Equivalent Characterization of a Class of Conditional Probabilistic Independencies

S.K.M. Wong and C.J. Butz

University of Regina, Regina, SK, Canada, S4S 0A2

Abstract. Markov networks utilize nonembedded probabilistic conditional independencies in order to provide an economical representation of a joint distribution in uncertainty management. In this paper we study several properties of nonembedded conditional independencies and show that they are in fact equivalent. The results presented here not only show the useful characteristics of an important subclass of probabilistic conditional independencies, but further demonstrate the relationship between relational theory and probabilistic reasoning.

1 Introduction

Belief networks [5, 6] utilize probabilistic conditional independencies to provide an economical representation of a joint distribution for managing uncertainty. A Bayesian network is a directed acyclic graph, explicitly specifying the *embedded* and *nonembedded* independency information, coupled with a corresponding set of conditional probability distributions. As performing inference in such a network may easily become intractable [2, 3], it is useful to transform a Bayesian network into a Markov network [5] albeit sacrificing the embedded independencies. That is, a Markov network is defined with respect to a hypergraph which explicitly specifies nonembedded conditional independencies only. This subclass of nonembedded independencies, called *generalized multivalued dependencies* (GMVDs) [13], is defined in a similar fashion to multivalued dependencies in relational database theory.

In this paper we study several properties of GMVDs and show that they are in fact equivalent. The results presented here not only show the useful characteristics of this important subclass of conditional independencies, but further demonstrate the intriguing relationship between relational database theory and probabilistic reasoning. It is perhaps worth mentioning that probabilistic conditional independencies do *not* have a complete axiomatization [8, 9] contrary to Pearl's conjecture [6]. However, it has been shown that GMVDs *do* in fact have a finite complete axiomatization [12].

Although the discussion in this paper draws heavily from [1], the exposition presented here is more general. For instance, in relational databases multivalued dependency [4] is a necessary and sufficient condition for a relation to be losslessly decomposed into two projections. However, multivalued dependency is a necessary but *not* a sufficient condition for the defining the notion of probabilistic conditional independence [13].

This paper is organized as follows. Section 2 contains background knowledge of our extended relational model [10, 11, 12, 13] for probabilistic reasoning. In Section 3, we analyze several properties of GMVDs and show that they are in fact equivalent.

2 Background

2.1 Hypergraphs and Hypertrees

A *hypergraph* [1, 7] is a pair $(\mathcal{N}, \mathbf{R})$, where $\mathcal{N}$ is a finite set of nodes and $\mathbf{R} = \{R_1, R_2, \ldots, R_n\}$ is a finite set of hyperedges which are arbitrary subsets of $\mathcal{N}$. An ordinary undirected graph without self loops is a hypergraph whose every hyperedge consists of two nodes. In this paper we assume $\mathcal{N} = \cup_{i=1}^{n} R_i$ and henceforth will often refer to the hypergraph $\mathbf{R}$ without mentioning the set $\mathcal{N}$ of nodes.

We call an element $R_i \in \mathbf{R}$, a *twig*, if there exists another distinct element $R_j \in \mathbf{R}$, such that $R_i \cap (\cup(\mathbf{R} - \{R_i\})) = R_i \cap R_j$. (By this definition, the hyperedge in a hypergraph consisting of a single hyperedge is not a twig). This means that the intersection of R_i and the hypergraph is contained in one hyperedge of the hypergraph. We call any such R_j a *branch* for the twig R_i, and note that a twig R_i may have many possible branches. A hypergraph is called a *hypertree* (an acyclic hypergraph [1, 7]) if its elements can be ordered, $R_1, R_2, \ldots, R_i$, such that R_i is a twig in the sub-hypergraph $R_1, R_2, \ldots, R_i$ for $i = 1, \ldots, n$. We call any ordering satisfying this condition a hypertree *construction ordering* for $\mathbf{R}$. (A hypertree construction ordering can also be represented as a *join tree* [1].) The first hyperedge R_1 in the hypertree construction ordering is called the root. Given a particular hypertree construction ordering, we can choose an integer $b(i)$, for $i = 2, \ldots, n$, such that $1 \leq b(i) \leq i-1$ and $R_{b(i)}$ is a branch for R_i in $R_1, R_2, \ldots, R_i$. We call such a function $b(i)$ satisfying this condition a *branching function* for $\mathbf{R}$. Note that a particular construction ordering may have many branching functions.

