$	6	$\frac{2}{9}$	$\frac{1}{3}\cdot\frac{2}{3}$		
$p_1q_1r_1s_0$	1	$\frac{1}{21}$	$\frac{1}{7}\cdot\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{4}\cdot\frac{1}{3}$
$p_1q_1r_1s_1$	3	$\frac{1}{9}$	$\frac{1}{3}\cdot\frac{1}{3}$		

Table 2: Relevant probabilities in weak transitivity case

Then table 2 shows that $P(pr|qs) = P(p|qs)P(r|qs)$ and $P(pr|q) = P(p|q)P(r|q)$. So, $I(p, qs, r)$ and $I(p, q, r)$ hold for this distribution. However, $P(s = s_1|p = p_1 \wedge q = q_1) = \frac{3}{4}$ and $P(s = s_1|q = q_1) = \frac{9}{16}$. Therefore, $P(s|pq) \neq P(s|q)$. So, $I(p, q, s)$ does not hold. Furthermore, $P(s = s_1|q = q_0 \wedge r = r_1) = \frac{3}{4}$ while $P(s = s_1|q = q_0) = \frac{2}{3}$. So, $I(s, q, r)$ does not hold either.

In the definition of d-transitivity the term $Z = \emptyset$ is added to the original definition in [11] because this axiom is supposed hold in independency models that are DAG-isomorph. When $U = \{a, b, c\}$ is a set of variables and $M = \{(a, \emptyset, b), (b, \emptyset, a)\}$ then the graph $(a) \rightarrow (c) \leftarrow (b)$ is a perfect map of M but it would not satisfy the original definition of d-transitivity.

4 Conditions for graph-isomorphism

In this section we show that an independency model is UG-isomorph if and only if it is closed under intersection, strong union, transitivity and the axioms that are valid for any independency model. Further we present a set of axioms that are necessary for an independency model to be DAG-isomorph. Next we show that a independency model is DAG-isomorph if all independency statements in the model can be derived from a causal input list and the model is closed under the axioms that are valid for any independency model. Last we show that any independency model is MUG-isomorph and also MDAG-isomorph.

4.1 UG-isomorphism

In this section, we will start with a description of some properties of vertex separation in undirected graphs. Then, we will investigate the constraints on an independency model to be UG-isomorph.

4.1.1 Properties of vertex separation

Let $W, X, Y, Z \subseteq U$ be sets of nodes and $c \in U$ a single node in a UG $G = (U, E(G))$. In this section we will show that the next properties hold for separation in undirected graphs.

$$
\begin{array}{rcll}
<X,Z,Y> & \Leftrightarrow & <Y,Z,X> & \{\text{Symmetry}\} \\
<X,Z,YW> & \Rightarrow & <X,Z,Y> & \{\text{Decomposition}\} \\
<X,Z,Y> \wedge <X,ZY,W> & \Rightarrow & <X,Z,WY> & \{\text{Contraction}\} \\
<X,ZW,Y> \wedge <X,ZY,W> & \Rightarrow & <X,Z,WY> & \{\text{Intersection}\} \\
<X,Z,Y> & \Rightarrow & <Y,ZW,X> & \{\text{Strong Union}\} \\
<X,Z,Y> & \Rightarrow & <X,Z,c> \vee <c,Z,Y> & \{\text{Transitivity}\}
\end{array}
$$

Lemma 4.1 *Separation in undirected graphs satisfies symmetry.*

Proof: Let $G = (U, E(G))$ be a UG. Let $X, Y, Z \subseteq U$ be sets of nodes. Then,

$$
\begin{array}{rl}
<X,Z,Y> & \Leftrightarrow \displaystyle\forall_{x\in X, y\in Y} \exists_{z\in Z}\; z \in p(x,y) \\
& \Leftrightarrow \displaystyle\forall_{x\in X, y\in Y} \exists_{z\in Z}\; z \in p(y,x) \\
& \Leftrightarrow <Y,Z,X>
\end{array}
$$

□

Lemma 4.2 *Separation in undirected graphs satisfies decomposition.*

Proof: Let $G = (U, E(G))$ be a UG. Let $W, X, Y, Z \subseteq U$ be sets of nodes. Then,

$$
\begin{array}{rl}
<X,Z,WY> & \Leftrightarrow \displaystyle\forall_{x\in X, y\in WY} \exists_{z\in Z}\; z \in p(x,y) \\
& \Rightarrow \displaystyle\forall_{x\in X, y\in Y} \exists_{z\in Z}\; z \in p(x,y) \\
& \Leftrightarrow <X,Z,Y>
\end{array}
$$

□

Lemma 4.3 *Separation in undirected graphs satisfies contraction.*

Proof: Let $G = (U, E(G))$ be a UG. Let $W, X, Y, Z \subseteq U$ be sets of nodes. Then,

$$
\begin{aligned}
<X,Z,Y> \wedge <X,ZY,W> \quad &\Rightarrow \quad \forall_{x\in X, y\in Y} \exists_{z\in Z}\, z \in p(x,y) \\
& \qquad \wedge \forall_{x\in X, w\in W} \exists_{z\in ZY}\, z \in p(x,w) \\
& \Rightarrow \quad \forall_{x\in X, y\in Y} \exists_{z\in Z}\, z \in p(x,y) \\
& \qquad \wedge \forall_{x\in X, w\in W} \exists_{z\in Z}\, z \in p(x,w) \\
& \Leftrightarrow \quad \forall_{x\in X, y\in WY} \exists_{z\in Z}\, z \in p(x,y) \\
& \Rightarrow \quad <X,Z,WY>
\end{aligned}
$$

The second step is valid since if $\exists_{y\in Y}\, y \in p(x,w)$ then $\exists_{z\in Z}\, z \in p(x,w)$ since $\exists_{z\in Z}\, z \in p(x,y)$.

□

Lemma 4.4 *Separation in undirected graphs satisfies intersection.*

Proof: Let $G = (U, E(G))$ be a UG. Let $W, X, Y, Z \subseteq U$ be sets of nodes. Then,

$$
\begin{aligned}
<X,ZW,Y> \wedge <X,ZY,W> \quad &\Leftrightarrow \quad \forall_{x\in X, y\in Y} \exists_{z\in WZ}\, z \in p(x,y) \\
& \qquad \wedge \forall_{x\in X, w\in W} \exists_{z\in YZ}\, z \in p(x,w) \\
& \Rightarrow \quad \forall_{x\in X, y\in Y} \exists_{z\in Z}\, z \in p(x,y) \\
& \qquad \wedge \forall_{w\in W} \exists_{z\in Z}\, z \in p(x,w) \\
& \Leftrightarrow \quad \forall_{x\in X, y\in WY} \exists_{z\in Z}\, z \in p(x,y) \\
& \Leftrightarrow \quad <X,Z,WY>
\end{aligned}
$$

The second step is valid because of the following. Consider a path from a node $x \in X$ to a node $y \in Y$. From $< X, ZW, Y >$ we have $\exists_{w\in W}\, w \in p(x,y) \vee \exists z \in Z z \in p(x,y)$. If $\exists_{w\in W}\, w \in p(x,y)$, let w be the node in W nearest to x. From $< X, ZY, W >$ we have $\exists_{z\in Z}\, z \in p(x,w) \vee \exists_{y\in Y}\, y \in p(x,w)$. The first term must be true since if only the last term would be true then $< X, ZW, Y >$ would not hold, for there would be a path from a node $x \in X$ to a node $y' \in Y$ not containing nodes from the set WZ. Therefore, $\exists_{z\in Z}\, z \in p(x,y)$. By the same line of reasoning $\exists_{z\in Z}\, z \in p(x,w)$ which establishes the second last step. □

Lemma 4.5 *Separation in undirected graphs satisfies strong union.*

Proof: Let $G = (U, E(G))$ be a UG. Let $W, X, Y, Z \subseteq U$ be sets of nodes. Then,

$$\begin{aligned}
<X,Z,Y> &\Leftrightarrow \forall_{x \in X, y \in Y} \ \exists_{z \in Z} \ z \in p(x,y) \\
&\Rightarrow \forall_{x \in X, y \in Y} \ \exists_{z \in Z} \ z \in p(x,y) \vee \exists_{w \in W} \ w \in p(x,y) \\
&\Leftrightarrow <X, ZW, Y>
\end{aligned}$$

□

Lemma 4.6 *Separation in undirected graphs satisfies transitivity.*

Proof: Let $G = (U, E(G))$ be a UG. Let $X, Y, Z \subseteq U$ be sets of nodes and $c \in U \backslash XYZ$ a single node. Now assume that $<X, Z, Y>$, $<X, Zc, Y>$ and $\neg <X, Z, c>$ and $\neg <c, Z, Y>$ all hold in G then,

$$\begin{aligned}
\neg <X,Z,c> \wedge \neg <c,Z,Y> &\Rightarrow \exists_{x \in X} \ \forall_{z \in Z} \ z \notin p(x,c) \wedge \exists_{y \in Y} \ \forall_{z \in Z} \ z \notin p(c,y) \\
&\Rightarrow \exists_{x \in X, y \in Y} \ \forall_{z \in Z} \ z \notin p(x,y) \\
&\Rightarrow \neg <X,Z,Y>
\end{aligned}$$

which is false. So weak transitivity holds by contradiction. □

4.1.2 Conditions for UG-isomorphism

An independency model M can be represented by a UG if the following property holds: $I(X, Z, Y) \in M$ iff $<X, Z, Y>$. Since separation for UGs satisfies symmetry, weak union, decomposition and contraction, the corresponding model will have to be closed under these axioms in order to be UG-isomorph. Recall that for any probability distribution symmetry, weak union, decomposition and contraction hold. So any independency model is closed under these axioms.

