Dual-Bounded Hypergraphs: A Survey*

Endre Boros† Khaled Elbassioni‡ Vladimir Gurvich† Leonid Khachiyan‡
Kazuhisa Makino§

Abstract

This short paper surveys recent results on the generation of implicitly given hypergraphs, as well as their applications in data mining, reliability theory, integer programming and combinatorics.

More precisely, we consider a monotone property π over the subsets of a finite set V, the corresponding family $\mathcal{S}_\pi$ of subsets satisfying property π, and the problem of generating (sequentially) the family $\mathcal{F}_\pi$ of all minimal subsets in $\mathcal{S}_\pi$, when only V is given explicitly, and π is represented by an oracle $\mathcal{O}_\pi$. We show that for a number of interesting monotone properties, the family $\mathcal{F}_\pi$ is uniformly dual-bounded *allowing for the incrementally efficient generation of the members of $\mathcal{F}_\pi$.*

Important applications include the efficient generation of minimal infrequent sets of a database (data mining), minimal connectivity ensuring collections of subgraphs from a given list (reliability theory), minimal feasible solutions to a system of monotone inequalities in integer variables (integer programming), minimal spanning collections of subspaces from a given list (linear algebra) and maximal independent sets in the intersection of matroids (combinatorial optimization).

In contrast to these results, the analogous problem of generating the family of maximal subsets not having property π is NP-hard for most of the monotone properties π considered in this paper.

*The research of the first four authors was supported in part by the National Science Foundation Grant IIS-0118635. The research of the first and third authors was also supported in part by the Office of Naval Research Grant N00014-92-J-1375. The second and third authors are also grateful for the partial support by DIMACS, the National Science Foundation's Center for Discrete Mathematics and Theoretical Computer Science.

†RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08854-8003; ({boros,gurvich} @rutcor.rutgers.edu).

‡Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway NJ 08854-8003; ({elbassio@paul, leonid@cs}.rutgers.edu).

§Division of Systems Science, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka, 560-8531, Japan; (makino@sys.es.osaka-u.ac.jp).

1 Introduction

Let V be a finite set of cardinality $|V| = n$. For a hypergraph $\mathcal{H} \subseteq 2^V$, let us denote by $\mathcal{I}(\mathcal{H})$ the family of its *maximal independent* subsets, i.e. maximal subsets of V not containing any hyperedges of $\mathcal{H}$. The complement of a maximal independent subset is called a *minimal transversal* of $\mathcal{H}$ (i.e. minimal subset of V intersecting all elements of $\mathcal{H}$). The collection $\mathcal{H}^d$ of minimal transversals is also called the *dual* of $\mathcal{H}$.

Let us consider a monotone property π over the subsets of V, and let $\mathcal{S}_\pi$ denote the corresponding collection of all subsets having property π. We shall assume that π is represented by a (polynomial time) *satisfiability oracle* $\mathcal{O}_\pi$, i.e. an algorithm which, given an input description of $\mathcal{O}_\pi$ of size $\|\mathcal{O}_\pi\|$ and a subset $X \subseteq V$, can decide whether or not X has property π, in time polynomial in n and $\|\mathcal{O}_\pi\|$. Since π is monotone, $Y \supseteq X \in \mathcal{S}_\pi$ implies $Y \in \mathcal{S}_\pi$. Thus, we may as well represent π by the hypergraph $\mathcal{F}_\pi \subseteq \mathcal{S}_\pi$ consisting of all minimal subsets having property π. Let us note that the oracle $\mathcal{O}_\pi$ could equivalently be called a *membership oracle* for the family $\mathcal{S}_\pi$, or a *superset oracle* for the family $\mathcal{F}_\pi$, since a subset $X \subseteq V$ has property π if and only if $X \in \mathcal{S}_\pi$, which is further equivalent to the existence of a subset $Y \subseteq X$, $Y \in \mathcal{F}_\pi$. Let us also note that the complementary family $\mathcal{I}_\pi = 2^V \setminus \mathcal{S}_\pi$ is an *independence system* (i.e. $Y \subseteq X \in \mathcal{I}_\pi$ implies $Y \in \mathcal{I}_\pi$, see e.g. [33]), the maximal elements of which are the maximal independent sets of $\mathcal{F}_\pi$.

Let us finally remark that any Sperner hypergraph $\mathcal{H}$ (i.e. one which has no two hyperedges $X, Y \in \mathcal{H}$ for which $X \subseteq Y$ or $Y \subseteq X$) defines uniquely a monotone property π such that $\mathcal{H} = \mathcal{F}_\pi$ holds.

Given a monotone property π, described by a satisfiability oracle $\mathcal{O}_\pi$, an interesting problem is to list incrementally all elements of $\mathcal{F}_\pi$:

GEN$(\mathcal{F}_\pi, \mathcal{X})$**:** *Given a subfamily $\mathcal{X} \subseteq \mathcal{F}_\pi$, either find a new minimal satisfying set $X \in \mathcal{F}_\pi \setminus \mathcal{X}$, or prove that the given partial list is complete $\mathcal{X} = \mathcal{F}_\pi$.*

Clearly, the entire family $\mathcal{F}_\pi$ can be generated by initializing $\mathcal{X} = \emptyset$ and iteratively solving the above problem $|\mathcal{F}_\pi|+1$ times. Note also that the size of $\mathcal{F}_\pi$ may be exponentially large both in n and the size of the oracle $\|\mathcal{O}_\pi\|$. For this reason, we shall consider algorithms that are *incrementally efficient*, i.e. which solve the above problem GEN($\mathcal{F}_\pi, \mathcal{X}$) efficiently in terms of n, $\|\mathcal{O}_\pi\|$ and the size of the partial list $|\mathcal{X}|$.