For example, the hypergraph $\mathbf{R} = \{R_1 = \{A_1, A_2, A_3\},\ R_2 = \{A_1, A_2, A_4\},\ R_3 = \{A_1, A_2, A_5\},\ R_4 = \{A_5, A_6\}\}$, depicted in Figure 1, is in fact a hypertree; for instance, the ordering R_1, R_2, R_3, R_4 is a hypertree construction ordering and $b(2) = 1, b(3) = 1, b(4) = 3$ defines one possible branching function.

A *path* from a node A_i to a node A_j is a sequence of hyperedges $R_1, R_2, \ldots, R_k$ $(k \geq 1)$ such that $A_i \in R_1$, $A_j \in R_k$, and $R_l \cap R_{l+1} \neq \emptyset$ if $1 \leq l < k$. We also say that the above sequence of hyperedges is a *hyperedge path* (or simply path when no confusion arises) from R_1 to R_k.

Two nodes (attributes) are *connected* if there is a path from one to the other. Similarly, two hyperedges are connected if there is a hyperedge path from one to the other. A set of nodes or hyperedges is connected if every pair is connected. The *connected components* are the maximal connected set of edges or hyperedges.

The following undirected graph terminology will also be useful in describing characteristics of GMVDs. A *clique* in a graph is a set of nodes such that every

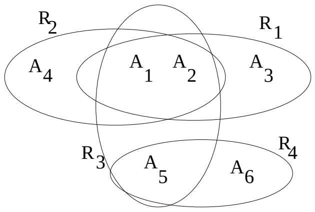

Fig. 1. A graphical representation of the hypergraph $\mathbf{R} = \{R_1, R_2, R_3, R_4\}$.

pair forms an edge of the graph. A *cycle* in a graph is a sequence $(A_1, A_2, \ldots, A_m)$ of nodes ($m \geq 3$) such that each A_i is distinct, except that $A_1 = A_m$, and (A_i, A_{i+1}) is an edge for $1 \leq i \leq m$.

Let $\mathbf{R}$ be a hypergraph. The *graph* of $\mathbf{R}$, denoted $G(\mathbf{R})$, has the same nodes as $\mathbf{R}$ and an edge between every pair of nodes that are in the same hyperedge of $\mathbf{R}$. Thus, the edges of $G(\mathbf{R})$ are precisely the set of all pairs (A_i, A_j) for which there is a hyperedge $R \in \mathbf{R}$ with $A_i, A_j \in R$.

2.2 Extended Relational Definitions

Let $\mathcal{N}$ be a finite set of distinct variables and let $X \subseteq \mathcal{N}$. Following [1], we define a X-tuple (or simply a *tuple* if X is understood) to be a function with domain X. Thus a tuple is a mapping that associates a value with each attribute in X. If $Y \subseteq X$ and t is a X-tuple, then $t[Y]$ denotes the Y-tuple obtained by restricting the mapping to Y. A X-relation (or a *relation over* X, or more simply a *relation* if X is understood) is a finite set of X-tuples. If r is a X-relation and $Y \subseteq X$, then the *projection* of r onto Y, denoted $r[Y]$, is the set of all tuples $t[Y]$ where t is in r.

We now extend the traditional relational concepts in order to express corresponding probabilistic concepts used in managing uncertain knowledge. Let r be a fixed relation representing the domain of a finite set of variables $\mathcal{N}$. A *joint probability distribution* [5, 6] over r, is a function ϕ on r assigning to each tuple $t \in r$ a real number $0 \leq \phi(t) \leq 1$ such that $\sum_{t \in r} \phi(t) = 1$. (We say the distribution is over $\mathcal{N}$ when the domain r is understood, and write ϕ as $\phi_{\mathcal{N}}$.) Suppose $\phi_{\mathcal{N}}$ is a joint probability distribution over $\mathcal{N} = \{A_1, A_2, \ldots, A_l\}$. We can succinctly express the probability distribution $\phi_{\mathcal{N}}$ as an *extended relation* $\Phi_{\mathcal{N}}$ with attributes $\{A_1, A_2, \ldots, A_l, f_{\phi_{\mathcal{N}}}\}$ as shown in Figure 2. Each row in $\Phi_{\mathcal{N}}$ corresponds to a tuple $t_i \in r$, and s is the cardinality of r.