Lemma 4.7 *Let M be an independency model closed under intersection. Let $G = (U, E(G))$ be the undirected graph obtained from the complete graph by deleting edges (a, b) such that $I(a, U\backslash\{a, b\}, b) \in M$ Then G is an I-map of M.*

Proof: The proof is based on the proof in [9]. Let $X, Y, Z \subseteq U$ be disjoint subsets.

$$\begin{aligned}
<X,Z,Y> &\Rightarrow \forall_{x \in X, y \in Y} \ <x, Z, y> && \{\text{Decomposition for sep.}\} \\
&\Rightarrow \forall_{x \in X, y \in Y} \ <x, U\backslash\{x,y\}, y> && \{\text{Strong Union for sep.}\} \\
&\Rightarrow \forall_{x \in X, y \in Y} \ I(x, U\backslash\{x,y\}, y) && \{\text{By method of construction}\} \\
&\Rightarrow I(X, U\backslash XY, Y) && \{\text{Intersection for } M\}
\end{aligned}$$

Let $R = U\backslash XZY$ then, $I(X, U\backslash XY, Y) = I(X, ZR, Y)$. Furthermore,

$$
\begin{array}{lll}
<X,Z,Y> & & \\
\Rightarrow & \forall_{x\in X, y\in Y} <x,Z,y> & \{\text{Decomposition for sep.}\} \\
\Rightarrow & \forall_{x\in X, y\in Y, r\in R} <x,Z,r> \vee <r,Z,y> & \{\text{Transitivity for sep.}\} \\
\Rightarrow & \forall_{x\in X, y\in Y, r\in R} <x,ZXYR\backslash\{x,r\},r> \vee <r,ZXYR\backslash\{y,r\},y> & \{\text{Strong Union for sep.}\} \\
\Rightarrow & \forall_{x\in X, y\in Y, r\in R} I(x,ZXYR\backslash\{x,r\},r) \vee I(r,ZXYR\backslash\{y,r\},y) & \\
 & & \{\text{By method of construction}\} \\
\Rightarrow & I(X,ZY,R_x) \wedge I(R_x,ZX,Y) & \{\text{Intersection for } M\}
\end{array}
$$

where $R_x \cup R_y = R$, $R_x \cap R_y = \emptyset$ and for all $r \in R$ $r \in R_y$ if $<r, ZXY\backslash\{y,r\}, y>$ and $r \in R_x$ if $<x, ZXY\backslash\{x,r\}, r>$. So now we have

$$I(X,ZY,R_x) \wedge I(R_y,ZX,Y) \wedge I(X,ZR,Y) \Rightarrow I(X,Z,Y)$$

using intersection and decomposition. □

Before we continue with main results, two lemmas necessary for a forthcoming proof are presented.

Lemma 4.8 *Let $G = (U, E(G))$ be a UG. Let $W, X, Y, Z \subseteq U$ be disjoint sets of nodes. Then,*

$$<X,Z,Y> \wedge <X,Z,W> \Leftrightarrow <X,Z,YW> \quad (5)$$

Proof: Since intersection, decomposition and strong union hold for separation in undirected graphs:

$$
\begin{array}{rcll}
<X,Z,Y> \wedge <X,Z,W> & \Rightarrow & <X,ZW,Y> \wedge <X,ZY,W> & \{\text{Strong union}\} \\
 & \Rightarrow & <X,Z,YW> & \\
 & & \{\text{Since } X,\ Z,\ Y \text{ and } W \text{ disjoint and Intersection}\} &
\end{array}
$$

$$<X,Z,YW> \Rightarrow <X,Z,Y> \wedge <X,Z,W> \quad \{\text{Decomposition}\}$$

□

Lemma 4.9 *Let M be an independency model over a set of variables U closed under intersection and strong union. Let $W, X, Y, Z \subseteq U$ be disjoint sets of variables. Then*

$$I(X,Z,Y) \wedge I(X,Z,W) \Leftrightarrow I(X,Z,YW) \quad (6)$$

Proof:

$$
\begin{array}{rcll}
I(X,Z,Y) \wedge I(X,Z,W) & \Rightarrow & I(X,ZW,Y) \wedge I(X,ZY,W) & \{\text{Strong union}\} \\
 & \Rightarrow & I(X,Z,YW) & \\
 & & \{\text{Since } X,\ Z\ ,Y \text{ and } W \text{ disjoint and Intersection}\} &
\end{array}
$$

$$I(X,Z,YW) \Rightarrow I(X,Z,Y) \wedge I(X,Z,W) \quad \{\text{Decomposition}\}$$

Since any independency model satisfies decomposition this is not mentioned as a constraint for the independency model. □

Lemma 4.10 *Let M be an independency model closed under intersection, strong union and transitivity. Let $G = (U, E(G))$ be the undirected graph obtained from the complete graph by deleting edges (a,b) iff $I(a, U\backslash\{a,b\}, b) \in M$. Then, G is a D-map of M.*

Proof: The proof is based on the proof in [9]. Let $X, Y, Z \subseteq U$ be disjoint subsets.

We have to prove for every X, Y and Z that $I(X,Z,Y) \Rightarrow < X,Z,Y >$. We have $\forall_{x \in X, y \in Y} I(x,Z,y) \Leftrightarrow I(X,Z,Y)$ by applying (6) several times. Further, we have $\forall_{x \in X, y \in Y} <x,Z,y> \Leftrightarrow <X,Z,Y>$ by applying (5) several times.

So it is sufficient to show that $I(a,S,b) \Rightarrow <a,S,b>$ where $a, b \in U$ are single variables and $S \subset U$ is a set of variables.

Let n be the cardinality of U. The hypothesis $I(a,S,b) \Rightarrow <a,S,b>$ is shown by induction on the cardinality of S:

– for $|S| = n-2$ the lemma, is true by the method of construction of G.

– Now assume the lemma is true for $|S| = k \leq n-2$.