Although, in general, problem GEN($\mathcal{F}_\pi, \mathcal{X}$), given a monotone property π, is NP-hard (see e.g. [27, 33, 37]), the following *joint generation* problem is most likely not (as we shall see in the next section):

GEN($\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi), \mathcal{X}, \mathcal{Y}$)**:** *Given two explicitly listed set families $\mathcal{X} \subseteq \mathcal{F}_\pi$ and $\mathcal{Y} \subseteq \mathcal{I}(\mathcal{F}_\pi)$, either find a new set in $(\mathcal{F}_\pi \setminus \mathcal{X}) \cup (\mathcal{I}(\mathcal{F}_\pi) \setminus \mathcal{Y})$, or prove that these families are complete: $(\mathcal{X}, \mathcal{Y}) = (\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi))$.*

It is clear that for a given monotone property π, represented by a satisfiability oracle $\mathcal{O}_\pi$, we can generate both $\mathcal{F}_\pi$ and $\mathcal{I}(\mathcal{F}_\pi)$ simultaneously by starting with $\mathcal{X} = \mathcal{Y} = \emptyset$ and solving GEN($\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi), \mathcal{X}, \mathcal{Y}$) $|\mathcal{F}_\pi| + |\mathcal{I}(\mathcal{F}_\pi)| + 1$ times, incrementing in each iteration either $\mathcal{X}$ or $\mathcal{Y}$ by the newly found subset $S \in (\mathcal{F}_\pi \setminus \mathcal{X}) \cup (\mathcal{I}(\mathcal{F}_\pi) \setminus \mathcal{Y})$, according to the answer of the oracle $\mathcal{O}_\pi$, until we have $(\mathcal{X}, \mathcal{Y}) = (\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi))$.

Though it may sound terribly redundant to use the above joint generation to generate $\mathcal{F}_\pi$ alone (and simply discard at the end all members of $\mathcal{I}(\mathcal{F}_\pi)$), it was in fact shown (see [26]) that no satisfiability oracle based algorithm can generate $\mathcal{F}_\pi$ in fewer than $|\mathcal{F}_\pi|+|\mathcal{I}(\mathcal{F}_\pi)|$ steps, in general.

Unfortunately, the above joint generation still may not be an efficient algorithm for solving either of GEN($\mathcal{F}_\pi, \mathcal{X}$) or GEN($\mathcal{I}(\mathcal{F}_\pi), \mathcal{Y}$) separately for the simple reason that we do not control which of the families $\mathcal{F}_\pi \setminus \mathcal{X}$ and $\mathcal{I}(\mathcal{F}_\pi) \setminus \mathcal{Y}$ contains each new hyperedge produced by the algorithm. Suppose we want to generate $\mathcal{F}_\pi$ and the family $\mathcal{I}(\mathcal{F}_\pi)$ is exponentially larger than $\mathcal{F}_\pi$. Then, if we are unlucky, we may get hyperedges of $\mathcal{F}_\pi$ with exponential delay, while getting large subfamilies of $\mathcal{I}(\mathcal{F}_\pi)$ (which are not needed at all) in between.

For the above reasons, we shall study in this paper Sperner hypergraphs (or equivalently, monotone properties) for which joint generation in fact provides an incrementally efficient framework for generating all hyperedges. Let us call a Sperner hypergraph $\mathcal{H} = \mathcal{F}_\pi$ (corresponding to a monotone property π) *dual-bounded* if the size of $\mathcal{I}(\mathcal{F}_\pi)$ is (quasi-)polynomially limited in the size of $\mathcal{F}_\pi$, the size of V and the size of the oracle $\|O_\pi\|$ representing π. Let us further call $\mathcal{H} = \mathcal{F}_\pi$ *uniformly dual-bounded*, if

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F}_\pi)| \leq (quasi\text{-})poly(|\mathcal{X}|, |V|, \|\mathcal{O}_\pi\|) \quad (1)$$

holds for any non-empty subfamily $\mathcal{X} \subseteq \mathcal{F}_\pi$.

In the next section we shall recall first from [6, 27] that GEN($\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi), \mathcal{X}, \mathcal{Y}$) can be reduced in polynomial time to dualization (or the generation of hypergraph transversals), which can be solved in quasi-polynomial time (see [25]). As a consequence, problem GEN($\mathcal{F}_\pi, \mathcal{X}$) can be solved in quasi-polynomial time for every subfamily $\mathcal{X} \subseteq \mathcal{F}_\pi$, whenever $\mathcal{F}_\pi$ is uniformly dual-bounded (see [11]). In Section 3 we survey several inequalities proving uniformly dual-boundedness for a number of families of hypergraphs (see [9, 10, 11, 13]). Finally, in Section 4 we recall several applications of the cited results and some generalizations (see [9, 10, 12, 23, 24]).

2 Dualization and Joint Generation

Let $\mathcal{H} \subseteq 2^V$ be a Sperner hypergraph given by an explicit list of its hyperedges. For $X \subseteq V$, let $\pi(X)$ be the property that X is a transversal of $\mathcal{H}$, that is, X intersects all hyperedges of $\mathcal{H}$. Then $\mathcal{F}_\pi = \mathcal{H}^d$ is the transversal hypergraph of $\mathcal{H}$, the above description provides a satisfiability oracle for π of size at most $n|\mathcal{H}|$, and problem GEN($\mathcal{F}_\pi, \mathcal{X}$) reduces to the well-known hypergraph transversal, or *dualization* problem:

DUAL($\mathcal{H}, \mathcal{X}$)**:** *Given two explicitly listed Sperner families $\mathcal{H} \subseteq 2^V$ and $\mathcal{X} \subseteq \mathcal{H}^d$, either find a new minimal transversal $X \in \mathcal{H}^d \setminus \mathcal{X}$ or show that $\mathcal{X} = \mathcal{H}^d$.*

The dualization problem can be efficiently solved for many classes of hypergraphs. For example, if the sizes of all the hyperedges of $\mathcal{H}$ are limited by a constant k, then dualization can be executed in incremental polynomial time (see e.g. [7, 21]), and even stronger, it can be solved efficiently in parallel [8]. In the quadratic case, i.e. when $k = 2$, there are even more efficient algorithms that run with polynomial delay, i.e. in $poly(|V|, |\mathcal{H}|)$ time (see e.g. [32, 33, 52]). Efficient algorithms also exist for the dualization of 2-monotonic, threshold, matroid, read-bounded, acyclic and some other classes of hypergraphs (see e.g. [5, 15, 18, 19, 21, 22, 36, 39, 40, 47, 48]).

Even though no incremental polynomial time algorithm for the dualization of arbitrary hypergraphs is known, an incremental *quasi-polynomial time* one exists (see [25], or [27] for more detail):

Proposition 1 ([25]) *Problem DUAL($\mathcal{H}, \mathcal{X}$) can be solved in $O(nm) + m^{o(\log m)}$ time, where $n = |V|$ and $m = |\mathcal{H}| + |\mathcal{X}|$.*

Remark that this quasi-polynomial time algorithm also indicates that monotone duality can be disproved with limited nondeterminism [22].

It was observed independently in [6, 27] that for any polynomial-time satisfiability oracle O_π problem GEN$(\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi), \mathcal{X}, \mathcal{Y})$ can be reduced in polynomial time to dualization:

Proposition 2 ([6, 27]) *GEN$(\mathcal{F}_\pi, \mathcal{I}(\mathcal{F}_\pi), \mathcal{X}, \mathcal{Y})$ can be solved in time $n(poly(|\mathcal{X}|, |\mathcal{Y}|) + T(|\mathcal{O}_\pi|)) + T_{dual}$, where $T(|\mathcal{O}_\pi|)$ is the worst-case running time of the oracle on any $X \subseteq V$, and T_{dual} denotes the time required to solve the dualization problem with $\mathcal{X}$ and $\mathcal{Y}$.*

From these results it follows easily that joint generation is an incrementally efficient way of generating $\mathcal{F}_\pi$, whenever this hypergraph is uniformly dual bounded:

Corollary 3 ([11]) *If the hypergraph $\mathcal{F}_\pi$ is uniformly dual-bounded, then problem GEN$(\mathcal{F}_\pi, \mathcal{X})$ is solvable in quasi-polynomial time for every subfamily $\mathcal{X} \subseteq \mathcal{F}_\pi$.*

3 Uniformly Dual-Bounded Hypergraphs

All monotone properties considered in this paper, can be described in terms a system of of $r \geq 1$ monotone inequalities

$$f_i(X) \geq t_i, \quad i = 1, \ldots, r, \tag{2}$$

over the subsets $X \subseteq V$ (or more generally, over elements of products of partially ordered sets), where $t_i \in \mathbb{R}$, $i = 1, \ldots, r$, are given thresholds, and $f_i : 2^V \mapsto \mathbb{R}$ are given functions. We assume that each set-function f_i is *monotone*, i.e. $f(X) \leq f(Y)$ whenever $X \subseteq Y$, and is defined via a (quasi-)polynomial-time evaluation oracle. For a subset $X \subseteq V$, let $\pi(X)$ be the property that X satisfies (2), then we have a satisfiability oracle for $\pi(X)$ (this requires only checking the feasibility of X for (2)). Below, we consider four different (but possibly overlapping) classes of functions for which the hypergraph $\mathcal{F}_\pi$ is uniformly dual-bounded.

3.1 Polymatroid functions

A monotone function $f : 2^V \mapsto \mathbb{Z}_+$ is called *polymatroid* if it is submodular, i.e.,

$$f(X \cup Y) + f(X \cap Y) \leq f(X) + f(Y)$$

holds for all subsets $X, Y \subseteq V$, and if $f(\emptyset) = 0$. Here we assume that f is integer-valued.

Theorem 4 ([10]) *Let $\mathcal{F}$ be the set of all minimal feasible solutions for a system of polymatroid inequalities (2) and let $\mathcal{X} \subseteq \mathcal{F}$ be an arbitrary subset of $\mathcal{F}$ of size $|\mathcal{X}| \geq 1$. Then*

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq \max(rn, r|\mathcal{X}|^{(\log t)/c(n,|\mathcal{X}|)}),$$

where $t = \max\{t_1, \ldots, t_r\}$, and $c = c(n, \beta)$ is the unique positive root of the equation $2^c(n^{c/\log \beta} - 1) = 1$. In particular, for $\mathcal{X} = \mathcal{F}$ we get $|\mathcal{I}(\mathcal{F})| \leq \max(rn, r|\mathcal{F}|^{(\log t)/c(n,|\mathcal{F}|)})$.

Let us remark here that the bound of Theorem 4 is reasonably sharp. As shown in [10], for any positive integer k there exists a polymatroid function f over $n = k2^k$ elements, and threshold $t = 2^k$, for which $|\mathcal{F}| = 2^{k^2}$, and $|\mathcal{I}(\mathcal{F})| = k\binom{2^k}{2}$. This gives an infinite family of polymatroid functions such that

$$|\mathcal{I}(\mathcal{F})| \geq |\mathcal{F}|^{(.551 \log t)/c(n,|\mathcal{F}|)},$$

as $k \to \infty$.

Let us also note that for many classes of polymatroid functions, $|\mathcal{F}|$ cannot be bounded by a quasi-polynomial estimate of the form $(n|\mathcal{I}(\mathcal{F})|)^{\text{poly} \log t}$. Let us consider for instance, a graph $G = l \times K_2$ consisting of l disjoint edges, and let $f(X)$ be the number of edges X intersects, for $X \subseteq V(G)$. Then f is a polymatroid function with $n = 2l$. If we let $t = l$ be the given threshold, then $|\mathcal{I}(\mathcal{F})| = l$ and $|\mathcal{F}| = 2^l$. Furthermore, there exist examples for polymatroid inequalities for which the problem of incrementally generating the maximal non-feasible sets (even for a single polymatroid inequality, i.e for $r = 1$) is NP-hard.

3.2 2-Monotonic functions

A monotone function $f : 2^V \mapsto \mathbb{R}$ is called *2-monotonic*, if there exists a permutation $\sigma \in \mathbb{S}_V$ of the ground set V such that $f(X \cup \{v\} \setminus \{u\}) \geq f(X)$ whenever $u \in X \subseteq V$, $v \notin X$ and v precedes u in their σ-order.

Theorem 5 ([9, 18]) *Let $\mathcal{F}$ be the set of all minimal feasible solutions for a system of 2-monotonic inequalities (2) and let $\mathcal{X} \subseteq \mathcal{F}$ be an arbitrary subset of $\mathcal{F}$ of size $|\mathcal{X}| \geq 1$. Then*

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq r \sum_{X \in \mathcal{X}} |X|.$$

In particular, $|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq rn|\mathcal{X}|$, and for $\mathcal{X} = \mathcal{F}$ we get $|\mathcal{I}(\mathcal{F})| \leq rn|\mathcal{F}|$.

Let us remark that the above inequality is reasonably tight, as it can be seen by several examples. It is also interesting to note that if the the number of inequalities r is bounded by a constant, then the size of $\mathcal{F}$ can also be polynomially bounded by the size of $\mathcal{I}(\mathcal{F})$, and both $\mathcal{F}$ and $\mathcal{I}(\mathcal{F})$ can be incrementally generated in polynomial time. On the other hand, if r is not a constant, problem $\mathrm{GEN}(\mathcal{I}(\mathcal{F}), \mathcal{Y})$ is known to be NP-hard [37].

3.3 k-Smooth polymatroid functions

Let $f : 2^V \mapsto \mathbb{Z}_+$ be an integer-valued monotone function. For an integer $k \in \mathbb{Z}_+$, let us say that the function f is *k-smooth* if for any $v \in V$ and any $X \subseteq V$, we have

$$f(X \cup \{v\}) - f(X) \leq k.$$

Theorem 6 *Let $\mathcal{F}$ be the set of all minimal feasible solutions for a system of k-smooth polymatroid inequalities (2) and let $\mathcal{X} \subseteq \mathcal{F}$ be an arbitrary subset of $\mathcal{F}$ of size $|\mathcal{X}| \geq 1$. Then*

$$\sum_{Y \in \mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})} |V \setminus Y| \leq r \sum_{X \in \mathcal{X}} |X||V \setminus X|^{k-1}.$$

In particular, we have $|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq rn^k|\mathcal{X}|$, which for $\mathcal{X} = \mathcal{F}$ gives $|\mathcal{I}(\mathcal{F})| \leq rn^k|\mathcal{F}|$.

It should be mentioned that the above bound is sharp for $r = 1$. On the other hand, there exist examples of 1-smooth polymatroid functions for which the size of $\mathcal{F}$ is exponentially large in the size of $\mathcal{I}(\mathcal{F})$. Despite this, for a single 1-smooth polymatroid inequality, both the sets $\mathcal{F}$ and $\mathcal{I}(\mathcal{F})$ can be generated in polynomial time, and also for constant k, the set $\mathcal{F}$ can be generated in polynomial time. In contrast, it can be shown that generating $\mathcal{I}(\mathcal{F})$ is already NP-hard, for *a single* 2-smooth polymatroid inequality, and for a system of *two* 1-smooth polymatroid inequalities.

3.4 Transversal functions

Given a (not necessarily) Sperner hypergraph $\mathcal{H} \subseteq 2^V$ and a non-negative real weight $w(H) \in \mathbb{R}_+$ associated with each hyperedge $H \in \mathcal{H}$, let us call the function $f : 2^V \mapsto \mathbb{R}_+$, defined by

$$f(X) = \sum\{w(H) \mid X \cap H \neq \emptyset,\ H \in \mathcal{H}\}$$

a *(weighted) transversal function.* Note that such a function is submodular, but not generally polymatroid since the weights are not necessarily integral. In contrast to the bound of Theorem 4, the following stronger bound is known for transversal functions.

Theorem 7 ([13]) *Let $f_1, \ldots, f_r : 2^V \mapsto \mathbb{R}_+$ be r transversal functions defined by r hypergraphs $\mathcal{H}_1, \ldots, \mathcal{H}_r$. Let $\mathcal{F}$ be the set of all minimal feasible solutions for the system of inequalities (2) and let $\mathcal{X} \subseteq \mathcal{F}$ be an arbitrary subset of $\mathcal{F}$ of size $|\mathcal{X}| \geq 1$. Then*

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq \sum_{i=1}^{r} \sum_{X \in \mathcal{X}} |\{H \in \mathcal{H}_i \mid H \cap X \neq \emptyset\}|.$$

In particular, it follows that $|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq r|\mathcal{H}_{max}||\mathcal{X}|$, which for $\mathcal{X} = \mathcal{F}$ gives $|\mathcal{I}(\mathcal{F})| \leq r|\mathcal{H}_{max}||\mathcal{F}|$, where $\mathcal{H}_{max}$ is the hypergraph with largest size.

As it was shown in [13], the bound of Theorem 7 is sharp within a polylogarithmic factor. A similar reverse inequality does not hold, in general, and the incremental generation of $\mathcal{I}(\mathcal{F})$ is NP-hard [12].

4 Applications

In this section we give some specific examples of the four classes of functions described in the previous section. We then conclude by an application in data mining, and a generalization of it over products of partially ordered sets.

4.1 Matroids

Several examples of polymatroid functions can be found, for instance, in [35] and [53]. The most general example considered in this paper is the following. Let M be a matroid with rank function $\rho : 2^M \mapsto |M|$, and let $M_1, \ldots, M_n$ be some subsets of M. For each $X \subseteq V \stackrel{\text{def}}{=} \{1, \ldots, n\}$, let $f(X) = \rho(\bigcup_{i \in X} M_i)$. Then f is a polymatroid function. In fact, every polymatroid function arises by this construction from some matroid, see [31, 44], and also [35]. Thus Theorem 4 and corollary 3 imply that, for a given integer threshold t, the family of minimal subsets of $\{M_1, \ldots, M_n\}$, the rank of whose union is at least t, can be listed incrementally in quasi-polynomial time. Since this result holds for systems of polymatroid inequalities, we can also combine any number of matroid examples.

4.2 Matroid intersections

Let $M_1, \ldots, M_r$ be r matroids on the ground set V of cardinality $|V| = n$. In [33] the question of generating the family $\overline{\mathcal{F}}$ of all maximal sets independent in all the matroids $M_1, \ldots, M_r$ was asked, and an $O(n^{r+2}|\overline{\mathcal{F}}| \sum_{i=1}^{r} T_i)$ algorithm was given, where T_i is the time required for independence testing in matroid M_i. In contrast to this, Theorem 6 together with corollary 3 implies that this problem can be solved in $O(n^2r|\overline{\mathcal{F}}|) + (rn|\overline{\mathcal{F}}|)^{o(\log(rn|\overline{\mathcal{F}}|))}$ oracle time. Indeed, let $\rho_i : 2^V \mapsto \{0, 1, \ldots, n\}$ be the rank function of matroid M_i, for $i = 1, \ldots, r$. Then the function $f_i : 2^V \mapsto \{0, 1, \ldots, n\}$, defined by

$$f_i(X) = \rho_i(V \setminus X) + |X|,$$

for $i = 1, \ldots, r$, is 1-smooth polymatroid, and a set $X \subseteq V$ is independent in M_i if and only if $f_i(V \setminus X) \geq |V|$. Thus letting $\mathcal{F} \stackrel{\text{def}}{=} \{V \setminus X \mid X \in \overline{\mathcal{F}}\}$, we conclude that $\mathcal{F}$ is the family of minimal solutions for the system of 1-smooth polymatroid inequalities

$$f_i(X) \geq |V|, \quad i = 1, \ldots, r,$$

and therefore we get by Theorem 6,

$$\sum_{Y \in \mathcal{I}(\mathcal{F})} |V \setminus Y| \leq r \sum_{X \in \mathcal{F}} |X|.$$

This implies the quasi-polynomial time generation of $\mathcal{F}$.

4.3 Spanning a linear space by linear subspaces

The transversal hypergraph problem is equivalent to the following set covering problem: Given an s-element ground set $\mathcal{R}$ and a family $\mathcal{V}$ of n subsets of $\mathcal{R}$, enumerate all minimal subfamilies of $\mathcal{V}$ which cover the entire set $\mathcal{R}$. Replacing $\mathcal{R}$ by the vector space $\mathbf{F}^s$ over some field $\mathbf{F}$, and replacing each given subset of $\mathcal{V}$ by a linear subspace of $\mathbf{F}^s$, we arrive at the following *space covering* problem:

> *Given a collection* $\mathcal{V} = \{\mathcal{V}_1, \ldots, \mathcal{V}_n\}$ *of n linear subspaces of* $\mathbf{F}^s$, *enumerate all minimal subsets X of* $V = \{1, \ldots, n\}$ *such that* $\mathrm{Span}\langle \bigcup_{i \in X} \mathcal{V}_i \rangle = \mathbf{F}^s$.

Generalizing further, consider the polymatroid inequality

$$f(X) = \dim(\bigcup_{i \in X} \mathcal{V}_i) \geq t, \tag{3}$$

where $t \in \{1, \ldots, s\}$ is a given threshold. Then the set $\mathcal{F}$ of minimal solutions to (3) is the collection of all minimal subsets of $\mathcal{V}$ the dimension of whose union is at least t. Theorem 4 then states that for all $t \in \{1, \ldots, s\}$, the size of $\mathcal{I}(\mathcal{F})$ can be bounded by a $\log t$-degree polynomial in n and $|\mathcal{F}|$, and thus all sets in $\mathcal{F}$ can be enumerated in incremental quasi-polynomial time.

We mention here two special cases of the above example. First, when each subspace $\mathcal{V}_i$ is spanned by a subset R_i of vectors from some fixed basis of $\mathbf{F}^s$, the value of $f(X)$ is just the size of $\bigcup_{i \in X} R_i$, which is a transversal function. Hence, by Theorem 7, we get the stronger inequality $|\mathcal{I}(\mathcal{F})| \leq s|\mathcal{F}|$.

Second, when the dimension of each input subspace $\mathcal{V}_i$, $i = 1, \ldots, n$, is bounded by some constant d, the function $f(X)$ is d-smooth. We can thus enumerate all sets in $\mathcal{F}$ in incremental polynomial time and the size of $\mathcal{I}(\mathcal{F})$ can be bounded by a d-degree polynomial in n and $|\mathcal{F}|$. In particular, when $t = s$ and the given subspaces $\mathcal{V}_i$ are all lines, i.e.,

$$\mathcal{V}_i = \mathrm{Span}\langle b_i \rangle \text{ for given vectors } b_i \in \mathbf{F}^s, \quad i = 1, \ldots, n,$$

then the set $\mathcal{F}$ of all column bases of the $s \times n$ matrix $[b_1, \ldots, b_n]$ can be enumerated in polynomial time.

4.4 Spanning collections of graphs

Let R be a finite set of s vertices and let $E_1, \ldots, E_n \subseteq R \times R$ be a collection of n graphs on R. Given a set $X \subseteq \{1, \ldots, n\}$ define $k(X)$ to be the number of connected components in the graph $(R, \bigcup_{i \in X} E_i)$. Then $k(X)$ is an anti-monotone supermodular function and hence for any integral threshold t, the inequality

$$f(X) = s - k(X) \geq t$$

is polymatroid. In particular, $\mathcal{F}$ is the family of all minimal collections of the input graphs $E_1, \ldots, E_n$ which interconnect all vertices in R. (If the n input graphs are just n disjoint edges, then $\mathcal{F}$ is the set of all spanning trees in the graph $E_1 \cup \ldots \cup E_n$.) Since $k(X)$ can be evaluated at any set X in polynomial time, Theorem 4 and Corollary 3 imply that for each $t \in \{1, \ldots, s\}$, all elements of $\mathcal{F}$ can be enumerated in incremental quasi-polynomial time. In particular, given a collection of n *equivalence relations* (partitions) on R, we can enumerate in incremental quasi-polynomial time all minimal subsets of the given relations whose transitive closure puts all elements of R in one equivalence class (or produces at most $s-t$ equivalence classes). Note that this example is a special case of the space covering problem, where $\mathbf{F} = \mathbf{GF}(2)$ and each subspace $\mathcal{V}_i$ is the span of the incidence vectors of the edges of the set E_i, for $i = 1, \ldots, n$.

As before, generating all maximal collections of $E_1, \ldots, E_n$ for which the number of connected components of $(R, \bigcup_{i \in X} E_i)$ exceeds a given threshold remains NP-hard.

4.5 Monotone systems of Boolean and integer inequalities

Consider a system of r linear inequalities in n binary variables

$$a_i^T x \geq b_i, \quad x \in \{0, 1\}^n, \quad i = 1, \ldots, r \tag{4}$$

where a_i is a given n-vector and b_i is a given scalar, for $i = 1, \ldots, r$. We assume that the system (4) is monotone: if $x \in \{0,1\}^n$ is feasible for (4) then any vector $y \in \{0,1\}^n$, $y \geq x$ is also feasible. (For instance, (4) is monotone if the vectors a_i are non-negative.) Let V be the set $\{1, \ldots, n\}$, and for $X \subseteq V$, let $x \in \{0,1\}^n$ be the incidence vector of X, and define $f_i(X) = a_i^T x$. Then f_i is 2-monotonic with respect to the permutation that puts the components of a_i in non-increasing order. Accordingly, from Theorem 5 we conclude that for any non-empty set $\mathcal{X} \subseteq \mathcal{F}$ of minimal feasible solutions to (4) we have the inequalities

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F})| \leq r \sum_{x \in \mathcal{X}} p(x) \leq nr|\mathcal{X}|, \qquad (5)$$

where $p(x)$ is number of positive components of x. In particular, for any feasible system (4) we have

$$|\mathcal{I}(\mathcal{F})| \leq nr|\mathcal{F}|. \qquad (6)$$

The above bounds are sharp when $r = 1$, for instance, for the inequality $x_1 + \ldots + x_n \geq n$. For large r, these bounds are accurate up to a factor polylogarithmic in r. For instance, for any positive integer k, the system of $r = 2^k$ inequalities in $n = 2k$ binary variables

$$x_{i_1} + \cdots + x_{i_k} \geq 1, \; i_1 \in \{1,2\}, \ldots, i_k \in \{2k-1, 2k\}$$

has 2^k maximal infeasible binary vectors and only k minimal feasible binary vectors, i.e.,

$$|\mathcal{I}(\mathcal{F})| = \frac{nr}{2(\log r)^2}|\mathcal{F}|.$$

As it was shown in [9], inequalities (5) and (6) also hold for any monotone system of r linear inequalities in n *integer* variables

$$a_i^T x \geq b_i, \; x \in \mathcal{C} \stackrel{\text{def}}{=} \{x \in \mathbb{Z}^n \mid 0 \leq x \leq c\}, \; i = 1, \ldots, r \qquad (7)$$

where c is a given n vector some or all components of which may be infinite. Furthermore, all minimal feasible integer solutions to a given monotone system of integer inequalities (7) can be generated in incremental quasi-polynomial time [9, 13], which should be contrasted with the conjecture of [33] that for non-negative a_i's and $c = (1, \ldots, 1)$ this problem cannot be solved in incremental polynomial time unless P=NP. On the other hand, the problem of generating all maximal infeasible binary vectors for (4) is NP-hard already for binary vectors a_i and $b_i = 2$ for $i = 1, \ldots, r$, see [37] and also [11] for more detail.

4.6 Maximal frequent and minimal infrequent sets for binary matrices

The notion of frequent sets in data-mining [49] can be related naturally to the transversal functions considered in Section 3.4. Let V be a finite set of binary attributes of a database. Let $\mathcal{D} : \mathcal{R} \times V \mapsto \{0,1\}$ be a given $s \times n$ binary matrix representing a set $\mathcal{R}$ of transactions over V. To each subset of columns $X \subseteq V$, we associate the subset $S(X) = S_{\mathcal{D}}(X) \subseteq \mathcal{R}$ of all those rows $i \in \mathcal{R}$ for which $\mathcal{D}(i,j) = 1$ in every column $j \in X$. The cardinality of $S(X)$ is called the *support* of X and it is easy to see that the function

$$f(X) = s - |S(X)|$$

is a transversal function with respect to the hypergraph $\mathcal{H}$ defined by the anti-incidence matrix $\mathcal{D}$.

A column set $X \subseteq V$ is called *t-frequent* if $|S(X)| \geq t$ and otherwise, is said to be *t-infrequent*. Thus the families $\mathcal{F}$ and $\mathcal{I}(\mathcal{F})$ of minimal feasible and maximal infeasible sets for the inequality $f(X) \geq s - t + 1$ correspond, respectively, to the minimal infrequent and maximal frequent sets for $\mathcal{D}$.

The generation of (maximal) frequent sets of a given binary matrix is an important task of knowledge discovery and data mining, e.g. it is used for mining association rules [1, 26, 41, 42, 45, 46, 54], correlations [16], sequential patterns [3], episodes [43], emerging patterns [20], and appears in many other applications. Most practical procedures to generate frequent sets are based on the anti-monotone *Apriori* heuristic (see [2]) and build frequent sets in a bottom-up way, running in time proportional to the number of frequent sets. It was demonstrated recently in [17] that these methods are inadequate in practice when there are (many) frequent sets of large size (see also [4, 30, 34]), due to the fact that the number of frequent sets can be exponentially larger than $|\mathcal{I}(\mathcal{F})|$. Thus these results show that it is perhaps more efficient to find the *boundary* of the frequent sets, i.e. the union $\mathcal{F} \cup \mathcal{I}(\mathcal{F})$ (proposed e.g. in [49]), and use it as a condensed representation of the data set, as suggested in [41]. Furthermore, no algorithm using membership queries "Is X frequent?" can generate all (maximal) frequent sets in fewer than $|\mathcal{F} \cup \mathcal{I}(\mathcal{F})|$ steps (see e.g. [26]). There were several other examples presented in [41] to show the usefulness of maximal frequent sets and minimal infrequent sets, e.g. providing error bounds for the confidence of an arbitrary Boolean rule, in terms of minimal infrequent sets.

It follows by the results of Section 3.4 that the hypergraph of minimal infrequent sets is dual bounded:

$$|\mathcal{I}(\mathcal{F})| \leq (s - t + 1)|\mathcal{F}| \qquad (8)$$

for all $t \in \{1, \ldots, s\}$. Let us note again that these inequalities are best possible. For instance, they are sharp when $\mathcal{D}$ is an $s \times (s-t+1)$ matrix in which every entry is 1, except the diagonal entries in the first $s-t+1$ rows, which are 0. In addition, (8) stays accurate, up to a factor of $\log s$, even when $s \gg n$ and $|\mathcal{F}|$ and $|\mathcal{I}(\mathcal{F})|$ are arbitrarily large. Let us consider for instance a binary matrix $\mathcal{D}$ with $s = 2^k$ rows and $n = 2k$ columns ($k \geq 1$, integer), such that each row contains exactly one 0 and one 1 in each pair of the adjacent columns $\{1, 2\}$, $\{3, 4\}$, $\ldots$, $\{2k-1, 2k\}$, and in all 2^k possible ways in the $s = 2^k$ rows. It is not difficult to see that for $t = 1$ there are 2^k maximal 1-frequent sets (every row of the matrix is the characteristic vector of a maximal 1-frequent set), and that there are only k minimal 1-infrequent sets, namely $\{2i-1, 2i\}$ for $i = 1, \ldots, k$. Thus for such examples we have $|\mathcal{I}(\mathcal{F})| = (s/\log s)|\mathcal{F}|$. The same examples also show that $|\mathcal{F}|$ cannot be bounded by a quasi-polynomial in $|\mathcal{I}(\mathcal{F})|$, n and s. Indeed, it was shown in [12, 38] that problem $\mathrm{GEN}(\mathcal{I}(\mathcal{F}), \mathcal{Y})$ is NP-hard even if $|\mathcal{Y}| = O(n)$ and $|\mathcal{I}(\mathcal{F})|$ is exponentially large in n.

Remark. Clearly, if the number of non-zero entries in each row of the matrix $\mathcal{D}$ is bounded by some constant k, then the number of frequent sets is bounded by n^k and therefore, they can be even listed efficinely in parallel. It is interesting to note that if the number of non-zero entries in each row of $\mathcal{D}$ is at least $n-k$ for some constant k, then for any constant t both maximal t-frequnet and minimal t-infrequent sets can be also generated efficiently in parallel, see [8] for more details.

4.7 Maximal frequent and minimal infrequent elements in products of partially ordered sets

Let us finally consider a natural generalization of frequent sets to databases defined over products of partially ordered sets (posets). Specifically, let $\mathcal{P} \stackrel{\mathrm{def}}{=} \mathcal{P}_1 \times \ldots \times \mathcal{P}_n$ be the product of n posets. Let us use $\preceq$ to denote the precedence relation in $\mathcal{P}$ and also in $\mathcal{P}_1, \ldots, \mathcal{P}_n$, i.e., if $p = (p_1, \ldots, p_n) \in \mathcal{P}$ and $q = (q_1, \ldots, q_n) \in \mathcal{P}$, then $p \preceq q$ in $\mathcal{P}$ if and only if $p_i \preceq q_i$ in $\mathcal{P}_i$, $i = 1, \ldots, n$. For $\mathcal{A} \subseteq \mathcal{P}$, denote by $\mathcal{A}^+ = \{x \in \mathcal{P} \mid x \succeq a, \text{ for some } a \in \mathcal{A}\}$ and $\mathcal{A}^- = \{x \in \mathcal{P} \mid x \preceq a, \text{ for some } a \in \mathcal{A}\}$, the filter and ideal generated by $\mathcal{A}$, respectively. Any element in $\mathcal{P} \setminus \mathcal{A}^+$ is called *independent of* $\mathcal{A}$. Let $\mathcal{I}(\mathcal{A})$ be the set of all maximal independent elements for $\mathcal{A}$:

$$\mathcal{I}(\mathcal{A}) \stackrel{\mathrm{def}}{=} \{p \in \mathcal{P} \mid p \notin \mathcal{A}^+ \text{ and } (q \in \mathcal{P}, q \succeq p, q \neq p \;\Rightarrow\; q \in \mathcal{A}^+)\}.$$

Consider a database $\mathcal{D} \subseteq \mathcal{P}$ of transactions, each of which is an n-dimensional vector of attribute values over $\mathcal{P}$. For an element $p \in \mathcal{P}$, let us denote by

$$S(p) = S_{\mathcal{D}}(p) \stackrel{\mathrm{def}}{=} \{q \in \mathcal{D} \mid q \succeq p\},$$

the set of transactions in $\mathcal{D}$ that *support* $p \in \mathcal{P}$. Note that, by this definition, the function $|S(\cdot)| : \mathcal{P} \mapsto \{0, 1, \ldots, |\mathcal{D}|\}$ is an anti-monotone function, i.e., $|S(p)| \leq |S(q)|$, whenever $p \succeq q$.

Given $\mathcal{D} \subseteq \mathcal{P}$ and an integer threshold t, let us say that an element $p \in \mathcal{P}$ is t-frequent if it is supported by at least t transactions in the database, i.e., if $|S_{\mathcal{D}}(p)| \geq t$. Conversely, $p \in \mathcal{P}$ is said to be t-infrequent if $|S_{\mathcal{D}}(p)| < t$. Since the function $|S_{\mathcal{D}}(\cdot)|$ is anti-monotone, we may restrict our attention only to *maximal frequent* and *minimal infrequent* elements. Denote by $\mathcal{F} = \mathcal{F}_{\mathcal{D},t}$ the set of all minimal t-infrequent elements of $\mathcal{P}$ with respect to the database $\mathcal{D}$. Then $\mathcal{I}(\mathcal{F}_{\mathcal{D},t})$ is the set of all maximal t-frequent elements.

As in the Boolean case, the separate and joint generation of maximal frequent and minimal infrequent elements of a database can be used for finding association rules in data mining applications. If the database $\mathcal{D}$ contains categorical (e.g., zip code, make of car), or quantitative (e.g., age, income) attributes, and the corresponding posets $\mathcal{P}_i$ are total orders, then the above generation problems can be used to mine the so called *quantitative* association rules [51]. More generally, each attribute a_i in the database can assume values belonging to some partially ordered set $\mathcal{P}_i$. For example, [50] describes applications where items in the database belong to sets of *taxonomies* (or *is-a hierarchies*), and proposes several algorithms for mining association rules among these hierarchical data (see also [28, 29]). Furthermore, many data analysis applications assume data values ranging over lattices of small size, see e.g. [14].

Extending the notion of uniformly dual-bounded hypergraphs to subsets of products of posets in the obvious way, it can be shown that if each poset $\mathcal{P}_i$ is a join semi-lattice (i.e. for every two elements $x, y \in \mathcal{P}_i$, there is a unique upper bound $x \vee y$), then for any integer threshold t, the set of minimal infrequent elements $\mathcal{F}_{\mathcal{D},t}$ of a database $\mathcal{D}$ is uniformly dual-bounded:

$$|\mathcal{I}(\mathcal{X}) \cap \mathcal{I}(\mathcal{F}_{\mathcal{D},t})| \leq (|\mathcal{D}| - t + 1)|\mathcal{X}|$$

for any nonempty $\mathcal{X} \subseteq \mathcal{F}_{\mathcal{D},t}$. It is also easily verified that for such uniformly dual-bounded subsets of a poset $\mathcal{P}$, the incremental generation of these subsets reduces in polynomial time to the following natural generalization of the hypergraph transversal problem on products of posets:

DUAL$(\mathcal{P}, \mathcal{A}, \mathcal{B})$**:** *Given a subset $\mathcal{A} \subseteq \mathcal{P}$ in a poset $\mathcal{P}$ and a collection of maximal independent elements $\mathcal{B} \subseteq \mathcal{I}(\mathcal{A})$, either find a new maximal independent element $p \in \mathcal{I}(\mathcal{A}) \setminus \mathcal{B}$, or prove that $\mathcal{A}$ and $\mathcal{B}$ form a dual pair: $\mathcal{B} = \mathcal{I}(\mathcal{A})$.*

The complexity of the dualization problem for products of general posets is still an open question. Two special cases are known to be quasi-polynomially solvable [23, 24], namely when the underlying precedence graph of each poset $\mathcal{P}_i$ is a rooted tree (is-a hierarchy), and when each poset $\mathcal{P}_i$ a join semi-lattice with constant width (where the width of $\mathcal{P}_i$ is the maximum size of an antichain in $\mathcal{P}_i$). It follows then that generating minimal infrequent elements in these classes of posets can be done in incremental quasi-polynomial time.

References

[1] R. Agrawal, T. Imielinski and A. Swami. Mining associations between sets of items in massive databases, *Proc. 1993 ACM-SIGMOD International Conference on Management of Data*, pp. 207-216.

[2] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A. I. Verkamo, Fast discovery of association rules, in U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy eds., *Advances in Knowledge Discovery and Data Mining*, 307-328, AAAI Press, Menlo Park, California, 1996.

[3] R. Agrawal and R. Srikant, Mining sequential patterns, *Proc. 11th International Conference on Data Engineering, 1995*, pp.3-14.

[4] R.J. