A Lattice Approach to Bayesian Networks

Abstract

Bayesian networks and conditional independence is studied via functional dependences. Armstrong's axioms known from the theory of relational databases is used to reformulate the concept of Bayesian networks into the theory of join-semidistributive lattices. Lattice theory provides us with a richer language to discuss causation than graph theory does.

1 Introduction

The theory of Bayesian networks is normally considered as a theory of conditional independence. In this paper we shall take a different point of view and consider the theory of Bayesian networks as a theory of functional dependence. Bayesian networks has become a popular model of causation among computer scientists. Functional dependencies has definitely played a role in philosophers attempts to understand the concepts cause and effect, so from a philosophical point of view the ideas presented here are not all that new and discussion and references can be found in (Pearl, 2000, Section 1.4). Nevertheless this point of view has not been explored in sufficient detail in relation to Bayesian networks. Although there are still a number of technical problems to be solved, the overall approach should be clear already at this point.

This work use ideas from three areas. The theory of functional dependencies in relational databases was pioneered by Armstrong (1974). Now the basic results can be found in any textbook like Gardarin & Valduriez (1989) on relational databases, but in the database community this branch of research is not really active anymore. The theory of lattices was developed by Birkhoff and his reprinted textbook Birkhoff (1991) on this subject is still worth reading. A comprehensive modern exposition can be found in Grätzer (2003). Lattice theory is now a mature research area with unified notation plenty of results and applications in all branches of mathematics. The study of causation has an enormous literature that started about 2500 years ago. The study of causation via Bayesian networks is much younger and a good overview of this approach to the study of causation can be found in Pearl (1988, 2000); Lauritzen (1998). Many of the definitions that we use may be found in these textbooks

2 Independence and functional dependence

In the theory of Bayesian networks one studies the relation $(A \perp B \mid C)_I$ (A and B are independent given C), where A, B and C are disjoint subsets of a set M of random variables. The power set of M is a Boolean lattice with inclusion as ordering, $\cup$ as join, and $\cap$ as met. We shall say that a relation $(\cdot \perp \cdot \mid \cdot)_I$ on a lattice $(L, \vee, \wedge)$ is a *semi-graphoid relation*, if it satisfies the following axioms:

Symmetry $(X \perp Y \mid Z)_I$ if and only if $(Y \perp X \mid Z)_I$.

***I*-Decomposition**

$(X \perp Y \vee W \mid Z)_I$ implies $(X \perp Y \mid Z)_I$.

Weak Union

$(X \perp Y \vee W \mid Z)_I$ implies $(X \perp W \mid Z \vee Y)_I$.

Contraction $(X \perp Y \mid Z)_I$ and $(X \perp W \mid Z \vee Y)_I$ implies $(X \perp Y \vee W \mid Z)_I$.

These propositions should hold for all $X, Y, Z, W \in L$. If L is the power set of a set and the relation $(\cdot \perp \cdot \mid \cdot)_I$ is only defined for disjoint sets then the definition coincides with the definition given in Pearl (1988).

In this paper we are also interested in the case where the subsets are not disjoint. In a power set of random variables we note that if A is independent of A given

C then A is a function of C almost surely. Hence we introduce the following additional axioms that are fulfilled for random variables.

Auto-independence

For all subsets A we have $(A \perp A \mid A)_I$.

Forced independence

For all subsets A, B and C we have that $(A \perp A \mid C)_I$ implies that $(A \perp B \mid C)_I$.

A semi-graphoid relation is said to be *super-graphoid* if it satisfies auto-independence and forced independence.

If $(A \perp A \mid C)_I$ we write $C \to_f A$ and say that A *depends functionally* on C. It is important to note that the notion of functional dependences may also be defined even in some cases where there are no probability measure and no semi-graphoid relation in play. For instance functional dependences are defined in relational databases. We can use the basic properties of semi-graphoid relations to prove properties of functional dependence.

Proposition 1 *If* $(\cdot \perp \cdot \mid \cdot)_I$ *is a super-graphoid relation then it satisfies:*

Functional Independence

If $X \to_f W$ *and* $(X \perp Y \mid Z)_I$ *then* $(W \perp Y \mid Z)_I$.

Causal Independence

If $Z \to_f W$ *and* $(X \perp Y \mid Z)_I$ *then* $(X \perp Y \mid W)_I$.

Theorem 2 *If* $(\cdot \perp \cdot \mid \cdot)_I$ *is a super-graphoid relation then it satisfies:*

Reflexivity *If* $Y \leq X$*, then* $X \to_f Y$.

Augmentation *If* $X \to_f Y$*, then* $X \vee Z \to_f Y \vee Z$.

Transitivity *If* $X \to_f Y$ *and* $Y \to_f Z$*, then* $X \to_f Z$.

***f*-Decomposition** If $X \to_f Y \vee Z$, then $X \to_f Y$ and $X \to_f Z$.

Union If $X \to_f Y$ and $X \to_f Z$, then $X \to_f Y \vee Z$.

Psudotransitivity If $X \to_f Y$ and $Y \vee Z \to_f W$, then $X \vee Z \to_f W$.

Proof. Reflexivity Assume that Y is a subset of X. We know that $X \to_f X$. Hence $(X \perp Y \mid X)_I$ because of forced independence. We have $X = Y \vee X$. Hence $Y \perp Y \mid X$.

Augmentation Assume that $X \to_f Y$. Now $X \vee Z \to_f Y \vee Z$ is equivalent to $((Y \vee Z) \perp (Y \vee Z) \mid (X \vee Z))_I$ so it is sufficient to prove that

$$(Y \perp (Y \vee Z) \mid (X \vee Z))_I$$

and

$$(Z \perp (Y \vee Z) \mid (X \vee Z \vee Y))_I .$$

The last proposition holds because $X \vee Z \vee Y \to_f Z$ by reflexivity together with forced independence. Therefore it is sufficient to prove that $X \vee Z \to_f Y$. Using forced independence we have $(Y \perp Y \vee Z \mid X)_I$, which by weak union implies $(Y \perp Y \mid X \vee Z)_I$.

***f*-Decomposition** Assume that $X \to_f Y \vee Z$. Then $((Y \vee Z) \perp (Y \vee Z) \mid X)_I$ and according to decomposition of independence we have $(Y \perp (Y \vee Z) \mid X)_I$ and applied once more $(Y \perp Y \mid X)_I$. The proof that $X \to_f Z$ works in the same way.

Transitivity Assume that $X \to_f Y$ and $Y \to_f Z$. Forced independence implies that $(Z \perp (Z \vee X) \mid Y)_I$ and by weak union $Z \perp Z \mid Y \vee X$. Forced independence then implies $(Z \perp (Y \vee Z) \mid (X \vee Y))_I$. Forced independence and $X \to_f Y$ together implies that $(Y \perp (Y \vee Z) \mid X)_I$. Contraction then implies $((Y \vee Z) \perp (Y \vee Z) \mid X)_I$ and $X \to_f Y \vee Z$. Now $X \to Z$ follows by decomposition.

The proofs that **Psudo-transitivity** and **Union** holds are omitted. ■

In database literature reflexivity, augmentation, and transitivity are called *Armstrong's Axioms* for functional dependence. Union, *f*-Decomposition and Psudo-transitivity can be deduced from Armstrong's Axioms, see Armstrong (1974).

Let L denote a lattice with a relation $\to$ such that Armstrong's axioms are satisfied. For simplicity we will assume that L is finite. For $X \in L$ we define $cl_{\to}(X)$ as $\bigvee Y_i$ where the join is taken over all Y_i such that $X \to Y_i$. In a set of random variables where $\to$ denotes functional dependence $cl_{\to}(X)$ is the set of all variables determined by the variables X. We say that X is closed if $cl_{\to}(X) = X$.

Theorem 3 *If* $(L, \vee, \wedge)$ *is a lattice with an preordering* $\to$ *satisfying Armstrong's axioms then the set of elements in* L *that are closed form a lattice with met* $X \wedge_{\to} Y = X \wedge Y$ *and join* $X \vee_{\to} Y = cl_{\to}(X \vee Y)$.

This theorem essentially dates back to Armstrong when L was a power set with inclusion as ordering and the proof is the same, so the proof is omitted here. The theorem as it is formulated here probably in the literature on lattices although the author has not been able to locate a good reference.

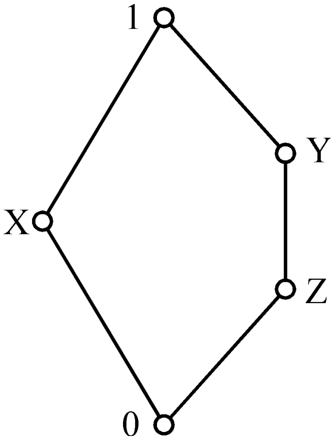

Figure 1: Hasse diagram of the lattice N_5 .

Theorem 4 *If* $(L,\vee,\wedge)$ *is a lattice with a super-graphoid relation* $(\cdot\perp\cdot\mid\cdot)_I$ *then this relation restricted to the lattice* $L_{\to_f}$ *is also super-graphoid.*

Proof. One just has to prove that $(X\perp Y\mid Z)_I$ if and only if $\left(X\perp cl_{\to_f}(Y)\mid Z\right)_I$ if and only if $\left(X\perp Y\mid cl_{\to_f}(Z)\right)_I$. This follows from Proposition 1. ■

The significance of this theorem is that if we start with a super-graphoid relation on a power set of random variables then this super-graphoid relation is also super-graphoid when restricted the set of random variables that are closed under functional dependence.

3 Canonical independence in a lattice

Even simple examples of functional dependence lattices may be complicated to describe if they are not based on simple causal relations between the variables.

Example 5 *This example concern fruit from a supermarket. Variable* X *tells whether the supermarket will sell it at normal price, or at a reduced price because it is close to the expiration date, or whether it is through out because the expiration date has been exceeded. Variable* Y *describes whether the fruit tastes very fresh, is eatable, or looks disgusting. The variable* Z *tells whether the fruit will make you sick or not. The functional dependences are given by* $Y\to_f Z$ *and* $X\vee Y=X\vee Z$*. The lattice is* N_5*. This is the standard example of a lattice that is not modular.*

Let L be a lattice. We may ask whether one can associate a random variable X_a to the element $a\in L$ in such a way that if we define conditional independence with respect to the distribution of the random variables, then $(X_a\perp X_a\mid X_b)$ if and only if $a\leq b$ in L. This is indeed possible. The *canonical construction* is as follows. For each $a\in L$ we define assign a binay random variable Y_a with a uniform distribuion. We define

$$X_a=\left(\left\{\begin{array}{ll} Y_\ell & \text{for } \ell\leq a\\ 0 & \text{for } \ell\nleq a\end{array}\right.\right)_{\ell\in L}.$$

Then $X_b\to_f X_a$ if and only if $b\geq a$. As any finite lattice is the closure lattice of a power set lattice of functional dependences it is natural to consider super-graphoid relations on lattices of any structure rather than restricting our attention to semi-graphoid relations on Boolean lattices.

Different super-graphoid relations on the same lattice may lead to the same lattice of functionally closed sets.

Example 6 *Let* X *and* Y *denote two random variables and consider two different joint distributions* P *and* Q *on* (X,Y)*. Under* P *the variables* X *and* Y *are independent. Under* Q *they are dependent but in a way so that that* X *is not a function of* Y *and* Y *is not a function of* X*. In both cases the lattice of functionally closed sets of variables is the power set. Under* P *we have* $(X\perp Y)_{I_P}$ *but under* Q *we have* $\neg(X\perp Y)_{I_Q}$*.*

We have seen that an super-graphoid defines a lattice and the there is a canonical way of constructing a super-graphoid relation that induces the ordering in the lattice. Therefore it is natural to ask how properties of the super-graphoid is reflected in the structure of the lattice and vice versa.

Example 7 *Consider three binary variables* $X,Y,$ *and* Z *related by* $Z=X\oplus Y$*. Then the closure lattice has the five elements* $0,X,Y,Z,$ *and* 1*. The lattice is called* M_5 *and is one of the standard example of a lattice that is modular but not distributive. This is the simplest case of a functional dependence structure that might be described as a causal loop. The variables* X *and* Y *are not equal but given* Z *one may perhaps say that* X *influences* Y *and* Y *influences* X*.*

4 Intersection

For a semi-graphoid relation $(\cdot\perp\cdot\mid\cdot)_I$ on a power set the following property for disjoint sets $X,Y,$ and Z is of importance.

Intersection

> $(X\perp Y\mid Z\cup W)_I$ and $(X\perp Z\mid Y\cup W)_I$ implies $(X\perp Y\cup Z\mid W)_I$.

Intersection holds for Bayesian networks and most other graphical models of independence. If the relation $(\cdot\perp\cdot\mid\cdot)_I$ is super-graphoid, and Y and Z are not

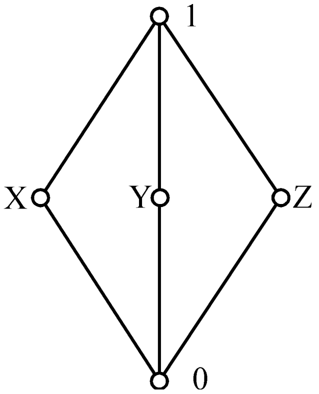

Figure 2: Hasse diagram of the lattce M_5 .

disjoint, then we may write $Y = (Y \backslash Z) \cup (Y \cap Z)$ and $Z = (Z \backslash Y) \cup (Y \cap Z)$. In this case we get that $(X \perp Y \mid Z \cup W)_I$ and $(X \perp Z \mid Y \cup W)_I$ implies $(X \perp Y \cup Z \mid (Y \cap Z) \cup W)_I$. For a lattice $(L, \vee, \wedge)$ it is natural to consider any of the two generalizations.

Strong Lattice Intersection

> $(X \perp Y \mid Z \vee W)_I$ and $(X \perp Z \mid Y \vee W)_I$ implies $(X \perp Y \vee Z \mid (Y \wedge Z) \vee W)_I$.

Weak Lattice Intersection

> $(X \perp Y \mid Z \vee W)_I$ and $(X \perp Z \mid Y \vee W)_I$ implies $(X \perp Y \vee Z \mid (Y \vee W) \wedge (Z \vee W))_I$.

Weak Lattice Intersection is fulfilled for conditional independence of random variables in the following sense.

Theorem 8 *Let X, Y, Z and W denote random variables with a joint distribution P. Assume that $(X \perp Y \mid Z \cup W)_I$ and $(X \perp Z \mid Y \cup W)_I$ under P. Then there exists a random variable V almost surely determined by both $Y \cup W$ and $Z \cup W$ such that $(X \perp Y \cup Z \mid V)$.*

The proof will be given in a longer version of this paper. A lattice is said to be *join-semidistributive* if $X \vee Z = Y \vee Z$ implies that $(X \wedge Y) \vee Z = X \vee Z$ for all X, Y and Z.

Theorem 9 *If the a super-graphoid relation $(\cdot \perp \cdot \mid \cdot)_I$ on the lattice satisfies strong lattice intersection then the lattice is join-semidistributive.*

A lattice said to be *semi-convex* if $X \wedge Y = X \wedge Z$ and $X \vee Z = Y \vee Z$ implies $X \leq Z$.

Proposition 10 *If a lattice is join-semidistributive then it is semi-convex. In particular the a graphical closure lattice is semi-convex.*

Proof. Assume that $X \wedge Y = X \wedge Z$ and $X \vee Z = Y \vee Z$ in a lattice. Then

$$\begin{aligned} X &\leq X \vee Z \\ &= (X \wedge Y) \vee Z \\ &= (X \wedge Z) \vee Z \\ &= Z. \end{aligned}$$

■

We note that the lattice M_5 is not semi-convex, and semi-convexity efficiently rule out the possibility of anything like a causal loop.

5 The graphical closure lattice of a deterministic Bayesian network

A Bayesian network is a set of random variables which are organized in a directed acyclic graph (DAG). The joint distribution of the random variables should satisfy that if variables are *d-separated* in the graph then they should be conditional independent with respect to the probability distribution. In a Bayesian network the joint distribution can be factored into the conditional distribution of each child given its parents. If a variable has no parents we just get a distribution with no conditioning. We may think of such a distribution as a prior distribution.

Definition 11 *A DAG is said to be Austrian if there exist nodes x, y and z such that x and y are parents to z and so that x is an ancestor of y.*

Note that there is a one-to-one correspondence between finite posets and non-Austrian graphs.

Definition 12 *A Bayesian network is* deterministic *if it is non-Austrian and each child with parents is determined by these parents. Variables with no parents are independent and not determined.*

Note that, formally, if a variable with no parents is determined by its parents it would mean that it could only take one value, and then any other variable would also only take one value.

Definition 13 *In a poset a set of variables A graphically determines a set of variables B if any maximal chain through a variable $v \in B$ intersects A in a variable prior to v. When A graphically determines B we write $A \succeq_g B$.*

Theorem 14 *Let A and B be subsets of a deterministic Bayesian network. If $A \succeq_g B$ then $A \rightarrow_f B$.*

Proof. Without loss of generality we may assume that B is a singleton $\{x\}$. The proof is by induction on the length of the longest backward maximal chain from x. Assume that the result holds if all backward maximal chains are shorter than n and that x has a maximal chain of length $n+1$. If $x \in A$ it obviously holds. Therefore, assume that $x \notin A$. We will show that all parents are determined by A and that x is therefore also determined by A. Let y be a parent of x. If $y \in A$ it is determined by A. If $y \notin A$ then all maximal chains through x and y will intersect A because $A \to_g \{x\}$. A maximal chain through x and y does not intersected A at x or at y and it does not intersect between x and y because there are no nodes between x and y in a non-Austrian DAG. Therefore any maximal chain through x and y must intersect A before y. According to the induction hypothesis that implies that y is determined by A. ■

The graphical closure $cl_g(A)$ of a set of variables A is is the largest set B such that $B \preceq_g A$.

Theorem 15 *The ordering* $\succeq_g$ *defines a lattice on the set of graphical closures of a poset.*

Proof. This follows from Theorem 3 by checking that Armstrong's axioms are fulfilled. ■

Theorem 16 *The graphical closure lattice is join-semidistributive.*

Proof. The inequality $(X \wedge Y) \vee Z \preceq_g X \vee Z$ holds in all lattices. Let a be an element in $X \vee Z = Y \vee Z$. Let w be a chain through a. Then there is a smallest element b in w such that $b \in X \vee Z = Y \vee Z$. If $b \in Z$ then obviously $b \in (X \wedge Y) \vee Z$. If $b \notin Z$ then $b \in cl_g(X \cup Z)$ implying $b \in X$ and, similarly, $b \in Y$. Hence, $b \in X \cap Y = X \wedge Y$. ■

Theorem 17 *If* $X \vee Y = A \vee B$ *then* $X \vee Y = (X \wedge A) \vee (X \wedge B) \vee (Y \wedge A) \vee (Y \wedge B)$.

Proof. The inequality $X \vee Y \succeq_g (X \wedge A) \vee (X \wedge B) \vee (Y \wedge A) \vee (Y \wedge B)$ holds for any lattice. Let a be an element in $X \vee Y = A \vee B$. Let w be a chain through a. Then there is a smallest element b in w such that $b \in X \vee Y = A \vee B$. Then b is in X and A, or in X and B, or in Y and A, or in Y and B. Hence, $b \in (X \wedge A) \vee (X \wedge B) \vee (Y \wedge A) \vee (Y \wedge B)$. This holds for all w so we conclude that $a \in (X \wedge A) \vee (X \wedge B) \vee (Y \wedge A) \vee (Y \wedge B)$. ■

Together with the join-semidistributivity this theorem implies that any set in the graphical closure lattice has a unique irreducible decomposition.

Definition 18 *We now recall the definition of modularity. An lattice element A is said to be* right modular *if $X \leq A$ implies*

$$X \vee (Y \wedge A) = (X \vee Y) \wedge A$$

for all Y. An lattice element B is said to be left modular *if $B \leq Z$ implies*

$$B \vee (Y \wedge Z) = (B \vee Y) \wedge Z$$

for all Y. Note that a lattice element A is often called modular if it is right modular. An element X in a lattice L is said to be distributive if

$$X \vee (Y \wedge Z) = (X \vee Y) \wedge (X \vee Z)$$

for all $Y, Z \in L$. It is co-distributive if

$$X \wedge (Y \vee Z) = (X \wedge Y) \vee (X \wedge Z)$$

for all $Y, Z \in L$. An element a is standard if

$$X \wedge (a \vee Z) = (X \wedge a) \vee (X \wedge Z)$$

for all $Y, Z \in L$.Similarly, a is co-standard if

$$X \vee (a \wedge Z) = (X \vee a) \wedge (X \vee Z)$$

for all $Y, Z \in L$. An element X is cancelable if $X \vee Y = X \vee Z$ and $X \wedge Y = X \wedge Z$ implies $Y = Z$ for all $Y, Z \in L$. It can be proved that an element is (co-)standard if and only if it is cancelable and (co-)distributive.

An upset in a DAG is obviously a graphical closure. The set of upsets is a distributive lattice. To get a better understanding of the relation between the upsets in a DAG and the ordering $\succeq_g$ we need the following definition.

Theorem 19 *In a lattice the set of standard elements is a distributive sub-lattice.*

We shall try to determine this sub-lattice.

Theorem 20 *For an element ℓ in the graphical closure lattice of a non-Austrian DAG the following five conditions are equivalent:*

1. *The element ℓ is standard.*
2. *The element ℓ is an upset in the DAG.*
3. *For all $Y, Z \in L$ such that $Y \wedge Z = 0$ we have $(\ell \vee Y) \wedge (\ell \vee Z) = \ell$.*
4. *The element ℓ is right modular.*
5. *The element ℓ is distributive.*

Proof.

2. ⇒ 1. Assume that ℓ is an upset. Then $\ell \vee X = \ell \cup X$ and it is easy to see that ℓ is standard.

2. ⇒ 1. Follows from the definition of standard elements.

5. ⇒ 3. Follows from the definition of distributive elements.

$\neg 2. \Rightarrow \neg 3.$ Assume that ℓ is not an upset. Then there exist a variable Y such that Y is after a variable X in ℓ. Let Z be the set of variables such that Z is not a function of ℓ and such that there exist a maximal chain from Z to Y such that Z is the last element in the chain that is not a function of ℓ. Then $Y \preceq_g \ell \vee Z$. We do not have $Y \preceq_g Z$ because Y has an ancestor in ℓ. Hence

$$Y \preceq_g (\ell \vee Y) \wedge (\ell \vee Z)$$

but

$$Y \not\preceq_g \ell \vee (Y \wedge Z) .$$

$\neg 2. \Rightarrow \neg 4.$ Assume that ℓ is not an upset. Then there exist a variable a such that a is after a variable X in ℓ. Define $Y = a \vee \ell$. Let Z be the set of variables such that Z is not a function of ℓ and such that there exist a maximal chain from Z to a such that Z is the last element in the chain that is not a function of ℓ. Then $a \preceq_g \ell \vee Z$ and therefore also $Y \preceq_g \ell \vee Z$. We do not have $a \preceq_g Z$ because a has an ancestor in ℓ. Hence

$$a \preceq_g (\ell \vee Y) \wedge Z$$

but

$$a \not\preceq_g \ell \vee (Y \wedge Z) .$$

because $Y \wedge Z = 0$.

1. ⇒ 5. Follows from the definitions.

$\neg 2. \Rightarrow \neg 5.$ Assume that ℓ is not an upset. Then there exist a variable a such that a is after a variable X in ℓ. Let Z be the set of variables such that Z is not a function of ℓ and such that there exist a maximal chain from Z to a such that Z is the last element in the chain that is not a function of ℓ. Define $Y = Z \vee a$. Then $X \vee Y = X \vee Z$ and $X \wedge Y = X \wedge Z = 0$. We have $Y \neq Z$ because $Y \succeq_g a$ but $Z \not\succeq_g a$.

5. ⇒ 4. Follows from the definitions.

■

Example 21 *The lattice N_5 is not the graphical closure lattice a non-Austrian DAG. To see this we note that X is left modular but both Y and Z are complements of X, so X is not standard.*

An element ℓ in a lattice is *join-irreducible* if $\ell = a \vee b$ implies $\ell = a$ or $\ell = b$.

Theorem 22 (Birkhoff (1991)) *A distributive lattice L can be identified with upsets in the poset of join-irreducible elements of L.*

Now we can combine these results. We know that a deterministic Bayesian network defines a lattice, and we have shown the left modular elements form a distributive sub-lattice that is equivalent with poset of join-irreducible elements but at the same time it can be identified with upsets in the Bayesian network. Hence, nodes in a Bayesian network can be identified with the join irreducible elements of the sub-lattice of left modular elements in the graphical closure lattice of the graph. Another way to say this is that a deterministic Bayesian network is completely described by its graphical closure lattice. All questions related to deterministic Bayesian networks can therefore be stated in terms of lattices.

A deterministic Bayesian network has the property that each node is determined by its parents. This property can be translated into a property of the graphical closure lattice. Although this is straight forward we shall abstain from doing this in this short note because at the moment we have not yet been able to express the resulting condition on the lattice in a pleasant algebraic form. The relation between the lattice and the following two conditions

Composition $(X \perp Y \mid W)_I$ and $(X \perp Z \mid W)_I$ implies $(X \perp Y \vee Z \mid W)_I$.

Weak transitivity If w is a join-irreducible element of the lattice and $(X \perp Y \mid Z \vee w)_I$ and $\neg (X \perp w \mid Z)_I$ and $\neg (w \perp Y \mid Z)_I$, then $(X \perp Y \mid Z)_I$.

are also not clarified yet although they can be translated into lattice language.

Theorem 23 *For an element ℓ in the graphical closure lattice of a non-Austrian DAG the following two conditions are equivalent:*

1. *The element ℓ is right modular.*
2. *The element ℓ is a closed downset in the DAG.*
3. *The element is co-distributive.*
4. *The element is co-standard.*

Proof.

1. ⇒ 2. Assume that ℓ is not an downset. Then there exist a variable a such that a is after a variable $X \preceq_g \ell$. Define $Y = a \vee \ell$. Let Y be the set of

variables such that Y is not a function of ℓ and such that there exist a maximal chain from Y to a such that Y is the last element in the chain that is not a function of ℓ. Then $a \preceq_g X \vee Y$. We do not have $a \preceq_g Y$ because a has an ancestor in ℓ. Hence

$$a \preceq_g (X \vee Y) \wedge \ell$$

but

$$a \npreceq_g X \vee (Y \wedge \ell) .$$

because $Y \wedge \ell = 0$.

2. $\Rightarrow$ 1. Let $X \preceq_g \ell$ and let a be an element in $(X \vee Y) \wedge \ell$. We have to show that a is element in $X \vee (Y \wedge \ell)$. Let w be a maximal chain. We have to prove that w intersects X or that w intersects $Y \wedge \ell$. Assume that w does not intersect X. Then w intersects Y but the whole backward part of w is within ℓ so the intersection of w and Y must be member of $Y \wedge \ell$.

The rest of the proof is strict forward. ■

Theorem 24 *The set of right modular elements of the graphical closure lattice of a non Austrian DAG is a Boolean lattice. In this lattice we have* $A \perp B \mid C$ *if and only if* $C \to A \wedge B$.

Proof. The right modular elements can be identified with closed downsets. If A is a set of initial nodes in the DAG then $cl_g(A)$ is a closed downset. Conversely let A be the initial elements in a downset ℓ. Then $\ell = cl_g(A)$. Hence, downsets can be identified with subsets of the set of initial variables and these form a Boolean lattice. ■

The right modular elements do not really tell anything about the functional dependence structure of the deterministic Bayesian network. They only identify the initial elements that all other variables are functions of. For an element X in a lattice $a(X)$ will denote the smallest co-standard element that determines X.

Theorem 25 *For a deterministic Bayesian network the following two conditions hold.*

1. *If* $(X \perp Y \mid Z)_I$ *and* $a(X \vee Y \vee Z) \to W$ *then* $(X \perp Y \mid Z \vee W)_I$.
2. *If* $(X \perp Y \mid Z \vee W)_I$ *and* $W \wedge a(X \vee Y \vee Z) = 0$ *then* $(X \perp Y \mid Z)_I$.

6 Hidden variables

For a deterministic Bayesian network one may restrict the attention to some of the random variables and consider the other variables as latent or hidden. On the other hand any Bayesian network can be embedded in a deterministic Bayesian network by introducing independent noise variables to all variables that are not determined by their parents. In general the network will not be deterministic anymore if some of the variables are latent. Also the ordering determined by left-modular elements in the functional dependence lattice will be sensitive to latency of variables. Therefore we need some concepts that are more robust to latency of variables than modularity.

Definition 26 *Let* $(M, \preceq)$ *be a poset with a supergraphoid relation* $(\cdot \perp \cdot \mid \cdot)_I$ *on the lattice of subsets. If* $X \subseteq M$ *then* $a_{\preceq}(X)$ *is the set of ancestors to elements in* X*. The relation* $\preceq$ *is said to be harmonic if the following two conditions are fulfilled:*

1. *If* $(X \perp Y \mid Z)_I$ *and* $a_{\preceq}(X \vee Y \vee Z) \to W$ *then* $(X \perp Y \mid Z \vee W)_I$.
2. *If* $(X \perp Y \mid Z \vee W)_I$ *and* $W \wedge a_{\preceq}(X \vee Y \vee Z) = 0$ *then* $(X \perp Y \mid Z)_I$.

For a Bayesian network or a hybrid graph the ordering in the graph defines a harmonic relation. It is tempting to use the word *causal* instead of harmonic, and the only reason this is not done here is that the word causal is already overloaded. The word harmonic is just a neutral temporary working terminology. We note that for a deterministic Bayesian network $a_{\preceq}(X)$ is simply the smallest right modular element in the lattice that contains X. If an ordering $\preceq$ is harmonic then it is also harmonic on $(\cdot \perp \cdot \mid \cdot)_I$ restricted to a subset of A. Therefore harmonic ordering are our candidates to the ordering of the variables when they are embedded in a deterministic Bayesian network.

We shall use the concept of locality to describe the deviation of an partially ordered set of variables from being a deterministic Bayesian network. The word locality is taken from the physics literature about quantum non-locallity.

Definition 27 *Let* $(M, \preceq)$ *be a poset with a supergraphoid relation* $(\cdot \perp \cdot \mid \cdot)_I$ *on the lattice of subsets. We assume that* $\preceq$ *is harmonic.*

Deterministic Locality *If any chains through* B *intersects* C *before* B*, then* B *is a function of* C.

Temporal Locality *If* $a \preceq b$ *and all chains from* a *to* b *intersects* C*, then* $(a \perp b \mid C)_I$.

Spatial Locality

If $X, Y \subseteq M$ *then* $(X \perp Y \mid a_{\preceq}(X) \wedge a_{\preceq}(Y))_I$.

Pearl & Verma (1990) introduced graphs with bidirected arrows and called them *hybrid graphs*. Independence in hybrid graphs is defined via the *d*-separation criterion. See Pearl & Verma (1990) for more details about this class of graphical models.

Theorem 28 *Let* $(A, \preceq)$ *be a poset with a super-graphoid relation* $(\cdot \perp \cdot \mid \cdot)_I$ *on the lattice of subsets. If* $\preceq$ *is harmonic and the super-graphoid relation satisfies the axioms of intersection, composition, and weak transitivity then there exists a hybrid graph with ordering* $\preceq$ *and independence structure* $(\cdot \perp \cdot \mid \cdot)_I$ *. If spatial locality is satisfied then a Bayesian network exists with ordering* $\preceq$ *and independence structure* $(\cdot \perp \cdot \mid \cdot)_I$ *. If also temporal locality is satisfied then the Bayesian network can be realized by a non-Austrian graph.*

The proof is too long to be given in this short paper.

7 Discussion

The determination of functional dependencies is an important part of designing databases in the relational model, and in database normalization and denormalization. The functional dependencies, along with the attribute domains, are selected so as to generate constraints that would exclude as much data inappropriate to the user domain from the system as possible.

There are various reasons to study functional dependences in databases. The two most important are:

Compression One can make a more compact representation of data when functional dependences are known. How to make such compact representations is the subject of the theory of relational databases.

Control If one has direct control over some of the attributes/variables in a database then the functional dependences will tell what the effect will be on other attributes. If for instance one has a database of all members of a club and want to send a letter to all members in a specific area, one may do this by sending to all members for which part of their zip code has a certain value. The relation between zip code, town and areas will tell how this can be done most efficiently.

We note that these two reasons are also the main reasons to study causality.

Some people claim that causation is about overruling, i.e. external intervention in an otherwise probabilistic structure. One problem with this approach is that this excludes the intervention it self to be discussed using the notion of causality. One way around this problem is the distinguish between controlled and uncontrolled variables in the Bayesian network. If we believe that probabilities are a way of quantifying our uncertainty then it does not make much sense to assign probabilities to our own decisions. Therefore there is a conceptual problem in assigning probabilities to controlled variables. The theory presented here is based on functional dependence lattices so we do not have to make any artificial distinction between controlled or uncontrolled variables.

Bayesian networks are based on graphs that are in principle discrete objects. Therefore Bayesian networks will never be able to describe infinite sets of variables. For instance an electromagnetic field can be described by a vector in each point in space and time. Maxwell's equations describe functional dependences between all these variables. In particular an ancestor set $a(X)$ can be identified with a backward light cone from X. With infinitely many variables one has to take topological considerations into account, and for this lattices is an efficient tool because, formally a topology is a lattice.

Finally we shall note that we do not claim that functional dependences can or should be used to model complicated random phenomena like quantum non-causallity, but we claim that this approach is useful in describing how, when, and why causal models can be used.

References

Armstrong, W. W. 1974. Dependence structures of Database Relationships. *Pages 580–583 of: IFIP World Congress 1974*. North-Holland.

Birkhoff, Garrett. 1991. *Lattice Theory*. Colloquium Publications, vol. 25. American Mathematical Society. Reprint.

Gardarin, G., & Valduriez, P. 1989. *Relational Databases and Knowledge Bases*. Addison-Wesley.

Grätzer, G. 2003. *General Lattice Theory*. Second edition edn. Birkhäuser.

Lauritzen, S. L. 1998. *Graphical Models*. Oxford Statistical Science Series, vol. 17. Oxford Science Publications.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. San Mateo, Califonia: Morgan Kaufmann Publ.

Pearl, J. 2000. *Causality*. Cambridge University Press.

Pearl, J., & Verma, T. S. 1990 (27-29 July). Equivalence and Synthesis of Causal Models. *In: Proceedings 6th Conference on Uncertainty in AI, Cambridge. Mass.*