Entropy Inequalities for Lattices

Peter Harremoës †

Copenhagen Business College, Nørre Voldgade 34, 1358 Copenhagen K, Denmark; harremoes@ieee.org;
Tel.: +45-39-56-41-71
† Current address: Rønne Alle 1, st., 2860 Søborg, Denmark.

Abstract: We study entropy inequalities for variables that are related by functional dependencies. Although the powerset on four variables is the smallest Boolean lattice with non-Shannon inequalities, there exist lattices with many more variables where the Shannon inequalities are sufficient. We search for conditions that exclude the existence of non-Shannon inequalities. The existence of non-Shannon inequalities is related to the question of whether a lattice is isomorphic to a lattice of subgroups of a group. In order to formulate and prove the results, one has to bridge lattice theory, group theory, the theory of functional dependences and the theory of conditional independence. It is demonstrated that the Shannon inequalities are sufficient for planar modular lattices. The proof applies a gluing technique that uses that if the Shannon inequalities are sufficient for the pieces, then they are also sufficient for the whole lattice. It is conjectured that the Shannon inequalities are sufficient if and only if the lattice does not contain a special lattice as a sub-semilattice.

1. Introduction

The existence of non-Shannon inequalities has received much attention since the first inequality of this type was discovered by Zhang and Yeung [1]. The basic observation is that any four random variables X, Y, Z and W satisfy the following inequality:

$$2I(Z;W) \leq I(X;Y) + I(X;Z \uplus W) + 3I(Z;W \mid X) + I(Z;W \mid Y). \tag{1}$$

Here, $C \uplus D$ denotes the random variable that takes a value of the form (c, d) if $c = C$ and $d = D$. As usual, $I(\cdot;\cdot)$ and $I(\cdot;\cdot \mid \cdot)$ denote mutual information and conditional mutual information given by:

$$I(X;Y) = H(X) + H(Y) - H(X \uplus Y), \tag{2}$$

$$I(X;Y \mid Z) = H(X \uplus Z) + H(Y \uplus Z) - H(X \uplus Y \uplus Z) - H(Z). \tag{3}$$

where H denotes the Shannon entropy. The inequality (1) is non-Shannon in the sense that it cannot be deduced from the positivity, monotonicity and submodularity of the entropy function on the variables X, Y, Z, and their joins, i.e., satisfaction of the following inequalities:

Positivity	$H(X) \geq 0,$	(4)
Monotonicity	$H(X \uplus Y) \geq H(X),$	(5)
Submodularity	$H(X \uplus Z) + H(Y \uplus Z) \geq H(X \uplus Y \uplus Z) + H(Z).$	(6)

Positivity and monotonicity were recognized by Shannon [2], while submodularity was first observed by McGill [3]. It is easy to show that any inequality involving only three variables rather than four can be deduced from Shannon's inequalities [4]. The powerset of four variables is a Boolean algebra with 16 elements, and any smaller Boolean algebra corresponds to a smaller number of variables, so in a trivial sense, the Boolean algebra with 16 elements is the smallest Boolean algebra with non-Shannon inequalities.

In the literature on non-Shannon inequalities, all inequalities are expressed in terms of sets of variables and their joins. Another way to formulate this is that the inequalities are stated for the free ∪-semi-lattice generated by a finite number of variables. In this paper, we will also consider intersections of sets of variables. We note that for sets of variables, we have the inequality:

$$I\left(X;Y\mid Z\right)\geq H\left(X\cap Y\mid Z\right). \tag{7}$$

Inequality (7) has even inspired some authors to use $I\left(\cdot\wedge\cdot\right)$ as notation for mutual information.

Although non-Shannon inequalities have been known for two decades, they have found remarkably few applications compared with the Shannon inequalities. One of the reasons is that there exists much larger lattices than the Boolean algebra with 16 elements for which the Shannon inequalities are sufficient. The simplest examples are the Markov chains:

$$X_1 \to X_2 \to X_3 \to \cdots \to X_n \tag{8}$$

where any variable X_j is determined by its predecessor, i.e., the conditional entropies $H\left(X_{j+1}\mid X_j\right)$ are zero for $j = 1, 2, \ldots, n-1$. For such a chain, one has:

$$H\left(X_1\right)\geq H\left(X_2\right)\geq H\left(X_3\right)\geq \cdots \geq H\left(X_n\right)\geq 0. \tag{9}$$

The inequalities (9) are all instances of the entropy function being monotone, and it is quite clear that these inequalities are sufficient in the sense that for any sequence of values that satisfies these inequalities, there exists random variables related by a deterministic Markov chain with these values as entropies.

In this paper, we look at entropy inequalities for random variables that are related by functional dependencies. Functional dependencies give a partial ordering of sets of variables into a lattice. Such functional dependence lattices have many applications in information theory, but in this paper, we will focus on determining whether a lattice of functionally-related variables can have non-Shannon inequalities. In order to achieve interesting results, we have to restrict our attention to special classes of lattices.

Entropy inequalities have been studied using matroid theory, but finite matroids are given by geometric lattices, i.e., atomistic semi-modular lattices (see the textbook of Stern [5] for definitions). For the study of non-Shannon inequalities, it is more natural to look at general lattices rather than geometric lattices because many important applications involve lattices that are not atomistic or not semi-modular. For instance, a deterministic Markov chain gives a lattice that is not atomistic. It is known that a function is entropic if and only if it is (approximately) equal to the logarithm of the index of a subgroup in a group [6]. Therefore, it is natural to study entropic functions on lattices and their relations to subgroup lattices.

In this paper, we bridge lattice theory, database theory and the theory of conditional independence, but sometimes, the terminology in these fields does not match. In such cases, we give preference to lattice theory over database theory and preference to database theory over the theory of conditional independence. For instance, there is a property for closure operators that is called extensivity in the theory of lattices. We translate extensivity into a property for functional dependence, and it turns out that extensivity can be used instead of the property for functional dependences, which is called augmentation. Extensivity is apparently a weaker condition than augmentation, but together with

the properties called monotonicity and transitivity, they are equivalent on finite lattices. Finally, we translate extensivity from functional dependencies to separoid relations that model the concept of conditional independence. In the literature on conditional independence, extensivity has been termed "normality" without any explanation why this term is used. We called it extensivity because it is equivalent to the notion of extensivity in lattice theory, which we consider as a more fundamental theory.

The paper is organized as follows. In Section 2, we describe the link between lattice theory and the theory of functional dependences in detail. We demonstrate how properties of closure operators associated with sub-semilattices correspond to the properties of functional dependence that are normally called Armstrong's axioms. In Section 3, we describe positive monotone submodular functions (polymatroid functions) and how they lead to separoid relations on lattices. These separoid relations generalize the notion of conditional independence known from Bayesian networks and similar graphical models. We demonstrate how properties of separoid relations correspond to properties of functional dependences.

In Section 4, we describe entropy functions on lattices and how they correspond to subgroup lattices of a group. We conjecture that the Shannon inequalities are sufficient for describing entropic polymatroid functions of a lattice if and only if the lattice does not contain a special lattice as a sub-semilattice. In Section 5, we develop some technical results related to "gluing" lattices together. The gluing technique is very useful for planar lattices, and in Section 6, we demonstrate that entropic functions on planar modular lattices can be described by Shannon's inequalities.

We finish with a short discussion, where we outline some future research directions. There is one appendix with some additional comments related to Armstrong's axioms. These are mainly intended for readers that are familiar with the theory of functional dependencies in databases. A second appendix contains a long list of lattices that are used to document that polymatroid functions on lattices with seven or fewer elements can be described by Shannon's inequalities.

Some of the results presented in this paper have been published in preliminary form and without proof [7,8], but since then, most of the results have now been strengthened or reformulated. In this paper, all proof details will be given.

2. Lattices of Functional Dependence

In this section, we shall briefly describe functional dependencies and their relation to lattice theory. The relation between functional dependence and lattices has been studied [7,9–13]. The relation between lattices and functional dependencies is closely related to minimal sets of Shannon-type inequalities [14,15]. Relations between functional dependencies and Bayesian networks have also been described [8,16]. Many problems in information theory and cryptography can be formulated in terms of functional dependencies.

Example 1. *Consider a group consisting of n agents. One might be interested in giving each agent in the group part of a password in such a way that no single agent can recover the whole password, but any two agents are able to recover the password. Here, the password should be a function of the variables known by any two agents, but must not be a function of a variable held by any single agent. The functional dependence structure is the lattice illustrated in the Hasse diagram in Figure 1. The node at the top illustrates the password. Each of the intermediate nodes represents the knowledge of an agent. The bottom node represents no knowledge.*

A $\wedge$-semilattice is a set equipped with a binary operator $\wedge$ that satisfies the following properties:

Commutativity	$X \wedge Y = Y \wedge X$,	(10)
Associativity	$(X \wedge Y) \wedge Z = X \wedge (Y \wedge Z)$,	(11)
Idempotency	$X \wedge X = X$.	(12)

For a $\wedge$-semilattice the relation $X \wedge Y = X$ defines a preordering that we will denote $X \leq Y$. If $(\mathcal{L}, \wedge)$ is a semilattice, then we say that $\mathcal{M}$ is sub-semilattice if $\mathcal{M}$ is closed under the $\wedge$ operation. Let $(\mathcal{L}, \wedge)$ denote a semilattice. Let $\downarrow X = \{Y \in \mathcal{L} \mid Y \leq X\}$. Then, $\downarrow (X \wedge Y) = (\downarrow X) \cap (\downarrow Y)$. Therefore, we can identify any finite semilattice with a $\cap$-semilattice in a powerset. Since we will usually identify semilattice elements with sets of variables, we will often use $\subseteq$ and $\cap$ to denote the ordering and the meet operation.
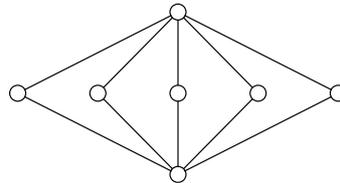
Figure 1. Hasse diagram of the lattice M_n for $n = 5$.

In this paper, we will assume that all semilattices and all lattices are finite. If a $\cap$-semilattice $(\mathcal{L}, \cap)$ has a maximal element, then a binary operator $\vee$ can be defined as:

$$X \vee Y = \bigcap_{\substack{Z \supseteq X \\ Z \supseteq Y}} Z \tag{13}$$

and then, $(\mathcal{L}, \cap, \vee)$ is a lattice.

Let $(\mathcal{L}, \subseteq)$ denote a lattice with $\mathcal{M}$ as a sub-semilattice with the same maximal element as $\mathcal{L}$. Then, a unary operator $cl : \mathcal{L} \to \mathcal{L}$ can be defined by:

$$cl(X) = \bigcap_{\substack{Z \supseteq X \\ Z \in \mathcal{M}}} Z \tag{14}$$

The operator cl is a closure operator [17], i.e., it satisfies:

Extensivity	$X \subseteq cl(X)$,	(15)
Monotonicity	$X \subseteq Y$ implies $cl(X) \subseteq cl(Y)$,	(16)
Idempotency	$cl(cl(X)) = cl(X)$.	(17)

For any closure operator cl, the element X is said to be closed if $cl(X) = X$. If X and Y are closed, then $X \cap Y$ is closed ([18], [Lemma 28]), so the closed elements of a lattice under a closure operator form a $\cap$-semilattice.