Let S' be a set for which $|S'| = k-1$. For $k \leq n-2$ there exists an element $c \in U \backslash S' \cup \{a,b\}$. So we have:

$$
\begin{array}{rcll}
I(a,S',b) & \Rightarrow & I(a,S'c,b) & \{\text{Strong union}\} \\
& \Rightarrow & I(a,S',c) \vee I(c,S',b) & \{\text{Transitivity}\} \\
& \Rightarrow & (I(a,S'b,c) \wedge I(a,S'c,b)) & \\
& & \quad \vee (I(c,S'a,b) \wedge I(a,S'c,b)) & \{\text{Strong union}\} \\
& \Rightarrow & (<a,S'b,c> \wedge <a,S'c,b>) & \\
& & \quad \vee (<c,S'a,b> \wedge <a,S'c,b> & \{\text{Induction hypothesis}\} \\
& \Rightarrow & <a,S',cb> \vee <a,S',cb> & \{\text{Intersection}\} \\
& \Rightarrow & <a,S',b> & \{\text{Decomposition}\}
\end{array}
$$

The second last step is true since the number of variables in the middle subset is k in each statement and by the induction hypothesis. □

Theorem 4.1 *For an independency model M to be UG-isomorph it is a necessary and sufficient condition for an independency model M to be closed under intersection, strong union and transitivity.*

Proof: First we proof sufficiency. By lemma 4.7 a UG can be constructed from an independency model that is closed under intersection such that the UG is an I-map the UG. By lemma 4.10 the UG constructed this way from an independency model that is closed under intersection, strong union and transitivity is a D-map. Therefore, the UG is a perfect map.

Now we proof necessity. Assume M would not be closed under intersection, strong union or transitivity. So, a clause $I(X,Z,Y) \wedge I(Q,S,R) \Rightarrow I(A,B,C) \vee I(D,E,F)$ exists satisfying intersection, strong union or transitivity such that all statements on the left-hand side are in M but none of the right-hand side statements are in M.

Suppose a UG G would exist such that G is a perfect map of M. Then, in G both $<X,Z,Y>$ and $<Q,S,R>$ would have to hold. Since intersection, strong union and transitivity hold for separation, this implies $<A,B,C> \vee <D,E,F>$ also have to hold. Since we assumed both $I(A,B,C)$ and $I(D,E,F)$ are not in M we have a contradiction. □

4.2 DAG-isomorphism

First, we will describe some properties of d-separation in DAGs. Next, we will investigate the conditions under which an independency model can be represented by a DAG. In the last part, we will introduce the notion of causal input lists and their properties.

4.2.1 Properties of d-separation

In this section we will show that the next properties hold for d-separation in DAGs. Let $G = (U, A(G))$ be a DAG. Let $W, X, Y, Z \subseteq U$ be sets of nodes and $a, b, c, d \in U$ single nodes.

$$
\begin{array}{rcll}
<X,Z,Y> & \Leftrightarrow & <Y,Z,X> & \{\text{Symmetry}\} \\
<X,Z,WY> & \Leftrightarrow & <X,Z,Y> \wedge <X,Z,W> & \{\text{(De)composition}\} \\
<X,Z,WY> & \Rightarrow & <X,ZW,Y> & \{\text{Weak union}\} \\
<X,Z,Y> \wedge <X,ZY,W> & \Rightarrow & <X,Z,WY> & \{\text{Contraction}\} \\
<X,ZW,Y> \wedge <X,ZY,W> & \Rightarrow & <X,Z,WY> & \{\text{Intersection}\} \\
<a,dc,b> \wedge <d,ab,c> & \Rightarrow & <a,d,b> \vee <a,c,b> & \{\text{Chordality}\} \\
<X,Z,Y> \wedge <X,Zc,Y> & \Rightarrow & <X,Z,c> \vee <c,Z,Y> & \{\text{Weak transitivity}\}
\end{array}
$$

Furthermore d-transitivity (see the list of axioms in section 3) is supposed to hold for d-separation. In the proofs that now follow we write $z \in t(x,y)$ if a $z \in Z$ exists such that z in the trail from x to y and z is not a head-to-head node in that trail. We write $\exists_{e \in U \setminus Z} \rightarrow e \leftarrow \in t(x,y)$ for a node e exists such that e is a head-to-head node in the trail and e is not in Z and all descendants of e are not in Z.

Lemma 4.11 *Separation in DAGs satisfies symmetry.*

Proof: Let $G = (U, A(G))$ be a DAG. Let $X, Y, Z \subseteq U$ be sets of nodes. Then,

$$
\begin{array}{rcl}
<X,Z,Y> & \Leftrightarrow & \forall_{x \in X, y \in Y, t(x,y)} \quad x \not\leftrightarrow y | Z \\
& \Leftrightarrow & \forall_{x \in X, y \in Y, t(x,y)} \quad y \not\leftrightarrow x | Z \\
& \Leftrightarrow & <Y,Z,X>
\end{array}
$$

□

Lemma 4.12 *Separation in DAGs satisfies (de)composition.*

Proof: Let $G = (U, A(G))$ be a DAG. Let $W, X, Y, Z \subseteq U$ be sets of nodes. Then,

$$
\begin{array}{rcl}
<X,Z,WY> & \Leftrightarrow & \forall_{x \in X, u \in WY, t(x,u)} \quad x \not\leftrightarrow u | Z \\
& \Leftrightarrow & \forall_{x \in X, y \in Y, t(x,y)} \quad x \not\leftrightarrow y | Z \\
& & \wedge \quad \forall_{x \in X, w \in W, t(x,w)} \quad x \not\leftrightarrow w | Z \\
& \Leftrightarrow & <X,Z,Y> \wedge <X,Z,W>
\end{array}
$$