Let ϕ_X and ϕ_Y be two distributions over X and Y respectively. We can express the product of these distributions $\phi_X \cdot \phi_Y$ as the *product join* of the extended relations Φ_X and Φ_Y, denoted by $\Phi_X \times \Phi_Y$, as follows:

> $\Phi_X \times \Phi_Y$ is the *extended* relation obtained by adding a new column with attribute $f_{\phi_X \cdot \phi_Y}$ to the relation $r = \Phi_X[X] \bowtie \Phi_Y[Y]$, where $\bowtie$ is the *natural*

$$\Phi_{\mathcal{N}} = \begin{array}{|llll l|}
\hline
A_1 & A_2 & \ldots & A_l & f_{\phi_{\mathcal{N}}} \\
\hline
t_1[A_1] & t_1[A_2] & \ldots & t_1[A_l] & t_1[f_{\phi_{\mathcal{N}}}] = \phi_{\mathcal{N}}(t_1[\mathcal{N}]) \\
t_2[A_1] & t_2[A_2] & \ldots & t_2[A_l] & t_2[f_{\phi_{\mathcal{N}}}] = \phi_{\mathcal{N}}(t_2[\mathcal{N}]) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
t_s[A_1] & t_s[A_2] & \ldots & t_s[A_l] & t_s[f_{\phi_{\mathcal{N}}}] = \phi_{\mathcal{N}}(t_s[\mathcal{N}]) \\
\hline
\end{array}$$

Fig. 2. A joint distribution $\phi_{\mathcal{N}}$ expressed as an *extended relation* $\Phi_{\mathcal{N}}$.

join operator [1]. (r is referred to as the domain of $\Phi_X \times \Phi_Y$.) For each tuple $t \in \Phi_X \times \Phi_Y$, t[XY] = r[XY], and $t[f_{\phi_X \cdot \phi_Y}] = \phi_X(t[X]) \cdot \phi_Y(t[Y])$.

If $\Phi_{\mathcal{N}}$ is an extended relation and $X \subseteq \mathcal{N}$, then the marginalization of $\Phi_{\mathcal{N}}$ onto X is the *marginal* distribution, denoted $\Phi_{\mathcal{N}}^{\downarrow X}$, with attributes $X \cup \{f_{\phi_{\mathcal{N}}^{\downarrow X}}\}$, defined by:

$$\Phi_{\mathcal{N}}^{\downarrow X} = \{t \mid t[X] \in \Phi_{\mathcal{N}}[X], \text{ and } t[f_{\phi_{\mathcal{N}}^{\downarrow X}}] = \phi_{\mathcal{N}}^{\downarrow X}(t[X]) = \sum_{t' \in \Phi_{\mathcal{N}}} \phi_{\mathcal{N}}(t') \text{ where } t[X] = t'[X]\}.$$

Due to limited space, we refer the reader to [10, 11, 12, 13] for examples illustrating the product join $\times$ and marginalization $\downarrow$ operators.

Let $\Phi_{\mathcal{N}}$ be an extended relation and $X \subseteq \mathcal{N}$. The *inverse* extended relation $(\Phi_{\mathcal{N}}^{\downarrow X})^{-1}$ is defined from $\Phi_{\mathcal{N}}^{\downarrow X}$ by renaming attribute $f_{\phi_{\mathcal{N}}^{\downarrow X}}$ as $f_{(\phi_{\mathcal{N}}^{\downarrow X})^{-1}}$, and

$$t[f_{(\phi_{\mathcal{N}}^{\downarrow X})^{-1}}] = \begin{cases} 1 \,/\, t[f_{\phi_{\mathcal{N}}^{\downarrow X}}] & \text{if } t[f_{\phi_{\mathcal{N}}^{\downarrow X}}] > 0 \\ 0 & \text{otherwise.} \end{cases}$$

If Φ is an extended relation over $\mathcal{N}$ and $X, Y \subseteq \mathcal{N}$, then we say that the *generalized multivalued dependency* (GMVD) [12], denoted $X \multimap\!\!\rightarrow Y$ holds for Φ if

$$\Phi = \Phi^{\downarrow XY} \otimes \Phi^{\downarrow XZ} = \Phi^{\downarrow X} \times \Phi^{\downarrow Y} \times (\Phi^{\downarrow X \cap Y})^{-1},$$

where $Z = \mathcal{N} - XY$, and $\otimes$ is called the *generalized join* operator.

For example, consider the joint distribution Φ_R on $R = \{A_1, A_2, A_3, A_4, A_5\}$ shown in Figure 3. Φ_R satisfies the GMVD $\{A_1\} \multimap\!\!\rightarrow \{A_2, A_3, A_4\}$ since $\Phi_R = \Phi_R^{\downarrow \{A_1, A_2, A_3, A_4\}} \otimes \Phi_R^{\downarrow \{A_1, A_5\}}$ as shown in Figure 4.

We say that an extended relation Φ_R over attributes $R = R_1 \cup R_2 \cup \ldots \cup R_n$ obeys the *generalized acyclic join dependency* (GAJD), denoted $\otimes\{R_1, R_2, \ldots, R_n\}$ or $\otimes\mathbf{R}$, if Φ_R can be expressed as

$$\Phi_R = (\ldots((\Phi^{\downarrow R_1} \otimes \Phi^{\downarrow R_2}) \otimes \Phi^{\downarrow R_3}) \ldots \otimes \Phi^{\downarrow R_n}),$$

where $R_1, R_2, \ldots, R_n$ is a hypertree construction ordering for $\mathbf{R} = \{R_1, R_2, \ldots, R_n\}$.

Let $\sum$ be a set of dependencies and σ be a single dependency. We say that $\sum$ *logically implies* σ if whenever every dependency in $\sum$ holds for an extended

$$\Phi_R \ = \ \begin{array}{|cccccc|} \hline A_1 & A_2 & A_3 & A_4 & A_5 & f_{\phi_R} \\ \hline 0 & 0 & 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1 & 1 & 0.2 \\ 0 & 1 & 0 & 0 & 1 & 0.2 \\ 0 & 1 & 1 & 0 & 1 & 0.2 \\ 1 & 2 & 1 & 1 & 0 & 0.2 \\ \hline \end{array}$$

Fig. 3. A joint distribution ϕ_R on $R = \{A_1, A_2, A_3, A_4, A_5\}$ expressed as the *extended relation* Φ_R.

$$\begin{array}{|ccccc|} \hline A_1 & A_2 & A_3 & A_4 & f_{\phi_R^{\downarrow\{A_1,A_2,A_3,A_4\}}} \\ \hline 0 & 0 & 0 & 0 & 0.2 \\ 0 & 1 & 1 & 1 & 0.2 \\ 0 & 1 & 0 & 0 & 0.2 \\ 0 & 1 & 1 & 0 & 0.2 \\ 1 & 2 & 1 & 1 & 0.2 \\ \hline \end{array} \otimes \begin{array}{|ccc|} \hline A_1 & A_5 & f_{\phi_R^{\downarrow\{A_1,A_5\}}} \\ \hline 0 & 1 & 0.8 \\ 1 & 0 & 0.2 \\ \hline \end{array} = \begin{array}{|cccccc|} \hline A_1 & A_2 & A_3 & A_4 & A_5 & f_{\phi_R^{\downarrow\{A_1,A_2,A_3,A_4\}} \cdot \phi_R^{\downarrow\{A_1,A_5\}} / \phi_R^{\downarrow\{A_1\}}} \\ \hline 0 & 0 & 0 & 0 & 1 & ((0.2)(0.8)/0.8) = 0.2 \\ 0 & 0 & 1 & 1 & 1 & ((0.2)(0.8)/0.8) = 0.2 \\ 0 & 1 & 0 & 0 & 1 & ((0.2)(0.8)/0.8) = 0.2 \\ 0 & 1 & 1 & 0 & 1 & ((0.2)(0.8)/0.8) = 0.2 \\ 1 & 2 & 1 & 1 & 0 & ((0.2)(0.2)/0.2) = 0.2 \\ \hline \end{array}$$

Fig. 4.
The distribution Φ_R in Figure 3 satisfies the GMVD $\{A_1\} -\!\circ\!\!\rightarrow \{A_2, A_3, A_4\}$ since $\Phi_R = \Phi_R^{\downarrow\{A_1,A_2,A_3,A_4\}} \otimes \Phi_R^{\downarrow\{A_1,A_5\}}$.

relation Φ, then σ also holds for Φ. That is, there is no "counterexample extended relation" Φ such that every dependency in $\sum$ holds for Φ but σ does not hold for Φ.

Since the concept of nonembedded conditional independence is the focus of our discussion, we conclude this section by distinguishing it from an embedded conditional independence. Consider the joint probability distribution ϕ_R on $R = \{A_1, A_2, A_3, A_4, A_5\}$ as shown in Figure 3. It can be verified that the conditional independence of $\{A_2\}$ and $\{A_3\}$ given $\{A_1\}$ does not hold with respect to ϕ_R. That is,

$$\phi_R \neq \frac{\phi_R^{\downarrow\{A_1,A_2\}} \cdot \phi_R^{\downarrow\{A_1,A_3,A_4,A_5\}}}{\phi_R^{\downarrow\{A_1\}}}.$$

However, the conditional independence of $\{A_2\}$ and $\{A_3\}$ given $\{A_1\}$ holds in the marginal distribution $\phi_R^{\downarrow\{A_1,A_2,A_3\}}$, namely:

$$\phi_R^{\downarrow\{A_1,A_2,A_3\}} = \frac{(\phi_R^{\downarrow\{A_1,A_2,A_3\}})^{\downarrow\{A_1,A_2\}} \cdot (\phi_R^{\downarrow\{A_1,A_2,A_3\}})^{\downarrow\{A_1,A_3\}}}{(\phi_R^{\downarrow\{A_1,A_2,A_3\}})^{\downarrow\{A_1\}}}.$$

We call such an independency an embedded conditional independency with respect to the distribution ϕ_R.

3 Equivalent Characterizations of GMVDs

In this section we derive several characterizations the set M of generalized multivalued dependencies (GMVDs) such that M is the consequence of a given generalized acyclic join dependency (GAJD). Since it has been shown [12] that every GMVD $X \multimap\!\!\!\rightarrow Y$ is equivalent to the GMVD $X \multimap\!\!\!\rightarrow Y - X$, we will only consider those GMVDs with $X \cap Y = \emptyset$ to simplify the notation. Thus M^+ will denote the set of all GMVDs $X \multimap\!\!\!\rightarrow Y$, with $X \cap Y = \emptyset$, that are logically implied by the set M.

If $\mathcal{H}$ is a hypergraph, then the set of GMVDs *generated by* $\mathcal{H}$ is the set of GMVDs $X \multimap\!\!\!\rightarrow Y$, where Y is the union of some connected components of the hypergraph $\mathcal{H} - X$ obtained from $\mathcal{H}$ by deleting the set X of nodes. That is, $\mathcal{H} - X = \{E - X \mid E \text{ is an edge of } \mathcal{H}\} - \{\emptyset\}$. We then say that X *separates off* Y from the rest of the nodes. A set M of GMVDs is *hypergraph generated* if there is a hypergraph that generates M. Similarly, M is *graph generated* if there is a graph (treated as a hypergraph) that generates M.

A GMVD $X \multimap\!\!\!\rightarrow Y$ ($X \cap Y = \emptyset$) *splits* two attributes A_i and A_j if one of them is in Y and the other is in $\mathcal{N} - XY$, where $\mathcal{N}$ is the set of all the attributes. A set M of GMVDs splits A_i and A_j if some GMVD in M splits them.

Lemma 1. *Two attributes A_i and A_j are split by a set M of generalized multivalued dependencies if and only if they are split by its closure.*

Lemma 1 indicates that two logically equivalent sets of GMVDs split exactly the same pairs of attributes.

Given a set M of GMVDs, we can construct a graph $G(M)$ with the attributes as nodes and an edge (A_i, A_j) between two attributes A_i and A_j if A_i and A_j are not split by M. For example, let $\mathcal{N} = \{A_1, A_2, A_3, A_4\}$ and $M = \{A_1 \multimap\!\!\!\rightarrow A_3, A_3 \multimap\!\!\!\rightarrow A_4\}$. The first GMVD splits A_2 and A_3, and A_3 and A_4, while the second GMVD splits A_1 and A_4, and A_2 and A_4. The set of edges in the graph $G(M)$ is $\{(A_1, A_2), (A_1, A_3)\}$.

Lemma 2. *Let M be a set of GMVDs, $G(M)$ its graph, and N the set of GMVDs generated by $G(M)$. Then $M^+ \subseteq N$.*

The converse to Lemma 2 does not necessarily hold. In the last example, the GMVD $\emptyset \multimap\!\!\!\rightarrow A_4$ is graph generated by $G(M)$ but is not logically implied by M. It will be shown that the converse holds exactly for those sets of GMVDs that form a cover of the set of GMVDs implied by a given generalized acyclic join dependency, where M_1 is a *cover* of M_2 if $M_1^+ = M_2^+$.

We say that M has the *intersection property* if whenever the GMVDs $X \multimap\!\!\!\rightarrow Z$ and $Y \multimap\!\!\!\rightarrow Z$ are implied by M (with Z disjoint from both X and Y), then $X \cap Y \multimap\!\!\!\rightarrow Z$ is also implied by M.

Let M be a set of GMVDs. Two disjoint sets X and Y are called *orthogonal* if the GMVD $\mathcal{N} - XY \multimap\!\!\!\rightarrow X$ (or equivalently, by the complementation rule for GMVDs [12], $\mathcal{N} - XY \multimap\!\!\!\rightarrow Y$) is implied by M. It follows from Lemma 1 and the rules for manipulating GMVDs [12] that two attributes A_i and A_j are orthogonal (i.e., the singleton sets $\{A_i\}$ and $\{A_j\}$ are orthogonal) if and only if they are split by M. It follows from the rules for manipulating GMVDs [12] that if X and Y are orthogonal, then for every pair A_i and A_j of attributes where $A_i \in X$ and $A_j \in Y$, necessarily A_i and A_j are orthogonal. We say that M has the orthogonal property if the converse also holds. M has the *orthogonal property* if every two sets X and Y are orthogonal whenever every attribute of X is orthogonal to every attribute of Y.

Theorem 3. *Let M be a set of generalized multivalued dependencies. The following are equivalent:*

(1) M is a cover of the set of GMVDs implied by some GAJD.
(2) M^+ is hypergraph generated.
(3) M^+ is graph generated.
(4) There is exactly one graph that generates M^+.
(5) M^+ is the set of GMVDs generated by $G(M)$.
(6) M^+ has the intersection property.
(7) M^+ has the orthogonal property.

Proof. We will show that (1) and (2), and (2) and (3) are equivalent. We then show (3) $\Rightarrow$ (6) $\Rightarrow$ (7) $\Rightarrow$ (5) $\Rightarrow$ (3). Hence, conditions $(1),(2),(3),(5),(6),(7)$ are equivalent. Finally, we show (5) $\Rightarrow$ (4) $\Rightarrow$ (3) which shows (4) is equivalent to the others.
(1) $\Leftrightarrow$ (2): The set of GMVDs implied by a generalized acyclic join dependency $\otimes\mathbf{R}$ is exactly the set of GMVDs generated by the hypergraph $\mathbf{R}$ [12]. Thus, if M is the set of GMVDs implied by the generalized acyclic join dependency $\otimes\mathbf{R}$, then we know that M^+ is the set of GMVDs generated by the hypergraph $\mathbf{R}$.
(2) $\Rightarrow$ (3): Let $\mathcal{H}$ be a hypergraph, and let $G = G(\mathcal{H})$ be the graph of $\mathcal{H}$. It is easy to see that a set X of nodes separates off another set Y in $\mathcal{H}$ if and only if X separates off Y in G. Therefore, the set of GMVDs generated by $\mathcal{H}$ is the same set of GMVDs generated by G.
(3) $\Rightarrow$ (2): Obvious, since every graph is a hypergraph.
(3) $\Rightarrow$ (6): Let $\mathcal{H}$ be the hypergraph that generates M^+. Suppose $X -\!\circ\!\!\to Z$ and $Y -\!\circ\!\!\to Z$ are in M^+. Since $X -\!\circ\!\!\to Z$ is in M^+, X separates off Z from $\mathcal{N} - XZ$. Thus, no node in Z is connected to a node in $\mathcal{N} - XZ$. Since $Y -\!\circ\!\!\to Z$ is in M^+, Y separates off Z from $\mathcal{N} - YZ$. Thus, no node in Z is connected to a node in $\mathcal{N} - YZ$. Therefore, no node in Z is connected to a node in $\mathcal{N} - [(X \cap Y) \cup Z]$. Therefore, $X \cap Y$ separates off Z from $\mathcal{N} - [(X \cap Y) \cup Z]$. By definition, $X \cap Y -\!\circ\!\!\to Z$ is in M^+.
(6) $\Rightarrow$ (7): Let $X = \{A_1, A_2, \ldots, A_k\}$ and $Y = \{B_1, B_2, \ldots, B_m\}$ be two disjoint sets with every A_i orthogonal to every B_j. Let $Z = \mathcal{N} - XY$, $X_i = X - A_i (i = 1, 2, \ldots, k)$, and $Y_j = Y - B_j (j = 1, 2, \ldots, m)$. Since every A_i is orthogonal to every B_j, we have $ZX_iY_j -\!\circ\!\!\to A_i$ for each i and j. Since M has the intersection property, we have $\cap\{ZX_iY_j \mid j = 1, 2, \ldots, m\} -\!\circ\!\!\to A_i$. But $\cap\{ZX_iY_j \mid j = 1, 2, \ldots, m\} = ZX_i$. Hence, $ZX_i -\!\circ\!\!\to A_i$, or equivalently by the complementation rule for GMVDs [12], $ZX_i -\!\circ\!\!\to Y$ for each i. Again from the intersection property,

$\cap\{ZX_i \mid i=1,2,\ldots,k\} \multimap\!\!\rightarrow Y$. Since $\cap\{ZX_i \mid i=1,2,\ldots,k\} = Z$, we have $Z \multimap\!\!\rightarrow Y$, or equivalently, $Z \multimap\!\!\rightarrow X$.
(7) $\Rightarrow$ (5): Suppose that M has the orthogonal property, and let N be the set of GMVDs graph generated from $G(M)$. By Lemma 2, $M^+ \subseteq N$. For the other inclusion, let $X \multimap\!\!\rightarrow Y$ be a GMVD in N, and let $Z = \mathcal{N} - XY$. By definition of N, there is no edge in $G(M)$ connecting a node in Y to a node in Z. Thus every attribute of Y is orthogonal to every attribute of Z. Then by the orthogonality property, Y is orthogonal to Z, and $X \multimap\!\!\rightarrow Y$ is in M^+.
(5) $\Rightarrow$ (3): Obvious.
(5) $\Rightarrow$ (4): Let G be the graph that generates M^+. Attributes A_i and A_j are split by M if and only if the edge (A_i, A_j) is not in G. But also, A_i and A_j are split by M if and only if the edge (A_i, A_j) is not in $G(M)$. Therefore, $G = G(M)$.
(4) $\Rightarrow$ (3): Obvious.

References

1. Beeri, C. and Fagin, R. and Maier, D. and Yannakakis, M.: On the Desirability of Acyclic Database Schemes. J. ACM **30**(3) (1983) 479–513
2. Cooper, G.F.: The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks. Artificial Intelligence. **42** (1990) 393-402
3. Dagum, P., Luby, M.: Approximating Probabilistic Inference in Bayesian Belief Networks is NP-hard. Artificial Intelligence. **60**(1) (1993) 141-153
4. Fagin, R.: Multivalued Dependencies and a New Normal Form for Relational Databases. ACM Transactions on Database Systems. **2**(3) (1977) 262–278
5. Hajek, P., Havranek, T., Jirousek, R.: Uncertain Information Processing in Expert Systems. CRC Press. (1992)
6. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann. San Francisco, California. (1988)
7. Shafer, G.: An Axiomatic Study of Computation in Hypertrees. University of Kansas. School of Business Working Papers (232). (1991)
8. Studeny, M.: Conditional Independence Relations Have No Finite Complete Characterization. Eleventh Prague Conference on Information Theory, Statistical Decision Foundation and Random Processes. (1990)
9. Wong, S.K.M., Wang, Z.W.: On Axiomatization of Probabilistic Conditional Independence. Tenth Conference on Uncertainty in Artificial Intelligence. (1994) 591–597
10. Wong, S.K.M., Butz, C.J., Xiang, Y.: A Method for Implementing a Probabilistic Model as a Relational Database. Eleventh Conference on Uncertainty in Artificial Intelligence. (1995) 556–564
11. Wong, S.K.M.: Testing Implication of Probabilistic Dependencies. Twelfth Conference on Uncertainty in Artificial Intelligence. (1996) 545–553
12. Wong, S.K.M.: The Relational Structure of Belief Networks. (submitted for publication) (1997)
13. Wong, S.K.M.: An Extended Relational Data Model for Probabilistic Reasoning. Journal of Intelligent Information Systems. **9** (1997) 181–202