Proposition 1. *Let $(\mathcal{L}, \subseteq)$ denote a finite lattice. Assume that a subset $\mathcal{M}$ of $\mathcal{L}$ is closed under the meet operation and has the same maximal element as $\mathcal{L}$. Then, $\mathcal{M}$ is a lattice under the ordering $\subseteq$ with the meet operation in $\mathcal{M}$ given by $\cap$ and join operation in $\mathcal{M}$ given by $X \uplus Y = cl(X \cup Y)$.*

Example 2. *If G is a group, then a subgroup is defined as a subset that is closed under the group operations. The closure of a subset of G is the subgroup generated by the subset. The lattice of subgroups forms a $\cap$-semilattice in the lattice of all subsets of the group. Let G denote a finite group. For any subgroup $\tilde{G} \subseteq G$, we associate the variable $X_{\tilde{G}}$ that maps an element $g \in G$ into the left coset $g\tilde{G}$. Then, the subgroup lattice of G is mapped into a lattice of variables where the subset ordering of subgroups is equivalent to functional dependences between the corresponding variables.*

Proposition 2. *If cl is a closure operator on a lattice, then the relation $cl(X) \supseteq Y$ and the relation $cl(X) \supseteq cl(Y)$ are equivalent. The relation $X \to Y$ given by $cl(X) \supseteq Y$ satisfies the following properties.*

Extensivity	$X \to Y$ implies $X \to X \uplus Y$,	(18)
Monotonicity	$X \supseteq Y$ implies $X \to Y$,	(19)
Transitivity	If $X \to Y$ and $Y \to Z$, then $X \to Z$.	(20)

Remark 1. *The monotonicity of $\to$ is called reflexivity in the literature on databases. We reserve the notion of reflexivity to the relation $X \to X$, in accordance with the terminology for ordered sets. In database theory, the property $X \to X$ is called self determination.*

In the literature on databases extensivity, (18) is replaced by an apparently stronger property called augmentation, but in a finite lattice augmentation can be proven from extensivity, monotonicity and transitivity. See Appendix A for details.

If the properties (18)–(20) are satisfied, we say that the relation $\to$ satisfies Armstrong's axioms [19]

Proof. Assume that $cl(X) \supseteq cl(Y)$. Using extensivity (15), we get $cl(Y) \supseteq Y$. The transitivity of $\supseteq$ then gives $cl(X) \supseteq Y$.

Assume $cl(X) \supseteq Y$. Then, the monotonicity (16) gives $cl(cl(X)) \supseteq cl(Y)$, and the idempotent gives $cl(X) \supseteq cl(Y)$.

To prove the extensivity (18) of $\to$, assume that $cl(X) \supseteq Y$. Using the extensivity (15), we also get $cl(X) \supseteq X$. Combining these two inequalities gives $cl(X) \supseteq X \uplus Y$, as desired.

The monotonicity (19) of $\to$ follows directly from the monotonicity (15) of cl.

The transitivity (20) of $\to$ follows from the transitivity of $\supseteq$. □

If $\mathcal{L}$ is a lattice with a relation $\to$ that satisfies Armstrong's axioms, then we say that a lattice element X is $\to$ closed if $X \to Y$ implies that $X \supseteq Y$.

Theorem 1. *Let $\mathcal{L}$ be finite lattice with a relation $\to$ that satisfies Armstrong's axioms. Then, the set of $\to$ closed elements form a $\cap$-semilattice with the same maximal element as $\mathcal{L}$. The relation $X \to Y$ holds if and only if $cl(X) \supseteq Y$, where cl denotes the closure operator with respect to the semilattice.*

Proof. Assume that X_1 and X_2 are closed and that $X_1 \cap X_2 \to Y$. The monotonicity (19) implies $X_i \to X_1 \cap X_2$, and then, the transitivity (20) implies that $X_i \to Y$. Since X_i is closed, we have $X_i \supseteq Y$. Since this holds for both $i = 1$ and $i = 2$, we have $X_1 \cap X_2 \supseteq Y$, implying that $X_1 \cap X_2$ is closed. The monotonicity (19) also implies that the maximal element of $\mathcal{L}$ is closed so that the set of closed elements $\mathcal{M}$ forms a $\cap$-semilattice with a closure operator $cl_{\mathcal{M}}$.

Let cl denote the closure with respect to $\mathcal{M}$. We will prove that $X \to cl(X)$. Let $X_1 = X$. Assume that X_1 is not $\to$ closed. The,n there exists Y_1 such that $X_1 \to Y_1$ and $X_1 \not\supseteq Y_1$. Using the extensivity (18), we get $X_1 \to X_1 \uplus Y_2$. Define $X_2 = X_1 \uplus Y_1$. Then, $X_1 \to X_2$ and $X_1 \subset X_2$. Iterate this construction so that:

$$X_1 \to X_2 \to \cdots \to X_n, \tag{21}$$

$$X_1 \subset X_2 \subset \cdots \subset X_n. \tag{22}$$

Since the lattice is finite, the construction must terminate, and when it terminates, X_n is closed. Using transitivity, we get $X \to X_n$ and $X \subseteq X_n$. Since $cl(X)$ is the smallest closed element greater than X, we have $X \to cl(X)$.

If $cl(X) \supseteq Y$, then $cl(X) \to Y$ by monotonicity (19), and then, $X \to Y$ by transitivity (20). If $X \to Y$, then $cl(X) \to Y$. Using that $cl(X)$ is $\to$ closed, we get $cl(X) \supseteq Y$. □

We will look at functional dependencies in databases. Assume that a set of records is labeled by elements in a set A. In statistics records are the individual elements of a sample. For each record

$a \in A$, the database contains the values of various attributes given by a number of functions from A to the set of possible attributes. Sets of such functions will be denoted by capital letters, and these will be our variables. We say that X determines Y and write $X \to Y$ if there exists some function f such that $Y(a) = f(X(a))$ for any record $a \in A$. Then, the relation $\to$ satisfies Armstrong's axioms. Armstrong proved that these axioms form a complete set of inference rules [19]. That means that if a set A of functional dependencies is given and a certain functional dependence $X \to Y$ holds in any database where all the functional dependencies in A hold, then $X \to Y$ holds in that database. Therefore, for any functional dependence $X \to Y$ that cannot be deduced using Armstrong's axioms, there exists a database where the functional dependence is violated [20,21]. As a consequence, there exists a database where a functional dependence holds if and only if it can be deduced from Armstrong's axioms. Using the result that Armstrong's axioms are equivalent to the closed sets forming a lattice, Armstrong's result is easy to prove.

Theorem 2. *For any finite lattice $\mathcal{L}$, there exists a database with a set of related variables such that the elements of the lattice corresponds to closed sets under functional dependence.*

Proof. As the set of records, we take the elements of the lattice $\mathcal{L}$. With each $Y \in \mathcal{L}$, we associate a function $f_Y : \mathcal{L} \to \mathcal{L}$ given by $f_Y(X) = Y \cap X$. If $Y_1 \supseteq Y_2$, then:

$$
\begin{aligned}
f_{Y_2}(X) &= Y_2 \cap X \\
&= Y_2 \cap (Y_1 \cap X) \\
&= f_{Y_2}\left(f_{Y_1}(X)\right)
\end{aligned}
\tag{23}
$$

so that $f_{Y_2} = f_{Y_2} \circ f_{Y_1}$. Therefore, $f_{Y_1} \to f_{Y_2}$.

Assume that $f_{Y_1} \to f_{Y_2}$. Let $X_1 = Y_1$ and $X_2 = Y_1 \uplus Y_2$. Then, $f_{Y_1}(X_1) = f_{Y_1}(X_2) = Y_1$, while $f_{Y_2}(X_1) = Y_1 \cap Y_2$ and $f_{Y_2}(X_2) = Y_2$. Using that $f_{Y_1} \to f_{Y_2}$, we get $Y_1 \cap Y_2 = Y_2$, so that $Y_1 \supseteq Y_2$. □

We have seen that for a subgroup lattice of a group, there exists a lattice of functional dependence. The opposite is also true. To each database with attributes related by functional dependence, there is a group. The construction is as follows. Let A denote a set of records. Let $G = Sym(A)$ be the symmetric group consisting of permutations of the records. If X is a function on A, then we define the stabilizer group G_X as the set of permutations that leave X invariant, i.e., permutations $\pi \in Sym(A)$ such that $X(\pi(a)) = X(a)$ for all $a \in A$. Then, $X \to Y$ if and only if $G_X \subseteq G_Y$. In this way, the functional dependence lattice of a database can be mapped into a lattice of subgroups of a group.

Combining Theorem 2 with the stabilizers subgroups of the symmetric group of a database, we get the following result that was first proven in 1946 by Whitman [22].

Corollary 1. *Any finite lattice can be represented as a functional dependence lattice generated by subgroups of a group.*

3. Polymatroid Functions and Separoids

Definition 1. *On a lattice, the submodularity of a function h is defined via the inequality $h(X) + h(Y) \geq h(X \uplus Y) + h(X \cap Y)$. If the submodular inequality holds with equality, we say that the function is modular. A polymatroid function on a lattice is a function that is non-negative, increasing and sub-modular.*

Example 3. *Let $\mathcal{L}$ be finite atomistic lattice with a ranking function $r : \mathcal{L} \to \mathbb{R}$. Then, L is a geometric lattice if and only if the function r is polymatroid ([5], Corollary 1.9.10).*

For a polymatroid function h on a lattice, one may introduce a function $I_h\left(\cdot;\cdot\mid\cdot\right)$ that corresponds to conditional mutual information by:

$$I_h\left(X;Y\mid Z\right)=h\left(X\uplus Z\right)+h\left(Y\uplus Z\right)-h\left(X\uplus Y\uplus Z\right)-h\left(Z\right). \tag{24}$$

One can rewrite $I_h\left(\cdot;\cdot\mid\cdot\right)$ as:

$$\begin{aligned}I_h\left(X;Y\mid Z\right)=h\left(X\uplus Z\right)+h\left(Y\uplus Z\right)-h\left(X\uplus Y\uplus Z\right)-h\left(\left(X\uplus Z\right)\cap\left(Y\uplus Z\right)\right)\\+h\left(\left(X\uplus Z\right)\cap\left(Y\uplus Z\right)\right)-h\left(Z\right).\end{aligned} \tag{25}$$

Since h is monotone and submodular, we have:

Positivity $$I_h\left(X;Y\mid Z\right)\geq 0. \tag{26}$$

It is straightforward to verify that:

Symmetry $$I_h\left(X;Y\mid Z\right)=I_h\left(Y;X\mid Z\right), \tag{27}$$
Chain rule $$I_h\left(X;Y\uplus Z\mid W\right)=I_h\left(X;Y\mid W\right)+I_h\left(X;Z\mid Y\uplus W\right). \tag{28}$$

We will say that a function $I(\cdot;\cdot\mid\cdot)$ that satisfies positivity (26), symmetry (27) and the chain rule (28) is a separoid function.

Proposition 3. *If $I\left(\cdot,\cdot\mid\cdot\right)$ is a separoid function, then the following property is satisfied.*

Monotonicity $$Y\subseteq Z\text{ implies }I\left(X;Y\mid Z\right)=0. \tag{29}$$

Proof. Assume that $Y\subseteq Z$. We can use the chain rule (28) to get:

$$\begin{aligned}I\left(X;Y\mid Z\right)&=I\left(X;Y\uplus Y\mid Z\right)\\&=I\left(X;Y\mid Z\right)+I\left(X;Y\mid Y\uplus Z\right)\\&=2\cdot I\left(X;Y\mid Z\right).\end{aligned} \tag{30}$$

Hence, monotonicity (29) is satisfied. □

The relation $I_h\left(X;X\mid Z\right)=0$ is equivalent to $h\left(X\uplus Z\right)=h\left(Z\right)$, and this relation will be denoted $X\to_h Z$. The first to observe that to_h defines a lattices was Shannon, who published a very short paper on this topic in 1953 [23]. Shannon did not mention the relation to the theory of functional dependences because that theory was only developed two decades later. Surprisingly, Shannon's paper was only cited once until 2002!

The relation $\to_h$ satisfies Armstrong's axioms, and the most instructive way to see this is via separoid relations. If h is a polymatroid function, then the relation $I_h\left(X,Y\mid Z\right)=0$ will be denoted $X\perp\!\!\!\perp_h Y\mid Z$. Following Dawid et al. [24,25], we say that a relation $(\cdot\perp\!\!\!\perp\cdot\mid\cdot)$ on a lattice $(\mathcal{L},\cap,\uplus)$ is a separoid relation, if it has the following properties:

Monotonicity $\quad\quad\quad Y\subseteq Z\text{ implies }X\perp\!\!\!\perp Y\mid Z,$ $\tag{31}$
Symmetry $\quad\quad\quad X\perp\!\!\!\perp Y\mid Z\text{ implies }Y\perp\!\!\!\perp X\mid Z,$ $\tag{32}$
Chain rule $\quad\quad\quad X\perp\!\!\!\perp Y\uplus Z\mid W,\text{ if and only if }X\perp\!\!\!\perp Y\mid W\text{ and }X\perp\!\!\!\perp Z\mid Y\uplus W.$ $\tag{33}$

Remark 2. *The term monotonicity was used for a different concept by Paolini [26]. In [24,25], a weaker condition than monotonicity was used, but their condition together with the chain rule implies monotonicity.*

With this definition we see that $\perp\!\!\!\perp_h$ is a separoid relation. The properties (31)–(33) should hold for all $X, Y, Z, W \in \mathcal{L}$. In this paper, we are particularly interested in the case where the subsets are not disjoint. In the literature on Bayesian networks and similar graphical models, the focus has been on disjoint sets where only the last two properties (32) and (33) are used to define a semi-graphoid relation [27]. See also [28], Remark 2.5, where it is noted that semi-graphoid relations can be defined on join semi-lattices.

A long list of properties for the notion of independence was given by Paolini [26], but Studený has proven that one cannot deduce all properties of statistical conditional independence from a finite list of axioms [28,29].

Proposition 4. *A separoid relation* $(\cdot \perp\!\!\!\perp \cdot \mid \cdot)$ *on a lattice satisfies the following properties.*

Extensivity	$X \perp\!\!\!\perp Y \mid Z$ implies $X \perp\!\!\!\perp Y \uplus Z \mid Z$,	(34)
Transitivity	If $X \perp\!\!\!\perp Y \mid W$ and $X \perp\!\!\!\perp Z \mid Y \uplus W$, then $X \perp\!\!\!\perp Z \mid W$.	(35)

Remark 3. *Property* (34)*, which we call extensivity, was called normality by Paolini* [26]*.*

Proof. To prove the extensivity (34), assume that $X \perp\!\!\!\perp Y \mid Z$, which is equivalent to $X \perp\!\!\!\perp Y \mid Z \uplus Z$. The monotonicity (31) gives $X \perp\!\!\!\perp Z \mid Z$. The conclusion $X \perp\!\!\!\perp Y \uplus Z \mid Z$ is obtained by the chain rule (33).

To prove the transitivity (35), assume that $X \perp\!\!\!\perp Y \mid W$ and $X \perp\!\!\!\perp Z \mid Y \uplus W$. The chain rule (33) applied twice gives $X \perp\!\!\!\perp Y \uplus Z \mid W$ and $X \perp\!\!\!\perp Z \mid W$. □

In a set of random variables, we note that if Y is independent of Y given X, then Y is a function of X almost surely. If $Y \perp\!\!\!\perp Y \mid X$, we write $X \to_{\perp\!\!\!\perp} Y$.

Theorem 3. *If* $(\mathcal{L}, \cap, \uplus)$ *is a lattice with a separoid relation* $(\cdot \perp\!\!\!\perp \cdot \mid \cdot)$*, then the relation* $\to_{\perp\!\!\!\perp}$ *satisfies Armstrong's axioms. The relation* $(\cdot \perp\!\!\!\perp \cdot \mid \cdot)$ *restricted to the lattice of closed lattice elements is separoid.*

Proof. The extensivity (18) of $\to_{\perp\!\!\!\perp}$ follows directly from the extensivity (34) of $\perp\!\!\!\perp$.

The monotonicity (19) follows directly from the monotonicity (31).

To prove the transitivity of $\to_{\perp\!\!\!\perp}$, assume that $X \to_{\perp\!\!\!\perp} Y$ and $Y \to_{\perp\!\!\!\perp} Z$. The monotonicity (31) implies that $Z \perp\!\!\!\perp X \mid Z \uplus Y$, which by the chain rule (33), implies $Z \perp\!\!\!\perp Z \uplus X \mid Y$. By the chain rule (33), we have $Z \perp\!\!\!\perp Z \mid Y \uplus X$. The monotonicity (31) also gives $Z \perp\!\!\!\perp Y \mid Y \uplus X$, which together with $X \to_{\perp\!\!\!\perp} Y$ implies that $Z \perp\!\!\!\perp Y \mid X$ by transitivity (35). The transitivity (35) then implies $Z \perp\!\!\!\perp Z \mid X$.

To prove that the relation $(\cdot \perp\!\!\!\perp \cdot \mid \cdot)$ restricted to the lattice of closed lattice elements is separoid, one just has to prove that $X \perp\!\!\!\perp Y \mid Z$ if and only if $X \perp\!\!\!\perp cl_{\perp\!\!\!\perp}(Y) \mid Z$ if and only if $X \perp\!\!\!\perp Y \mid cl_{\perp\!\!\!\perp}(Z)$. This follows from Armstrong's results. □

The significance of this theorem is that if we start with a separoid relation on a lattice, then this separoid relation is also a separoid when restricted to elements that are closed under the relation $\to_{\perp\!\!\!\perp}$.

Theorem 4. *Any finite lattice can be represented as a closure system of a separoid relation defined on a powerset.*

Proof. For any finite lattice $\mathcal{L}$, one identifies the elements with subgroups of a group G. If the group G is assigned a uniform distribution, then the variable corresponding to a subgroup will also have a uniform distribution. With this distribution, a variable is independent of itself given another variable if and only if the other variable determines the first variable. Therefore, statistical independence with respect to the uniform distribution on G gives a separoid relation for which the closure is the original lattice. □

Assume that X and Y are $\rightarrow_h$ closed. Then:

$$h\left(cl_{\rightarrow_h}(X \uplus Y)\right) + h(X \cap Y) = h(X \uplus Y) + h(X \cap Y) \tag{36}$$
$$\leq h(X) + h(Y).$$

Therefore, h restricted to the $\rightarrow_h$ closed elements is polymatroid. We may summarize these observations in the following proposition.

Proposition 5. *If h is a polymatroid function defined on the lattice $(\mathcal{L}, \subseteq)$, then the relation $\rightarrow_h$ satisfies Armstrong's axioms. The function h restricted to the lattice of $\rightarrow_h$ closed elements is polymatroid.*

We recall that a pair of point (Y, Z) is said to be a modular pair, and we write $Y \mathcal{M} Z$ if $Y \cap Z \subseteq X \subseteq Z$ implies that:

$$(X \uplus Y) \cap Z = X. \tag{37}$$

If all pairs are modular, we say that the lattice is modular, and we have:

The modular law $$\qquad\qquad X \uplus (Y \cap Z) = (X \uplus Y) \cap Z. \tag{38}$$

when $X \subseteq Z$.

Proposition 6. *If $(\cdot \perp\!\!\!\perp \cdot \mid \cdot)$ is a separoid relation on a lattice and:*

$$Y \perp\!\!\!\perp Z \mid Y \cap Z \tag{39}$$

then $Y \mathcal{M} Z$ in the lattice of closed elements. In particular, if h is a polymatroid function on a lattice and:

$$h(Y) + h(Z) = h(Y \cap Z) + h(Y \uplus Z), \tag{40}$$

then $Y \mathcal{M} Z$ in the lattice of closed elements.

Proof. If $Y \cap Z \subseteq X \subseteq Z$, then we have the following sequence of implications.

$$Y \perp\!\!\!\perp Z \mid Y \cap Z \tag{41}$$
$$Y \perp\!\!\!\perp X \uplus Z \mid Y \cap Z \tag{42}$$
$$Y \perp\!\!\!\perp Z \mid X \uplus (Y \cap Z) \tag{43}$$
$$Y \perp\!\!\!\perp Z \mid X \tag{44}$$
$$X \uplus Y \perp\!\!\!\perp Z \mid X \tag{45}$$
$$(X \uplus Y) \cap Z \perp\!\!\!\perp (X \uplus Y) \cap Z \mid X \tag{46}$$

Hence,
$$X \rightarrow_{\perp\!\!\!\perp} (X \uplus Y) \cap Z \, and \, cl(X) = cl((X \uplus Y) \cap Z). \tag{47}$$

$\square$

If $\perp\!\!\!\perp$ is separoid, then according to the extensivity (34), the relation $X \perp\!\!\!\perp Y \mid Z$ implies:

$$X \uplus Z \perp\!\!\!\perp Y \uplus Z \mid Z \tag{48}$$

so that $Z \supseteq_{\perp\!\!\!\perp} (X \uplus Z) \cap (Y \uplus Z) \supseteq_{\perp\!\!\!\perp} Z$. Following Dawid [24], we define the relation $X \perp\!\!\!\perp_M Y \mid Z$ by:

$$Z = (X \uplus Z) \cap (Y \uplus Z). \tag{49}$$

Theorem 5. *If a polymatroid function h on a lattice is modular, then the lattice of $\rightarrow_h$ closed elements is modular. If the lattice is modular, then $X \perp\!\!\!\perp_h Y \mid Z$ if and only if $X \perp\!\!\!\perp_M Y \mid Z$ in the lattice of closed elements.*

Proof. If the function h is modular, then all pairs of elements are modular in the lattice of h-closed elements, so the lattice of closed elements is modular. In a modular lattice:

$$I_h(X, Y \mid Z) = h((X \uplus Z) \cap (Y \uplus Z)) - h(Z) \tag{50}$$

so that $X \perp\!\!\!\perp_h Y \mid Z$ holds when $Z \rightarrow_h (X \uplus Z) \cap (Y \uplus Z)$. $\quad\square$

The following result appears in [24] with a longer proof.

Corollary 2. *For a lattice, the relation $X \perp\!\!\!\perp_M Y \mid Z$ is separoid if and only if the lattice is modular.*

Proof. Assume that the lattice is modular. Then, the ranking function r is modular, and $X \rightarrow_r Y$ if and only if $X \supseteq Y$. Therefore, $X \perp\!\!\!\perp_M Y \mid Z$ is equivalent to the separoid relation $I_r(X, Y \mid Z) = 0$.

Assume that the relation $\perp\!\!\!\perp_M$ is separoid. Since $X \perp\!\!\!\perp_M Y \mid X \cap Y$, we have that $X \mathcal{M} Y$. Since all pairs are modular, the lattice is modular. $\quad\square$

4. Entropy in Functional Dependence Lattices

Let $\mathcal{L}$ denote a lattice with maximal element m. Let $\Gamma(\mathcal{L})$ denote the set of polymatroid functions on $\mathcal{L}$. The set $\Gamma(\mathcal{L})$ is polyhedral, and often, we may normalize the polymatroid functions by replacing $h(\cdot)$ by $h(\cdot)/h(m)$. In this way, we obtain a polytope that we will denote $\Gamma_1(\mathcal{L})$.

Definition 2. *A function $h \in \Gamma(\mathcal{L})$ is said to be entropic if there exists a function f from $\mathcal{L}$ into a set of random variables such that $h(X) = H(f(X))$ for any element X in the lattice.*

Let $\Gamma_1^*(\mathcal{L})$ denote the set of normalized entropic functions on $\mathcal{L}$, and let $\bar{\Gamma}_1^*(\mathcal{L})$ denote the closure of $\Gamma_1^*(\mathcal{L})$.

Definition 3. *A lattice is said to be a Shannon lattice if any polymatroid function can be realized approximately by random variables, i.e., $\Gamma_1(\mathcal{L}) = \bar{\Gamma}_1^*(\mathcal{L})$.*

One may then check whether a lattice is a Shannon lattice by checking that the extreme polymatroid functions are entropic or can be approximated by entropic functions.

Example 4. *Let G denote a finite group. For any subgroup $\tilde{G} \subseteq G$, we associate the variable $X_{\tilde{G}}$ that maps an element $g \in G$ into the left coset $g\tilde{G}$. The number of possible values of $X_{\tilde{G}}$ is $|G : \tilde{G}| = \frac{|G|}{|\tilde{G}|}$. Assume that the subgroups are given a functional dependence structure where a variable X is given by a function $A \rightarrow_B$. If A has n elements, then the groups of permutations G have n! elements. The subgroup that leaves X invariant has:*

$$\Pi_{b \in B}(n \cdot P(X = b))! \tag{51}$$

element. Therefore:

$$\begin{aligned}
\ln(|G : G_X|) &= \ln\left(\frac{n}{\Pi_{b \in B}(n \cdot P(X = b))!}\right) \\
&\approx -n \cdot \sum_{b \in B} P(X = b) \ln(P(X = b)) \\
&= n \cdot H(X).
\end{aligned} \tag{52}$$

If U is the uniform distribution on the finite group G, then the distribution of $X_{\tilde{G}}$ is uniform, and the entropy is $H\left(X_{\tilde{G}}\right) = \ln\left(|G|\right) - \ln\left(|\tilde{G}|\right)$. It has been proven that the set of entropic functions generated form a convex cone. Therefore, the normalized polymatroid functions generated by groups has $\bar{\Gamma}_1^*\left(\mathcal{L}\right)$ as closure [4].

From Definition 3, we immediately get the following result.

Proposition 7. *If $\mathcal{L}$ is a Shannon lattice and M is a subset that is a $\cap$-semi-lattice, then M is a Shannon lattice. In particular, all sub-lattices of a Shannon lattice are Shannon lattices.*

Proof. Assume that $\mathcal{L}$ is a Shannon lattice and that M is a sub-lattice. Let $h : M \rightarrow \mathbb{R}$ denote a polymatroid function. For $\ell \in \mathcal{L}$, let $\tilde{\ell}$ denote the $m \in M$ that minimize $h(m)$ under the constraint that $m \supseteq \ell$. Define the function $\tilde{h}(\ell) = h(\tilde{\ell})$. Now, $\tilde{h}$ is an extension of h, and with this definition, $\tilde{h}$ is non-negative and increasing. For $x, y \in \mathcal{L}$, we have:

$$
\begin{aligned}
\tilde{h}(X) + \tilde{h}(Y) &= h\left(\tilde{X}\right) + h\left(\tilde{Y}\right) \\
&\geq h\left(\tilde{X} \uplus \tilde{Y}\right) + h\left(\tilde{X} \cap \tilde{Y}\right) \\
&\geq \tilde{h}(X \uplus Y) + \tilde{h}(X \cap Y)
\end{aligned}
\tag{53}
$$

because $\tilde{X} \uplus \tilde{Y} \geq X \uplus Y$ and $\tilde{X} \cap \tilde{Y} \geq X \cap Y$. Hence, $\tilde{h}$ is submodular. By the assumption, $\tilde{h}$ is entropic, so the restriction of $\tilde{h}$ to M is also entropic. $\square$

With these results it hand, we can start hunting for non-Shannon lattices. We take a lattice that may or may not be a Shannon lattice. We find the extreme normalized polymatroid functions. These extreme polymatroid functions can be found either by hand or by using some suitable software that can find extreme points of a convex polytope specified by a finite set of inequalities. For instance, the R program with package rcdd can find all extreme points of a polytope. For each extreme point, we determine the lattice of closed elements using Proposition 5. These lattices of closed sets will often have a much simpler structure than the original lattice, and the goal is to check if these lattices are Shannon lattices or not. It turns out that there are quite a few of these reduced lattices, and they could be considered as the building blocks for larger lattices.

We recall that an element i is $\uplus$-irreducible if $i = X \uplus Y$ implies that $i = X$ or $i = Y$. An $\cap$-irreducible element is defined similarly. An element is double irreducible if it is both $\uplus$-irreducible and $\cap$-irreducible. The lattice denoted M_n is a modular lattice with a smallest element, a largest element and n double irreducible elements arranged in-between.

Theorem 6. *For any n, the lattice M_n is a Shannon lattice.*

Proof. The proof is essentially the same as the solution to the cryptographic problem stated at the beginning of Section 2. The idea is that one should look for groups with a subgroup lattice M_n and then check that the subgroups of such group have the right cardinality.

Let the values in the double irreducible elements be denoted $h_1, h_2, \ldots, h_n$. If $n = 1$, the extreme polymatroid functions are $h_1 = 0$ and $h_1 = 1$, and these points are obviously entropic. If $n = 2$, the extreme points are $(h_1, h_2) = (0, 1)$, $(h_1, h_2) = (1, 0)$ and $(h_1, h_2) = (1, 1)$, which are all entropic.

Assume $n \geq 3$. Then, the values should satisfy the inequalities:

$$
0 \leq h_i \leq 1,
\tag{54}
$$

$$
h_i + h_j \geq 1.
\tag{55}
$$

If $(h_1, h_2, \ldots, h_n)$ is an extreme point, then each variable should satisfy one of the inequalities with equality. Assume $h_i = 0$. Then, sub-modularity implies that $h_j = 1$ for $j \neq i$. The extreme point $(1, 1, \ldots, 1, 0, 1, \ldots, 1)$ is obviously entropic. If $h_i = 1$, this gives no further constraint on the other

values, so it corresponds to an extreme point on a lattice with one less variable. Finally, assume that $h_i + h_j = 1$ for all i, j. Then, $h_i = 1/2$ for all i. □

Corollary 3. *Any polymatroid function that only takes the values* $0, 1/2$ *and* 1 *is entropic.*

Proof. Assume that the polymatroid function h only takes the values $0, 1/2$, and 1. Then, h defines a separoid relation, and the closed elements form a lattice isomorphic to M_n for some integer n. The function h is entropic on M_n, so h is also entropic on the original lattice. □

Lemma 1. *If h is submodular and increasing on* ∩*-irreducible elements, then h is increasing.*

Proof. Assume that h is submodular and increasing on ∩-irreducible elements. We have to prove that if $X \supseteq Z$, then $h(X) \geq h(Z)$. In order to obtain a contradiction, assume that Z is a maximal element such that there exist an element X such that $X \supseteq Z$, but $h(X) < h(Z)$. We may assume that X cover Z. Since h is increasing at ∩-irreducible elements, Z cannot be ∩-irreducible. Therefore, there exists a maximal element b such that $Y \supseteq Z$, but $Y \not\supseteq X$. Since X cover Z, we have $X \cap Y = Z$. According to the assumptions, $h(X) + h(Y) \geq h(X \uplus Y) + h(X \cap Y)$ and $h(X \uplus Y) \geq h(Y)$ because Z is a maximal element that violates that h is increasing. Therefore, $h(X) \geq h(X \cap Y) = h(Z)$. □

Theorem 7. *Any lattice with seven or fewer elements is a Shannon lattice.*

Proof. Up to isomorphism, there only exist finitely many lattices with seven elements or less. These are listed in the Appendix B. Each of these lattices has finitely many extreme polymatroid functions. These extreme polymatroid functions can be found by hand or by using the R program with package rcdd. All the extreme polymatroid functions on these lattices can be represented by a trivial lattice, or by the two-element chain **2**, or by M_5, or by M_6, or by M_7. All these lattices are representable, and thereby, they are Shannon lattices. □

The number of lattices grows quite fast with the number of elements, and the number of elements is not the best way of comparing lattices.

The Boolean lattice with four atoms is the smallest non-Shannon Boolean algebra. Nevertheless, there are smaller non-Shannon lattices. Figure 2 illustrates the Matúš lattice, which is a lattice with just 11 elements that violates Inequality (1). This corresponds to the fact that the lattice in Figure 2 is not equivalent to a lattice of subgroups of a finite group. The lattices that are equivalent to lattices of subgroups of finite groups have been characterized [30], but the characterization is too complicated to describe here. Using the ideas from [31], one can prove that the Matúš lattice in Figure 2 has infinitely many non-Shannon inequalities. Therefore, any lattice that contains the Matúš lattice as a ∩-semilattice also has infinitely many non-Shannon inequalities.
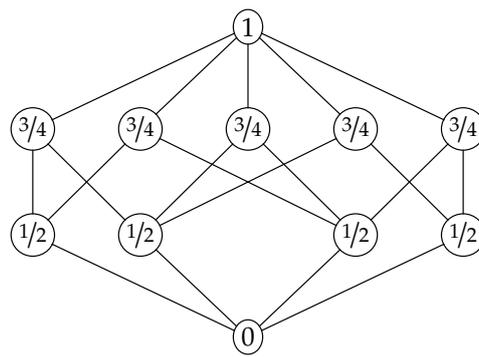
Figure 2. The Matúš lattice with a non-entropic polymatroid function. This lattice is named in honor of František Matúš, who passed away shortly before the submission of this manuscript.

Conjecture 1. *A lattice is a Shannon lattice if and only if the lattice does not contain the Matúš lattice as a ∩-semilattice.*

The result of Matúš has recently found a parallel in matroid theory. An infinite set of inequalities is needed in order to characterize presentable matroids [32–34].

5. The Skeleton of a Lattice

In this section, we will develop a cutting-and-gluing technique that can be used to handle many lattices, but it is especially useful for planar lattices. We present the notion of tolerance. Further details about this concept can be found in the literature [5,35].

Definition 4. *A symmetric and reflexive relation Θ on a lattice is called a tolerance relation if $X_1 \Theta X_2$ and $Y_1 \Theta Y_2$ imply:*

$$(X_1 \cap X_2)\Theta\,(Y_1 \cap Y_2) \tag{56}$$

and

$$(X_1 \uplus X_2)\Theta\,(Y_1 \uplus Y_2). \tag{57}$$

If Θ is a tolerance relation, then for any X, the set $\{Y \in \mathcal{L} \mid X\Theta Y\}$ is an interval in the lattice. These intervals are called the blocks of Θ, and the blocks will be denoted $[X]_\Theta$. For a tolerance relation, the blocks may be considered as elements of the factor $\mathcal{L}/\Theta$, and this factor has a natural structure as a lattice. Congruence relations are special cases of tolerance relations, but in general, the blocks of a tolerance relation may overlap. We note that if the intersection of two blocks is non-empty, then the intersection is a sublattice. If $X \in \mathcal{L}/\Theta$, then $\mathcal{L}_X$ will denote the block in $\mathcal{L}$ determined by X. We defined a glued tolerance relation as a tolerance relation where X cover Y in $\mathcal{L}/\Theta$, implying that $\mathcal{L}_X \cap \mathcal{L}_Y \neq \varnothing$.

A tolerance relation can be identified with a subset of $\mathcal{L} \times \mathcal{L}$, so tolerance relations are ordered by subset ordering. The trivial tolerance relation is the one where $x\Theta y$ holds for all $x, y \in \mathcal{L}$, and this tolerance relation is the greatest tolerance relation. A glued tolerance relation contains any covering pair, and glued tolerance relations are characterized by this property. Therefore, the intersection of two glued tolerance relations is a glued tolerance relation. Therefore, the set of glued tolerance relations forms a lattice. The smallest glued tolerance relation is denoted $\Sigma(\mathcal{L})$ and is called the skeleton of the lattice. An example of a planar modular lattice is given in Figure 3 and the skeeton is given in Figure 4.
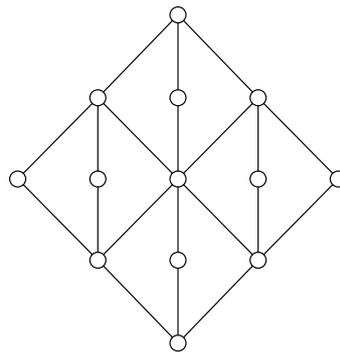
Figure 3. A planar modular lattice.

Lemma 2. *Let $\mathcal{L}$ be a lattice with an increasing function h. If the function h satisfies:*

$$h\left(X\right) + h\left(Y\right) \geq h\left(X \cap Y\right) + h\left(X \uplus Y\right) \tag{58}$$

for all X, Y where $X \cap Y$ is covered by X and Y, then the function h is submodular on $\mathcal{L}$.