□

Lemma 4.13 *Separation in DAGs satisfies weak union.*

Proof: Let $G = (U, A(G))$ be a DAG. Let $W, X, Y, Z \subseteq U$ be sets of nodes. Then,

$$
\begin{aligned}
<X, Z, WY> &\Leftrightarrow \forall_{x\in X, y\in Y, t(x,y)} x \not\leftrightarrow y|Z \wedge \forall_{x\in X, w\in W, t(x,w)} x \not\leftrightarrow w|Z \\
&\Rightarrow \forall_{x\in X, y\in Y, w\in W, t(x,y)} x \not\leftrightarrow y|Z \wedge (w \notin t(x,y) \vee w \in t(x,y)) \\
&\Rightarrow \forall_{x\in X, y\in Y, t(x,y)} x \not\leftrightarrow y|ZW \\
&\Rightarrow <X, ZW, Y>
\end{aligned}
$$

□

Lemma 4.14 *Separation in DAGs satisfies contraction.*

Proof: Let $G = (U, A(G))$ be a DAG. Let $W, X, Y, Z \subseteq U$ be sets of nodes. Then,

$$
\begin{aligned}
& <X, Z, Y> \wedge <X, ZY, W> \\
&\Leftrightarrow \forall_{x\in X, y\in Y, t(x,y)} x \not\leftrightarrow y|Z \wedge \forall_{x\in X, w\in W, t(x,w)} x \not\leftrightarrow w|YZ \\
&\Rightarrow \forall_{x\in X, y\in Y, t(x,y)} x \not\leftrightarrow y|Z \\
&\quad \wedge \forall_{x\in X, w\in W, t(x,w)} z \in t(x,w) \vee y \in t(x,w) \vee \exists_{e\in U\setminus YZ} \rightarrow e \leftarrow\in t(x,w) \\
&\Rightarrow \forall_{x\in X, y\in Y, t(x,y)} x \not\leftrightarrow y|Z \wedge \forall_{x\in X, w\in W, t(x,w)} x \not\leftrightarrow w|Z \\
&\Rightarrow <X, Z, WY>
\end{aligned}
$$

The second last step is true since $z \in t(x,w)$ implies $x \not\leftrightarrow w|Z$ by definition of blocked trails. For the same reason $\exists_{e\in U\setminus YZ} \rightarrow e \leftarrow\in t(x,w)$ implies $x \not\leftrightarrow w|Z$. And $y \in t(x,w)$ implies $x \not\leftrightarrow w|Z$ because the trail runs from w via y to x and $x \not\leftrightarrow y|Z$ since $<X, Z, Y>$. □

Lemma 4.15 *Separation in DAGs satisfies intersection.*

Proof: Let $G = (U, A(G))$ be a DAG. Let $W, X, Y, Z \subseteq U$ be sets of nodes. Then,

$$
\begin{aligned}
<X, ZW, Y> \wedge <X, ZY, W> &\Rightarrow \forall_{x\in X, y\in Y, t(x,y)} x \not\leftrightarrow y|ZW \\
&\quad \wedge \forall_{x\in X, w\in W, t(x,w)} x \not\leftrightarrow w|ZY \\
&\Rightarrow \forall_{x\in X, y\in Y, t(x,y)} x \not\leftrightarrow y|W \vee x \not\leftrightarrow y|Z \\
&\quad \wedge \forall_{x\in X, w\in W, t(x,w)} x \not\leftrightarrow w|Y \vee x \not\leftrightarrow w|Z \\
&\Rightarrow \forall_{x\in X, y\in Y, t(x,y)} x \not\leftrightarrow y|Z \wedge \forall_{x\in X, w\in W, t(x,w)} x \not\leftrightarrow w|Z \\
&\Rightarrow \forall_{x\in X, y\in WY, t(x,y)} x \not\leftrightarrow y|Z \\
&\Rightarrow <X, Z, WY>
\end{aligned}
$$

The third step is valid because of the following. Assume a trail $t(x,y)$ between a node $x \in X$ and a node $y \in Y$ would be bloked by W but not by Z. Then, $t(x,y)$ would have to contain a node in W. Consider the node $w \in t(x,y)$ closest to x such that $w \in W$ and the part of the trail between this node w and x $t(x,w) \subseteq t(x,y)$. From $<X, ZY, W>$ we know that this trail $t(x,w)$ is blocked by ZY. But, $x \leftrightsquigarrow y|Z$ implies $x \leftrightsquigarrow w|Z$ thus $x \not\leftrightsquigarrow w|Y$. Therefore, there is a $y' \in Y$ such that $y' \in t(x,w)$. But, then a trail from y' to x would exist that is not blocked by Z nor by W (we choose w closest to x) implying $<X, ZW, Y>$ does not hold. So, by contradiction we have every trail $t(x,y)$ is blocked by Z. For reasons of symmetry, the proof for trails $t(x,w)$ being blocked by Z goes analogous. □

Lemma 4.16 *Separation in DAGs satisfies chordality.*

Proof: Let $G = (U, A(G))$ be a DAG. Let $a, b, c, d \subseteq U$ be nodes. Then,

$$
\begin{aligned}
<a, cd, b> \quad & \Rightarrow \quad \forall_{t(a,b)} \; a \not\leftrightsquigarrow b|cd \\
& \Rightarrow \quad \forall_{t(a,b)} \; a \not\leftrightsquigarrow b|c \vee \forall_{t(a,b)} \; a \not\leftrightsquigarrow b|d \\
& \qquad \vee \exists_{t_1(a,b) \neq t_2(a,b)} \; c \in t_1(a,b) \wedge d \notin t_1(a,b) \;\; \wedge \;\; d \in t_2(a,b) \wedge c \notin t_2(a,b) \\
& \qquad\qquad \wedge \; \neg \exists_{e \in U \setminus \{c,d\}} \rightarrow e \leftarrow \in t_1(a,b)\} \vee \rightarrow e \leftarrow \in t_2(a,b) \\
& \Rightarrow \quad \forall_{t(a,b)} \; a \not\leftrightsquigarrow b|c \vee \forall_{t(a,b)} \; a \not\leftrightsquigarrow b|d \vee \exists_{t(c,d)} \; c \leftrightsquigarrow d|ab \\
& \Rightarrow \quad \forall_{t(a,b)} \; a \not\leftrightsquigarrow b|c \vee \forall_{t(a,b)} \; a \not\leftrightsquigarrow b|d \\
& \Rightarrow \quad <a, d, b> \vee <a, c, b>
\end{aligned}
$$

The second last step follows from $<c, ab, d>$ which excludes the existence of a trail $t(c,d)$ that is not blocked by ab. □

Lemma 4.17 *Separation in DAGs satisfies weak transitivity.*

Proof: Let $G = (U, A(G))$ be a DAG. Let $X, Y, Z \subseteq U$ be sets of nodes and $c \in U$ be a single node. Assume $<X, Z, Y>$, $< X, Zc, Y >$ and $\neg <X, Z, c>$ and $\neg <c, Z, Y>$ hold in G. Then,

$$
\begin{aligned}
\neg <X, Z, c> \wedge \neg <c, Z, Y> \quad & \Leftrightarrow \quad \exists_{x \in X, t(x,c)} \; x \leftrightsquigarrow c|Z \wedge \exists_{y \in Y, t(y,c)} \; y \leftrightsquigarrow c|Z \\
& \Rightarrow \quad \exists_{x \in X, y \in Y, t(x,y)} \; x \leftrightsquigarrow y|Z \\
& \qquad\qquad \{\text{By appending the trails via } c\} \\
& \qquad \vee \exists_{x \in X, y \in Y, t(x,y)} \; x \leftrightsquigarrow y|Zc \\
& \qquad\qquad \{\text{If } c \text{ in the trail forms a} \rightarrow\leftarrow \text{node}\} \\
& \Rightarrow \quad \neg <X, Z, Y> \vee \neg <X, Zc, Y>
\end{aligned}
$$

So the assumption was wrong and we have weak transitivity by contradiction. □

Furthermore it is assumed in [11] that separation in DAGs satisfies d-transitivity. Unfortunately, the proof in [11] might not be quite understood.

4.2.2 Conditions for DAG-isomorphism

A lot of work has been done on finding a set of axioms that characterizes a DAG-isomorph independency model in the same way the UG-isomorph independency model is characterized by its closure under strong union, intersection and transitivity. The fruitlessness of these efforts have resulted in the observation that DAG-isomorph independency models cannot be characterized by a set of axioms that are all Horn-clauses [3], i.e. axioms that have only one independency statement on the right-hand side of the clause. This research also resulted in the following theorem.

Theorem 4.2 *For an independency model M to be DAG-isomorph it is a necessary and sufficient condition for an independency model M to be closed under chordality, weak transitivity and d-transitivity.*

A proof can be found in [11] but might not be quite understood. The value of the d-transitivity axiom is that in the search for finding new axioms that characterize DAG-isomorph models this axiom can be used to derive them.

If an independency model M satisfies chordality, weak transitivity and d-transitivity, then a DAG that is a perfect map of M can be generated from this model by starting with an arcless graph. Next connect any two nodes a and b for which a —b. The results in an undirected graph called the skeleton of the DAG. For any triplet a, b and c for which a —b, c —b, $a \not- c$ and a —$c|b$ assign the directions $a \to b$ and $c \to b$. To the rest of the arcs a direction can be assigned such that no new independencies, i.e. head-to-head nodes, are induced.

4.3 MUG-isomorphism

A MUG is a set of undirected graphs. This section relies heavily on the work done by Dan Geiger in [4]. We only consider MUGs that contain UGs of a special form: the UG $G = (U', E(G))$ can be divided into three disjoint sets of nodes X, Y and Z such that $XYZ = U'$. The UG is formed by starting with a complete graph and deleting all edges between the nodes in X and Y. So, in the UG the nodes in XZ form a clique and the nodes in YZ form a clique. For such MUGs we have the following properties.

Lemma 4.18 *Let $W, X, Y, Z \subseteq U$ be sets of nodes. Then for MUG-separation the following properties hold:*

$<X, Z, Y>$	$\Leftrightarrow$	$<Y, Z, X>$	{*Symmetry*}
$<X, Z, WY>$	$\Rightarrow$	$<X, Z, Y> \wedge <X, Z, W>$	{*Decomposition*}
$<X, Z, WY>$	$\Rightarrow$	$<X, ZW, Y>$	{*Weak union*}

Proof: Remember the definition of MUG-separation: Z separates X from Y if there exists a UG $G = (V(G), E(G))$ in the MUG for which

1. Z separates X from Y in G and
2. $\forall_{a \in X \cup Y \cup Z} a \in V(G)$.

The above mentioned axioms share the properties that

- they hold for UG separation. This is necessary by the first condition of definition of MUG-separation.

- they are unary, i.e. they have only one statement on the left-hand side of the equation. So, contraction ($<X, ZW, Y> \wedge <X, Z, W> \Rightarrow <X, Z, WY>$) and intersection ($<X, ZW, Y> \wedge <X, ZY, W> \Rightarrow <X, Z, WY>$), which hold for UG separation, do not hold for MUG-separation since they can introduce new variables.

- they don't introduce new variables like transitivity and strong union do. Therefore, these last two axioms do not obey the second condition of the definition. We can have a MUG for example that exist of the two statements $I(a, c, b)$ and $I(a, d, b)$ with which two UGs are associated. But $\neg <a, cd, b>$ since both graphs don't contain all the variables a, b, c and d.

Therefore, the axioms symmetry, decomposition and weak union hold for MUG-separation.□

The properties mentioned in the previous lemma are in fact sufficient to derive any independency statement from a UG generated from an independency statement $I(X, Z, Y)$.

Theorem 4.3 *Let M be an independency model. Let MG be a MUG such that for every $I(X, Z, Y)$ MG contains a UG $G = (XYZ, E(G))$ with $(a, b) \in E(G)$ iff not $a \in X$ and $b \in Y$. Then, MG is a perfect map of M.*

Proof: Let $X, Y, Z \subseteq U$ be disjoint subsets. For a MUG to be a perfect map $I(X, Z, Y) \Leftrightarrow <X, Z, Y>$ must hold. First we show that $I(X, Z, Y) \Rightarrow <X, Z, Y>$. By the method of construction this must be clear.

Now we show that $<X, Z, Y> \Rightarrow I(X, Z, Y)$. Let $\sigma = <X, Z, Y>$ be a separation statement for a certain UG G in the MUG generated from $I(X, Z, Y)$ and let $\sigma' = <X', Z', Y'>$ be an arbitrary separation statement in G. Let $\sigma, \sigma_0, \sigma_1, .., \sigma_n, \sigma'$ be a sequence of separation statements that are derived using the axioms symmetry, weak union and decomposition. Since σ implies $I(X, Z, Y)$ and the axioms symmetry, weak union and decomposition hold for the independency model, we have by induction on the number of derivation steps that σ' implies $I(X', Z', Y')$. □

The only restriction for an independency model to be MUG-isomorph is that it obeys the axioms symmetry, weak union and decomposition. Since every independency model obeys these axioms, every independency model is MUG-isomorph.

By the method of construction, the number of graphs needed to represent a probability distribution is exponential in the number of variables on average. Therefore, the amount of memory to represent a distribution is also exponential on average. Also the time needed to find out if $I(X, Z, Y)$ holds in a MUG is exponential. Considering these properties, MUGs are not of much practical use.

4.4 MDAG-isomorphism

This section relies heavily on the work done by Dan Geiger in [4]. An MDAG can be constructed from an independency model by adding a DAG for every statement $I(X, Z, Y)$ in the model. This DAG contains nodes for every $a \in XYZ$ and the nodes in X form a clique, the nodes in Y form a clique and the nodes in Z form a clique. Further this DAG contains arcs from every node $z \in Z$ to every node $x \in X$ and every node $y \in Y$. This DAG does not contain any head-to-head node $a \rightarrow c \leftarrow b$ unless a and b are adjacent in the DAG. Therefore, the MDAG constructed this way is isomorph to the MUG corresponding to the model for every DAG in the MDAG a UG in the MUG exists with the same nodes and where the same separation statements hold. The properties that hold for MUGs also hold for this type of MDAGs.

Theorem 4.4 *Let M be an independency model. Let MG be a MUG such that for every $I(X,Z,Y)$ MG contains a UG $G = (XYZ, E(G))$ with $(a,b) \in E(G)$ iff not $a \in X$ and $b \in Y$. Let MG' be an MDAG such that for every $I(X,Z,Y)$ MG contains a DAG $G = (XYZ, A(G))$ with arcs between every pair of nodes but nodes in X and Y and no arcs pointing into nodes of Z unless from a node in Z itself. Then, MG and MG' have the same properties.*

Proof: For every UG in MG there is a DAG in MG' with the same set of nodes. No DAG contains a head-to-head node with uncoupled parents. By the observation that d-separation in a DAG with no head-to-head nodes becomes vertex separation in the underlying UG of such a DAG, we conclude that MDAG-separation in an MDAG that does not contain DAGs with head-to-head nodes becomes MUG-separation in the MUG that contains the set of underlying UGs of the DAGs in the MDAG. □

So, symmetry, weak union and decomposition also hold for MDAG-separation and the MDAG constructed in the way described is also a perfect map of the independency model it is constructed from. However, it needs exponential amount of memory to store and exponential time to find out if a particular independency statement holds or not. Therefore, MDAGs are not of practical use also.

5 The graphs and methods compared

In this section, we give a comparison between the different types of graphs and the methods of construction. In Table 3 an overview is given of the methods of construction of graphs described in this paper. For all methods it is supposed an independency model M is given. In the first column the name of the method is given, in the second column the type of graph, and in the last column a short description of the algorithm. In Table 4, an overview of the properties of the construction methods is given. The name in the first column refers to the name the first column of Table 3. In the second column the axioms are given under which the model must be closed to get a map as denoted in the last column.

Looking in the table may lead to think multigraphs are the best graphs to deal with since there are no restrictions on the independency model and the resulting graph is a perfect map of the model. However, these graphs are of no practical use since they require exponential amount of memory and computer power. The other two methods that result in perfect maps have the disadvantage that no class of probability distributions exists for which the corresponding independency model is closed under the axioms in Table 4. In other words, we cannot decide from some characteristics of the probability distribution whether the distribution is graph-isomorph. Therefore, we have to check for every statement in the independency model if these axioms apply. Now observe that there are exponential in $|U|$ elements in an independency model over U on average. So, it takes exponential time to check if a model is closed under these axioms.

Method of Construction	Resulting graph	Algorithm
UG-construction	$G = (U, E(G))$	$(a,b) \in E(G)$ iff $I(a, U\backslash ab, b) \notin M$
DAG-construction	$G = (U, A(G))$	1. Build UG $G' = (U, E(G'))$ with $(a,b) \in E(G')$ iff $\forall_{S \subseteq U \backslash ab} I(a,S,b) \notin M$ 2. Let $a \rightarrow b$ $c \rightarrow b$ if $(a,b),(c,b) \in E(G)$ and $(a,c) \notin E(G)$ and $\exists_{S \subseteq U \backslash abc} I(a,S,b) \in M$ 3. Let $(a,b) \in A(G)$ iff $(a,b) \in E(G')$ and no new head-to-head nodes arise
MUG-construction	$MG = (U, S(MG))$	For every $I(X,Z,Y) \in M$ add a UG $G = (XYZ, E(G))$ to $S(MG)$ such that $(a,b) \in E(G)$ iff not $a \in X$ and $b \in Y$
MDAG-construction	$MG = (U, S(MG))$	For every $I(X,Z,Y) \in M$ add a DAG $G = (XYZ, A(G))$ to $S(MG)$ such that $(a,b) \in A(G)$ or $(b,a) \in A(G)$ iff not $a \in X$ and $b \in Y$ and no arc points from a node in XY to a node in Z

Table 3: Overview of methods of graph construction

Method of Construction	Axioms the independency model must satisfy	Resulting type of map
UG-construction	Intersection	I-map
UG-construction	Intersection Strong Union Transitivity	perfect-map
DAG-construction	Chordality Weak Transitivity d-transitivity	perfect map
MUG-construction	none	perfect map
MDAG-construction	none	perfect map

Table 4: Overview of the properties of methods of construction

UG-construction is a method to obtain an I-map of the independency model. When a model is to be constructed from statistical data it is not a strong condition that the independency model must satisfy intersection since it satisfies intersection if the probability distribution is positive definite. Any discrete distribution can be made positive definite by assigning the zero-probability values a small positive number and correct the probabilities for the other values.

6 Conclusions

In this paper we have given an overview of the work done in constructing Bayesian belief networks using conditional independencies. Firstly, we investigated the notion of conditional independence. It turned out that weak transitivity does not hold for binary variables as suggested in literature. Secondly, we investigated different types of graphs, methods to construct these graphs and classes of independency models they can represent. Multigraphs turned out to be unpractical representations of independency models because of their computational unattractiveness. Undirected graphs can be constructed in $O(n^2)$ if the probability model is over a positive definite distribution. The resulting graph will be a minimal I-map.

References

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

[2] S.L. Lauritzen F.V. Jensen and K.G. Olesen. Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly*, 4:269–82, 1990.

[3] D. Geiger. The non-axiomatizablility of dependencies in directed acyclic graphs. Technical Report CSD-870000, UCLA, Cognitive Systems Laboratory, Computer Science Department, 1987.

[4] D. Geiger. Towards the formalization of informational dependencies. Technical Report CSD-880053, UCLA, Cognitive Systems Laboratory, Computer Science Department, 1988.

[5] 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.

[6] 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.

[7] Steffen L. Lauritzen. *Lectur notes in Coningency Tables.* Institue for Elctronic Systems Department of Methematics and Computer Science, University of Aalborg, Denmark, 1989.

[8] Z. Luo and A. Gammerman. *Bayesian learning in medical database, Technical Report TR910003.* Dept. of Computer Science, Heriot-Watt University, Edinburgh, UK, 1991. Technical Report.

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

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

[11] T. Verma and J. Pearl. Influence diagrams and d-separation. Technical Report CSD-880052, UCLA, Cognitive Systems Laboratory, Computer Science Department, 1988.

[12] J. Whittaker. *Graphical Models in Applied Mathematical Multivariate Statistics.* John Wiley & Sons Ltd., 1990.