Proof. First, we prove that if the function h satisfies:

$$h\left(X\right) + h\left(Y\right) \geq h\left(X \cap Y\right) + h\left(X \uplus Y\right) \tag{59}$$

for all X, Y where $X \cap Y$ is covered by X, then the function h is submodular on $\mathcal{L}$.
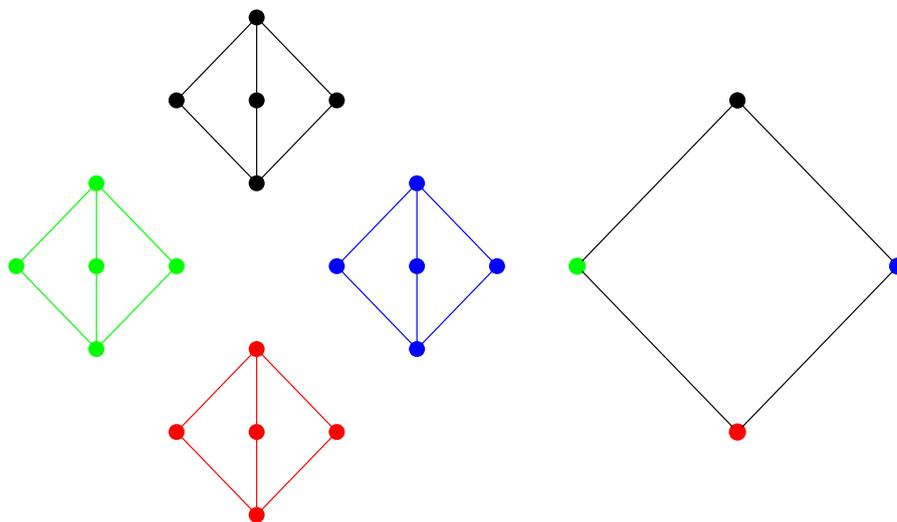
Figure 4. The skeleton of the lattice in the previous figure. It consist of four blocks glued together by the factor lattice illustrated to the right.

Let A and A denote two lattice elements. Define sequences $X_1 \subseteq X_2 \cdots \subseteq X_n = A$ and $Y_1 \subseteq Y_2 \cdots \subseteq Y_n = A \uplus B$ by first defining $X_1 = A \cap B$ and $Y_1 = B$. Assume that X_1 is an element that covers $A \cap B$ and such that $X_1 \leq A$. Let $X_{i+1} \subseteq A$ be a cover of $A \cap Y_i$, and let $Y_{i+1} = X_{i+1} \uplus Y_i$. Then:

$$h\left(X_{i+1}\right) + h\left(Y_i\right) \geq h\left(Y_{i+1}\right) + h\left(X_{i+1} \cap Y_i\right). \tag{60}$$

Adding all these inequalities leads to:

$$h(A) + h(B) \geq h(A \uplus B) + h(A \cap B) + \sum_{i=0}^{n-1} \left(h(X_{i+1} \cap Y_i) - h(X_i) \right) \tag{61}$$

and the inequality is obtained because h is increasing to that $h(X_{i+1} \cap Y_i) - h(X_i) \geq 0$ and because $X_{i+1} \cap Y_i \supseteq X_i$ by construction of the sequences.

To see that, we just need to check submodularity when B covers $A \cap B$ proven in the same way. □

Proposition 8. *Let $\mathcal{L}$ be a lattice with a tolerance relations Θ, and let $h : \mathcal{L} \to \mathbb{R}$ denote some function. Then, h is polymatroid if and only if the restriction of h to any block $\mathcal{L}_x$ is polymatroid.*

If h is entropic, then the restriction to each block is entropic. Characterizing the blocks of a lattice has been done for certain classes of lattices, but here, we shall only mention a single result.

Theorem 8 ([36]). *The blocks of a modular lattice are the maximal atomistic intervals.*

In particular, the skeleton of a modular lattice consists of blocks that are geometric lattices.

6. Results for Planar Lattices

In this section, we will restrict our attention to planar lattices. There are several reasons for this restriction. First of all, any poset with a planar Hasse diagram is a lattice if and only if it has a least element and a greatest element [37]. As a consequence, any ∩-semilattice of a planar lattice is also a planar lattice. Certain cut-and-glue techniques are also very efficient for planar lattices. Finally, both planar distributive lattices and planar modular lattices have nice representations that will play a central role in our proofs.

Theorem 9. *Let h denote a polymatroid function on a planar lattice $\mathcal{L}$. Then, h has an entropic representation if and only if the restriction to each block of $\Sigma(\mathcal{L})$ has an entropic representation.*

Proof. The proof is via induction over the number of elements in the lattice. For a trivial lattice, there is nothing to prove. Assume that the theorem has been proven for all lattices with fewer elements than the number of elements of $\mathcal{L}$. Assume that h is a polymatroid. Since the lattice is planar, it has a left boundary chain $\varnothing \subset L_1 \subset L_2 \cdots \subset L_m$ and a right boundary chain $\varnothing \subset R_1 \subset R_2 \cdots \subset R_n$ where $L_m = R_n$ is the maximal element of $\mathcal{L}$. Let R_k be the minimal element of the right boundary chain such that $L_1 \subseteq R_k$. We note that $R_k = L_1 \uplus R_{k-1}$. Let L_j denote the largest element in the left boundary chain such that $L_j \subseteq R_k$. Then, there is a chain from L_j to R_k, and we have a glued tolerance relation with two blocks $\mathcal{L}_0 = \{ X \in \mathcal{L} \mid X \subseteq R_k \}$ and $\mathcal{L}_1 = \{ X \in \mathcal{L} \mid X \supseteq L_j \}$ and with the two element chain lattice **2** as the factor lattice. These two blocks are glued together along a chain $L_j = y_1 \subset y_2 \subset \cdots \subset y_t = R_k$ that $\mathcal{L}_1$ and $\mathcal{L}_0$ share. There are two cases: either $R_k \subset R_n$ or $R_k = R_n$.

Assume that $R_k \subset R_n$. Then, the glued tolerance relation is non-trivial. Since h restricted to $\{ X \in \mathcal{L} \mid X \supseteq L_j \}$ and $\{ X \in \mathcal{L} \mid X \subseteq R_k \}$ are probabilistically representable, we may without loss of generality assume that there exist two groups G_1 and G_0 such that to $X \in \mathcal{L}_i$, there is a subgroup $G_i(X) \subseteq G_i$ such that $h(X) = \ln |G_i : G_i(X)|$. We associate the variable $X_{G_i(X)}$ that maps an element $g \in G$ into the left coset $gG_i(X)$. The goal is to find a joint distribution to a set of variables associated with each $X \in \mathcal{L}$. We note that all variables in $\mathcal{L}_0$ are functions of r_k, so if we map $X_{G_1(r_k)}$ into $X_{G_0(r_k)}$, all other variables in $\mathcal{L}_2$ are determined. In particular, the chain $y_1 \subset y_2 \subset \ldots y_t$ is determined by $r_k = y_t$. The sequences $X_{G_1(y_i)}$ are mapped into the sequence $X_{G_0(y_i)}$ recursively, starting with mapping $X_{G_1(y_1)}$ into $X_{G_0(y_1)}$. This is possible since $X_{G_1(y_1)}$ and $X_{G_2(y_1)}$ are uniform distributions on sets of the same size. Now, there are equally many values of $X_{G_1(y_2)}$ and $X_{G_0(y_2)}$ that map into the same

values of $X_{G_1(y_1)}$ and $X_{G_0(y_1)}$, so the the values of $X_{G_1(y_2)}$ and $X_{G_0(y_2)}$ can be mapped into each other. We continue like that until all the random variables along the chain $y_1 \subset y_2 \subset \dots y_t$ have been identified.

If $r_k = r_n$, then we make a similar construction with the role of the left chain and the right chain reversed. If this leads to a non-trivial glued tolerance relation, we glue representations together as we did above.

If both the left chain and the right chain lead to trivial glued tolerance relations, then $L_1 \uplus r_1$ is the maximal element of $\mathcal{L}$, and the whole lattice consists of a single block in $\Sigma(\mathcal{L})$. In this case, the content of the theorem is trivial. □

Theorem 10. *All planar modular lattices are Shannon lattices.*

Proof. Without loss of generality, we may assume that the lattice consists of just one block for the tolerance relation $\Sigma(\mathcal{L})$. A modular block is atomistic, but if a modular planar lattice is atomistic, it is equivalent to the trivial lattice or to the lattice **2**, or to the lattice $\mathbf{2} \times \mathbf{2}$, or to one of the lattices M_n. □

Our construction actually tells us more. If the lattice is distributive, it is glued together with blocks that are either equivalent to **2** or to the lattice $\mathbf{2} \times \mathbf{2}$. Therefore, the lattice is a sublattice of a product of two chains, as illustrated in Figure 5. This result was first proven by Dilworth [38]. Other characterizations of planar distributive lattices can be found in the literature [39]. Since the extreme polymatroid functions on the lattices **2** and the lattice $\mathbf{2} \times \mathbf{2}$ only take the values zero and one, the same is true for any planar distributive lattice.

A modular planar lattice will also contain blocks of the type M_n. Therefore, a modular planar lattice can be obtained from a distributive planar lattice by adding double irreducible elements [40], as illustrated in Figure 6.
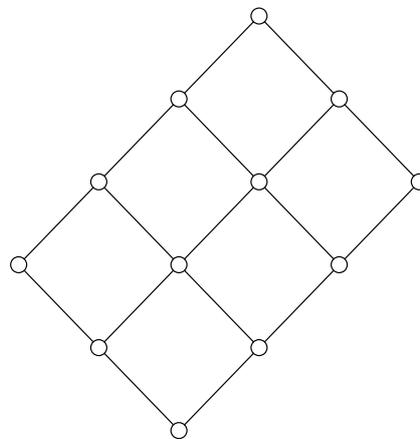
Figure 5. A product of two chains.

Since M_n has extreme polymatroid functions that take the values 0, $1/2$ and 1, the extreme functions are modular. Gluing such modular functions together leads to extreme polymatroid functions that are modular. Therefore, all extreme polymatroid functions on a planar modular lattice can be represented by a planar modular lattice with a modular function. Therefore, the independence structure is given by $(X \perp\!\!\!\perp Y \mid Z)$ when $Z = (X \uplus Z) \cap (Y \uplus Z)$.
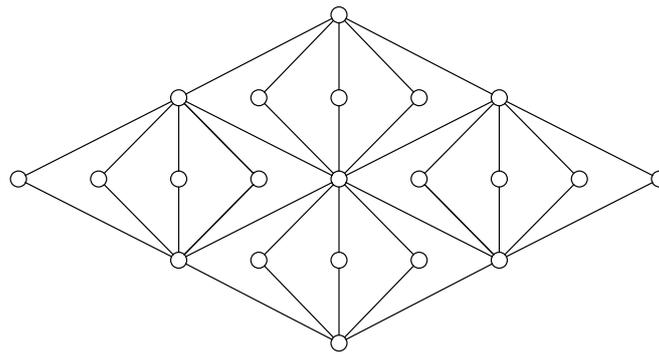
Figure 6. A planar modular lattice.

The extreme polymatroid functions on a planar modular lattice can be represented as follows. Let $X_1, X_2, \ldots X_m, Y_1, Y_2, \ldots, Y_n$ denote independent random variables uniformly distributed over $\mathbb{Z}_p$ for some large value of p. Let Z_{ij} denote the random variable:

$$\biguplus_{\ell \leq i} X_\ell \uplus \biguplus_{\ell \leq j} Y_\ell. \tag{62}$$

and let Z_{ijk} denote the random variable:

$$\biguplus_{\ell \leq i} X_\ell \uplus \biguplus_{\ell \leq j} Y_\ell \uplus \left(X_{i+1} + k \cdot Y_{j+1} \right) \tag{63}$$

for $k > 0$. The way to index the variables can be seen in Figure 7. Then, the entropy is proportional to the ranking function. A polymatroid function h that has a representation given by an Abelian group satisfies the Ingleton inequalities [41], i.e., inequalities of the form:

$$\begin{aligned} h(X) + h(Y) + h(X \uplus Y \uplus V) + h(X \uplus Y \uplus W) + h(V \uplus W) \leq \\ h(X \uplus Y) + h(X \uplus V) + h(X \uplus W) + h(Y \uplus V) + h(Y \uplus W). \end{aligned} \tag{64}$$

Therefore, the Shannon inequalities imply the Ingleton inequalities as long as the polymatroid function is defined on a planar modular lattice. Paajanen [42] has proven that under some conditions, the entropy function of a nilpotent p-group can be represented by an Abelian group. The core of the proof was that the subgroup lattice of a nilpotent p-group is also the subgroup lattice of an Abelian group. Many of these lattices are planar, and in these cases, the results by Paajanen follow from our results on planar graphs.
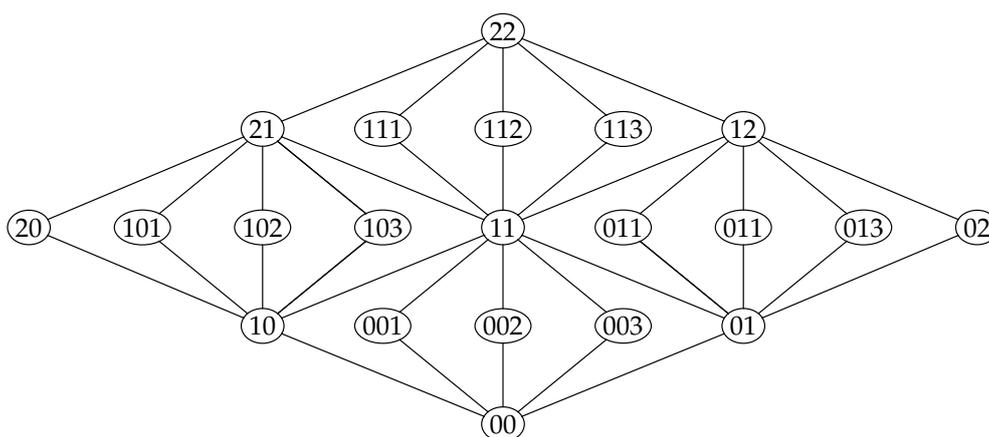
Figure 7. A planar modular lattice with indexing of the elements.

7. Discussion

In this paper, we have proven that the three basic Shannon inequalities are sufficient for certain lattices. It would be a major step forward if one could make a complete characterization of lattices without non-Shannon inequalities, but this may be too ambitious. In order to obtain results, one may have to restrict to certain classes of lattices like general modular lattices or geometric lattices. For handling such lattices, one would have to develop new techniques that may also be of wider interest.

Lattices seem to fall into two types. For one type, one does not have non-Shannon inequalities. For the other type, there are infinitely many non-Shannon inequalities. We do not know of any lattice with non-Shannon inequalities where the entropic functions are characterized by finitely many inequalities. Apparently, the complexity increases from three basic inequalities to infinitely many inequalities, and this transition seems to happen due to the Matúš lattice. Similarly matroids in general have no finite characterization, and conditional independence does not have a finite characterization. It appear to be the case that the leap from low complexity to infinite complexity happens for the same reason and seems to be related to the structure of the Matús lattice. In this paper, we have provided some basic results and a common terminology that should be useful for further exploration of this research area.

Bayesian networks and similar graphical models have not been discussed in the present paper. Nevertheless, Bayesian networks are closely related to functional dependencies, so important properties of Bayesian networks can be translated into lattice language. This will be the topic of a separate publication [43], but some preliminary results have already been published [7].

We have seen how a separoid relation generates a notion of functional dependence. For modular lattices, we have also seen that the lattice structure generates a separoid relation. It is an open question to what extent general lattices are born with a canonical notion of conditional independence that can be formalized in terms of separoids. For functional dependencies corresponding to Bayesian networks, this question has been studied in detail [16], but more general results related to these questions would be of great importance to our understanding of concepts related to cause and effect.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Augmentation

In database theory extensivity (18) is replaced by the following property:.

Augmentation $X \to Y$ implies $X \uplus Z \to Y \uplus Z$.   (A1)

If $Z = X$, the augmentation (A1) reduces to the extensivity (18). In a finite lattice, the extensivity (18) together with the other Armstrong axioms imply the augmentation (A1). To see this, first, we observe that in a finite lattice extensivity (18), monotonicity (19) and transitivity (20) imply that $X \to Y$ is equivalent to $cl(X) \supseteq Y$. The monotonicity (16) gives $cl(X \uplus Z) \geq Y$. Using the monotonicity (16) and extensivity (15), we also get $cl(X \uplus Z) \geq cl(Z) \geq Z$. Combining these two inequalities gives $cl(X \uplus Z) \geq Y \uplus Z$, as desired.

The condition in Theorem 1 that the lattice is finite can be relaxed to the ascending chain condition, because this is essentially what is used to conclude that the chain (22) must stop. The observation that augmentation can be relaxed to extensivity could be used to simplify some algorithms for database normalization.

Appendix B. Lattices of Size 1–7

Here, we give a complete list of the Hasse diagrams of lattices with seven or fewer elements.

Appendix B.1. Lattice of Size 1

The trivial lattice is the only lattice of Size 1.

Appendix B.2. Lattice of Size 2

The two element chain **2** is the only lattice of Size 2.

Appendix B.3. Lattices of Size 3

The three element chain **3** is only one lattice of Size 3, and and it is distributive. The extreme polymatroid functions can be represented by the lattice **2**.

Appendix B.4. Lattices of Size 4

There are two lattices of Size 4, and they are both distributive. Their extreme polymatroid functions can be represented by the lattice **2**.

2 × 2
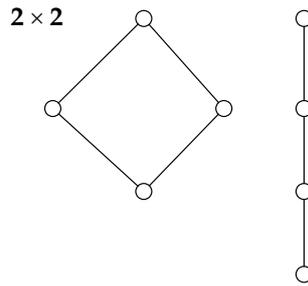
Appendix B.5. Lattices of Size 5

The lattice M_5 is modular, but not distributive. It has a a new non-trivial polymatroid function as the extreme point. The other extreme points can be represented by M_3 and the lattice **2**.

M_3

N_5
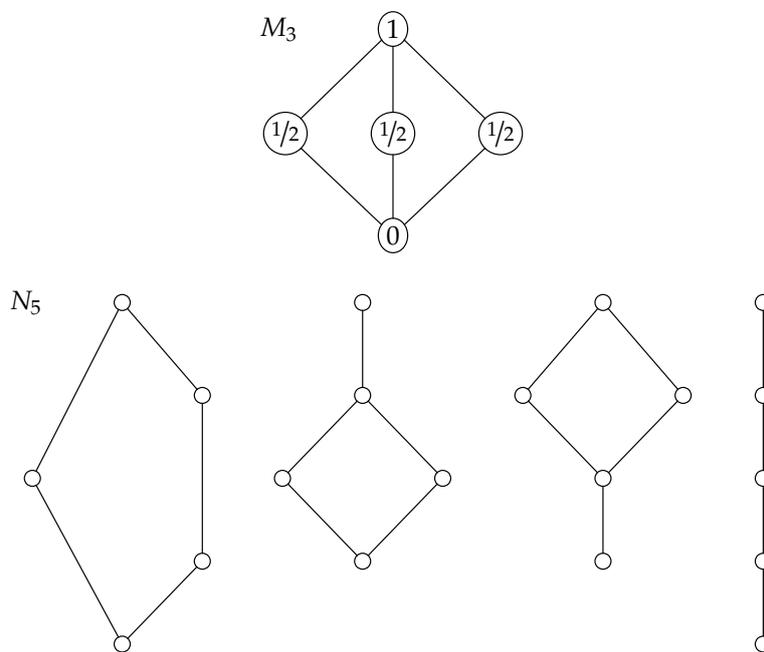
Appendix B.6. Lattices of Size 6

The lattice M_6 has a new non-trivial polymatroid extreme point. The other extreme points can be represented by M_3 and the lattice: **2**.

M_4
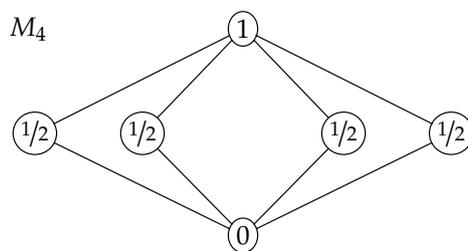
The next five lattices have extreme points that can be represented by M_5 or the lattice **2**. The first two lattices are modular, but not distributive. The next three are not modular.

V_1 L_4 L_5
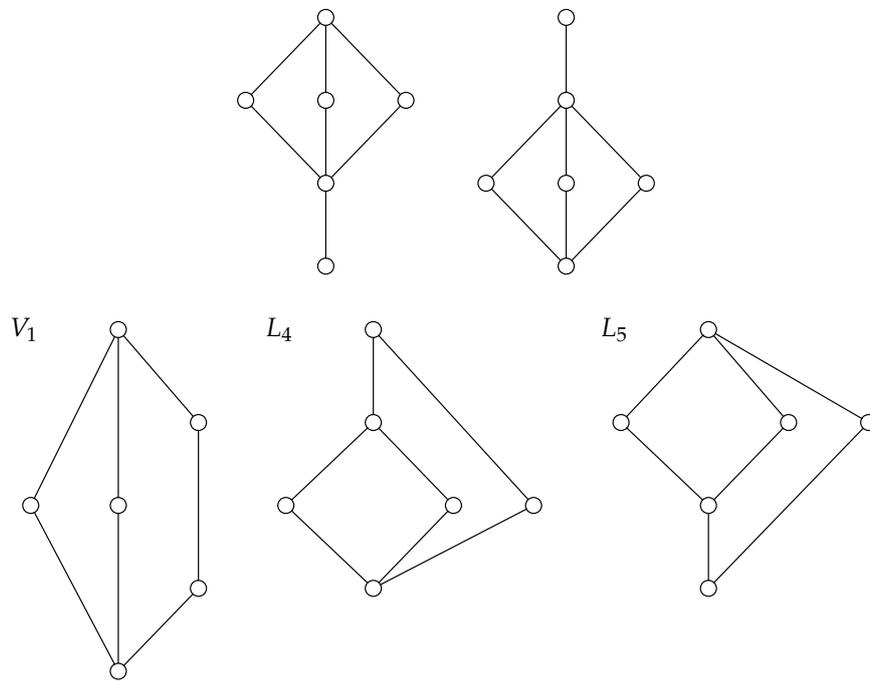
The extreme points of the last nine lattices are all represented by the lattice **2**. The first four are not modular.
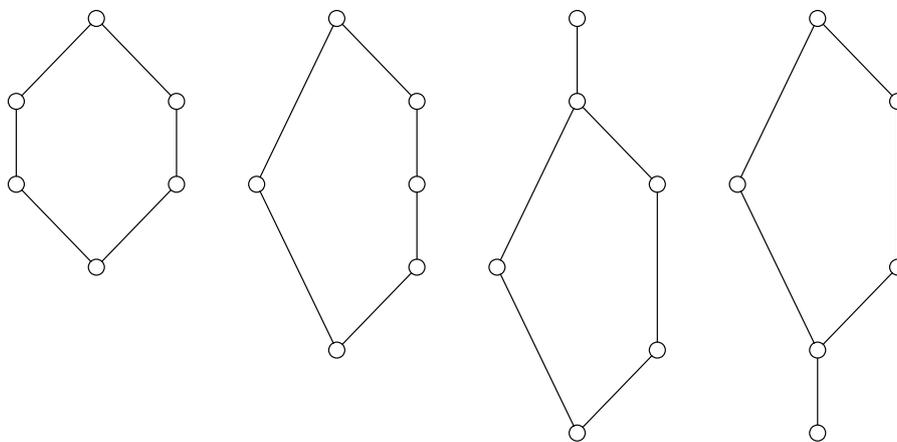
The last five are distributive.
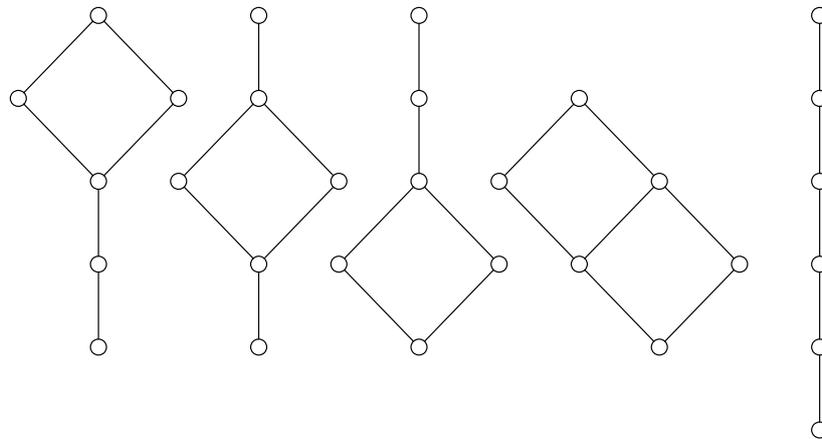
Appendix B.7. Lattices of Size 7

The lattice M_5 has a new polymatroid extreme point. The other extreme points can be represented by M_4, M_3 and the lattice **2**.
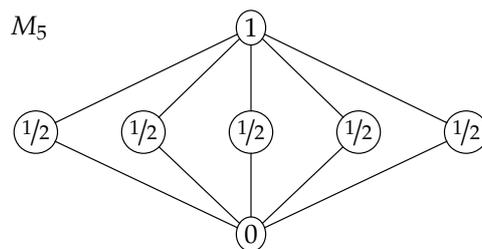
The next seven lattices have extreme points that can be represented by M_4, M_3 or the lattice **2**. The first two lattices are modular. The last five lattices are not modular.
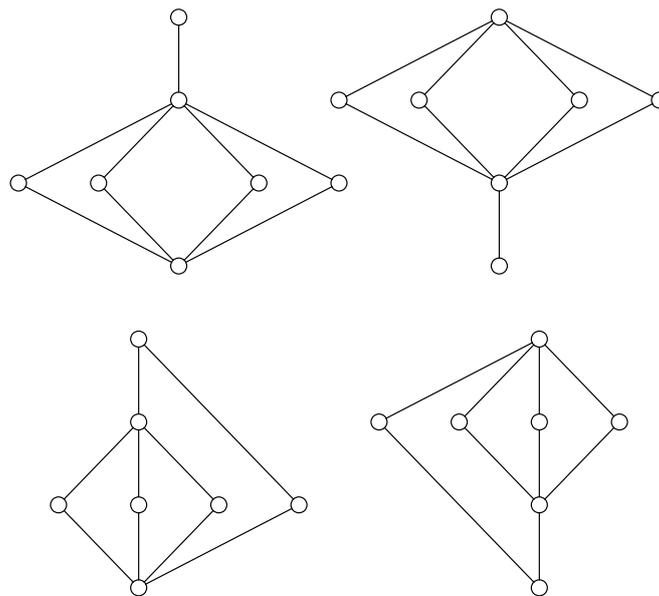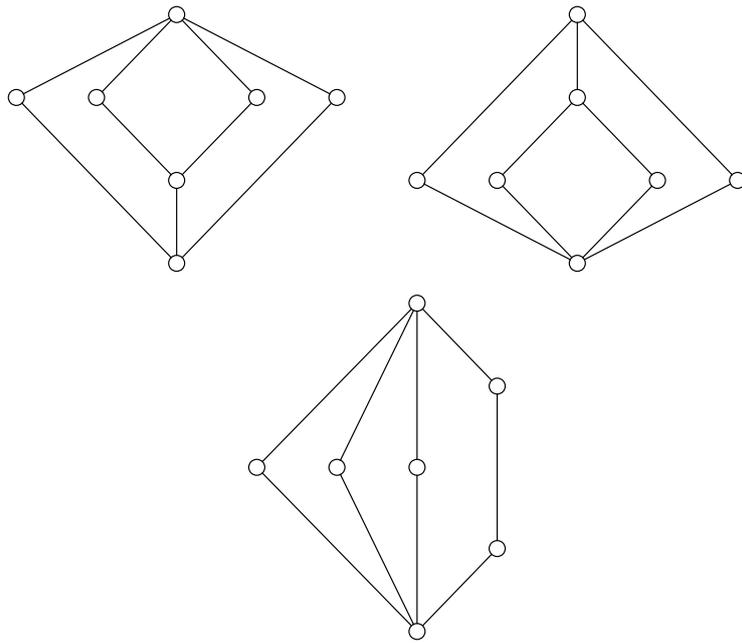
The following lattices have extreme polymatroid functions that can be represented by M_3 or the lattice **2**. The first five lattices are modular.
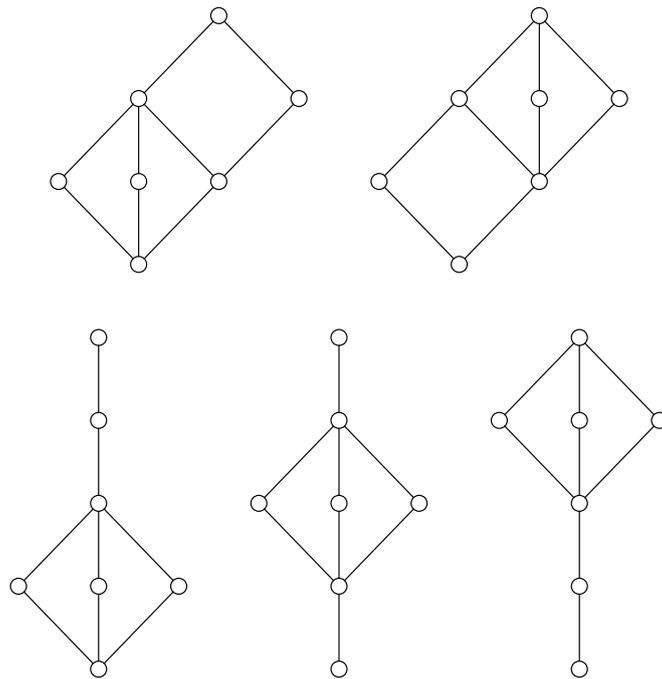
The next lattices are not modular.
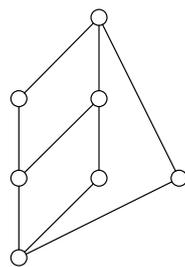
S_7
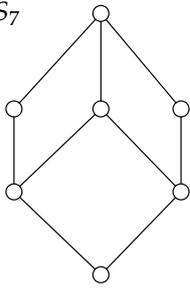
S_7^*
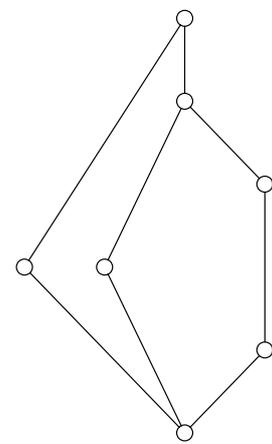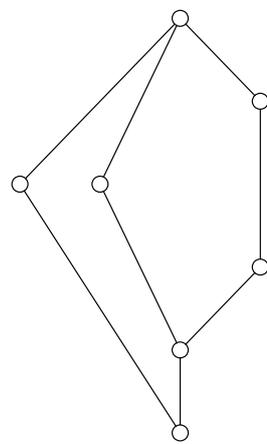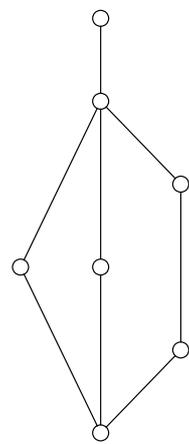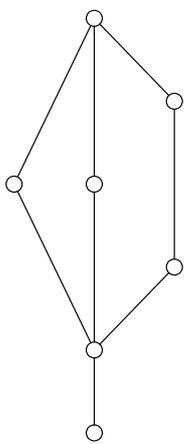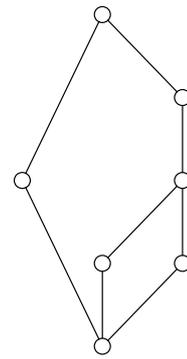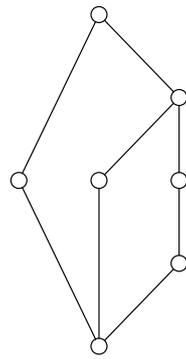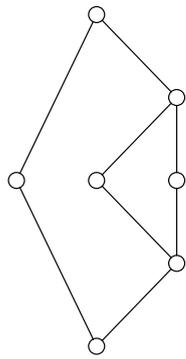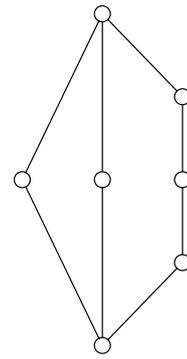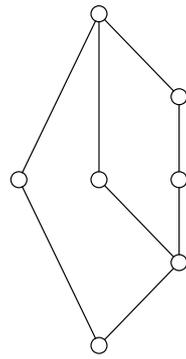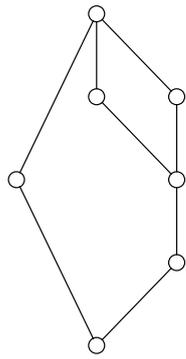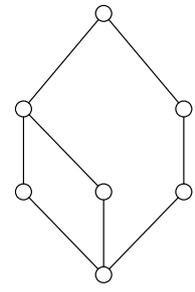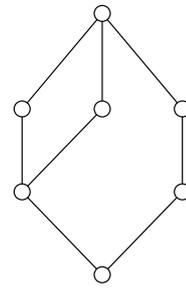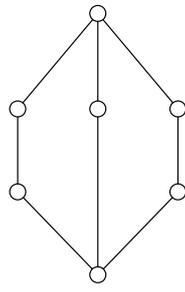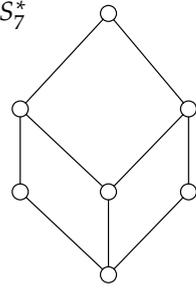
The last 22 lattices of Size 7 only have trivial extreme points. The first 14 lattices are not modular.

L_3
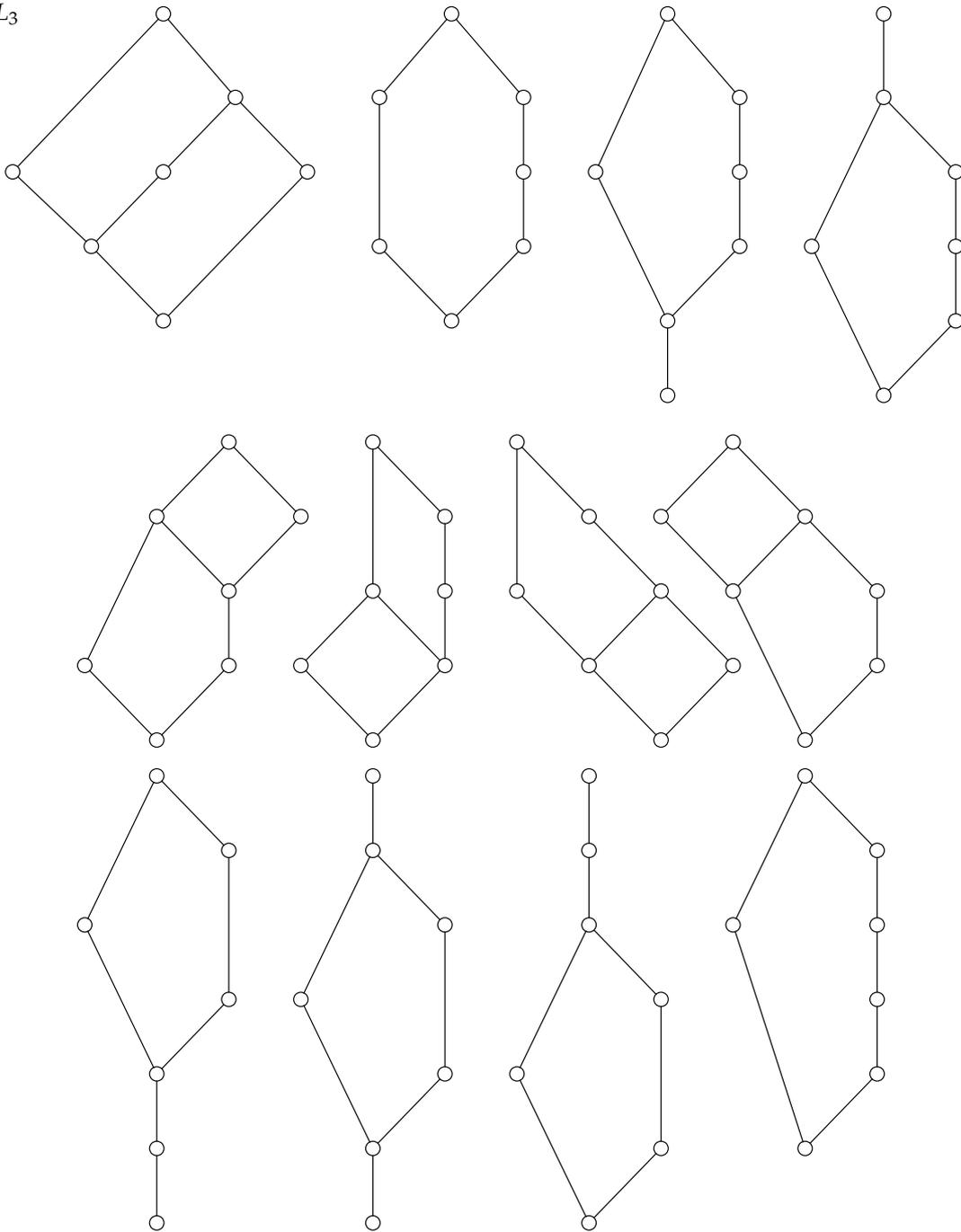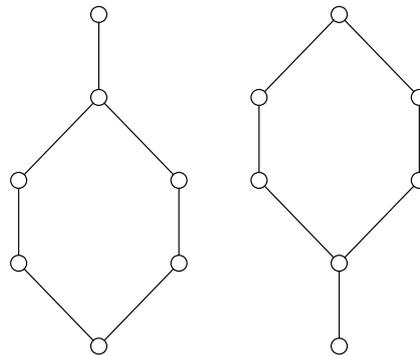
The last eight lattices are distributive.
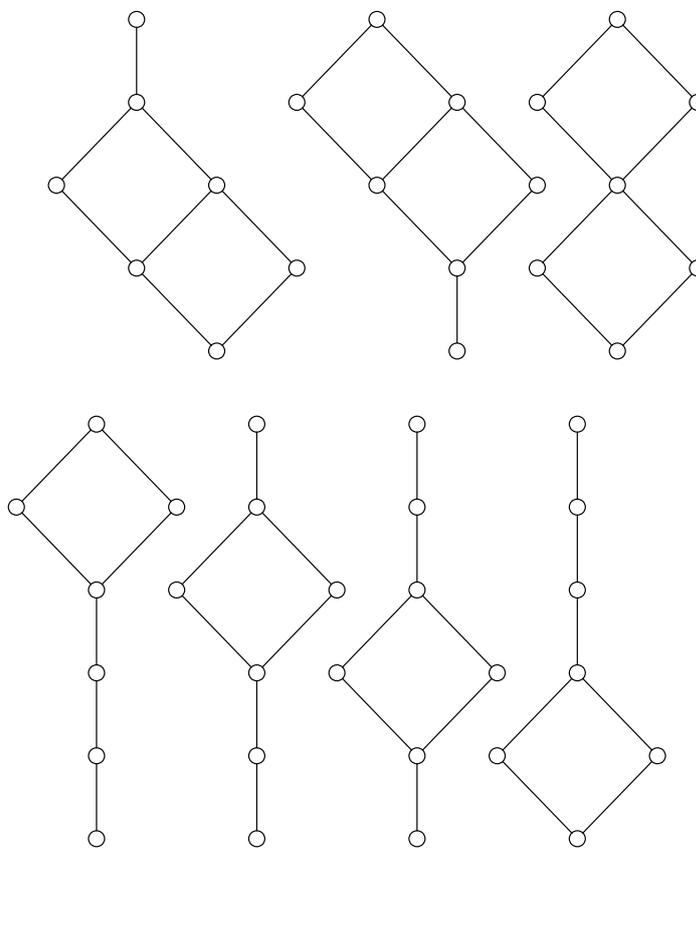
References

1. Zhang, Z.; Yeung, R.W. On characterization of entropy function via information inequalities. *IEEE Trans. Inform. Theory* **1998**, *44*, 1440–1452. [CrossRef]
2. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656. [CrossRef]
3. McGill, W. Multivariate information transmission. *Psychometrika* **1954**, *19*, 97–116. [CrossRef]
4. Yeung, R.W. *A First Course in Information Theory*; Kluwer: New York, NY, USA, 2002.
5. Stern, M. *Semimodular Lattices. Theory and Applications*; Number 73 in Encyclopedia of Mathematics and Its Applications; Cambridge University Press: Cambridge, UK, 1999.

6.   Chan, T.H.; Yeung, R.W. On a Relation between Information Inequalities and Group Theory. *IEEE Trans. Inform. Theory* **2002**, *48*, 1992–1995. [CrossRef]

7.   Harremoës, P. Functional Dependences and Bayesian Networks. In *Proceedings of the WITMSE 2011*; Number Report C-2011-45 in Series of Publications C; Rissanen, J., Myllymäki, P., Teemu Roos, I.T., Yamanishi, K., Eds.; Department Computer Science, University of Helsinki: Helsinki, Finland, 2011; pp. 35–38.

8.   Harremoës, P. Lattices with non-Shannon inequalities. In Proceedings of the 2015 IEEE International Symposium on Information Theory, Hong Kong, China, 14–19 June 2015; pp. 740–744. [CrossRef]

9.   Lee, T.T. An algebraic theory of relational databases. *Bell Syst. Tech. J.* **1983**, *62*, 3159–3204. [CrossRef]

10.   Demetrovics, J.; Libkin, L.; Muchnik, I.B. Functional dependencies and the semilattice of closed classes. In Proceedings of the 2nd Symposium on Mathematical Fundamentals of Database Systems (MFDBS '89), Visegrád, Hungary, 26–30 June 1989; Springer: Berlin, Germany, 1989; pp. 136–147.

11.   Matúš, F. Abstract functional dependency structures. *Theor. Comput. Sci.* **1991**, *81*, 117–126. [CrossRef]

12.   Demetrovics, J.; Libkin, L.; Muchnik, I.B. Functional Dependencies in Relational Databases: A Lattice Point of View. *Discret. Appl. Math.* **1992**, *40*, 155–185. [CrossRef]

13.   Levene, M. A Lattice View of Functional Dependencies in Incomplete Relations. *Acta Cybern.* **1995**, *12*, 181–207.

14.   Thakor, S.; Chan, T.; Grant, A. A minimal set of Shannon-type inequalities for functional dependence structures. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 679–683. [CrossRef]

15.   Chan, T.; Thakor, S.; Grant, A. A Minimal Set of Shannon-type Inequalities for MRF Structures with Functional Dependencies. In Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018; pp. 1759–1763.

16.   Harremoës, P. Time and Conditional Independence; IMFUFA-Tekst, IMFUFA Roskilde University, 1993; Volume 255. Original in Danish Entitled Tid og Betinget Uafhængighed. An English Translation of Some of the Chapters. Available online: http://www.harremoes.dk/Peter/afh/afhandling.pdf (accessed on 10 October 2018).

17.   Caspard, N.; Monjardet, B. The lattices of closure systems, closure operators, and implicational systems on a finite set: A survey. *Discret. Appl. Math.* **2003**, *127*, 241–269. [CrossRef]

18.   Grätzer, G. *General Lattice Theory*, 2nd ed.; Birkhäuser: Basel, Switzerland, 2003.

19.   Armstrong, W.W. Dependency Structures of Data Base Relationships. In Proceedings of the IFIP Congress, Stockholm, Sweden, 5–10 August 1974; pp. 580–583.

20.   Ullman, J.D. *Principles of Database and Knowledge-Base Systems*; Computer Science Press: Stanford, CA, USA, 1989; Volume 1.

21.   Levene, M.; Loizou, G. *A Guide Tour of Relational Databases and Beyond*; Springer: Berlin, Germany, 1999.

22.   Whitman, P.M. Lattices, equivalence relations, and subgroups. *Bull. Am. Math. Soc.* **1946**, *52*, 507–522. [CrossRef]

23.   Shannon, C. The lattice theory of information. *Trans. IRE Prof. Group Inf. Theory* **1953**, *1*, 105–107. [CrossRef]

24.   Dawid, A.P. Separoids: A mathematical framework for conditional independence and irrelevance. *Ann. Math. Artif. Intell.* **2001**, *32*, 335–372. [CrossRef]

25.   Constantinou, P.; Dawid, A.P. Extended Conditional Independence and Applications in Causal Inference. *Ann. Stat.* **2017**, *45*, 1–36. [CrossRef]

26.   Paolini, G. Independence Logic and Abstract Independence Relations. *Math. Logic Q.* **2015**, *61*, 202–216. [CrossRef]

27.   Pearl, J. *Probabilistic Reasoning in Intelligent Systems*; Morgan Kaufmann Publ.: San Mateo, CA, USA, 1988.

28.   Studený, M. *Probabilistic Conditional Independence Structures*; Springer: Belin, Germany, 2005.

29.   Studený, M. Conditional Independence Relations Have No Finite Complete Characterization, 1990. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.7014&rep=rep1&type=pdf (accessed on 10 October 2018).

30.   Schmidt, R. *Subgroup Lattices of Groups*; Walter de Gruyter: Berlin, Germany, 1994.

31.   Matúš, F. Infinitely many information inequalities. In Proceedings of the 2007 IEEE International Symposium on Information Theory, Nice, France, 24–29 June 2007; pp. 2101–2105. [CrossRef]

32.   Vámos, P. The Missing Axiom of Matroid Theory is Lost Forever. *J. Lond. Math. Soc.* **1978**, *18*, 403–408. [CrossRef]

33. Mayhew, D.; Whittle, G.; Newman, M. Is the Missing Axiom of Matroid Theory Lost Forever? *Q. J. Math.* **2014**, *65*, 1397–1415. [CrossRef]

34. Mayhew, D.; Newman, M.; Whittle, G. Yes, the "missing axiom" of matroid theory is lost forever. *Trans. Am. Math. Soc.* **2018**, *370*, 5907–5929. [CrossRef]

35. Czédli, G. Factor lattices by tolerance. *Acta Sci. Math.* **1982**, *44*, 35–42.

36. Hermann, C. S-verklebte Summen von Verbänden. *Math. Z.* **1973**, *130*, 255–274. [CrossRef]

37. Quackenbush, R.W. Planar Lattices. In Proceedings of the University of Houston Lattice Theory Conference 1973, Houston, TX, USA, 22–24 March 1973.

38. Dilworth, R.P. A decomposition theorem for partially ordered sets. *Ann. Math.* **1950**, *51*, 161–166. [CrossRef]

39. Chen, C.C.; Koh, K.M. A characterization of finite distributive planar lattices. *Discret. Math.* **1973**, *5*, 207–213. [CrossRef]

40. Quackenbush, G.G.W. The variety generated by planar modular lattices. *Algebra Universalis* **2010**, *63*, 187–201. [CrossRef]

41. Guille, L.; Chan, T.; Grant, A. The Minimal Set of Ingleton Inequalities. *IEEE Trans. Inform. Theory* **2011**, *57*, 1849–1864. [CrossRef]

42. Paajanen, P. Finite *p*-Groups, Entropy Vectors, and the Ingleton Inequality for Nilpotent Groups. *IEEE Trans. Inf. Theory* **2014**, *60*, 3821–3824. [CrossRef]

43. Harremoës, P. Influence Diagrams as Convex Geometries. Available online: http://www.harremoes.dk/Peter/FunctionalDAG.pdf (accessed on 21 September 2018).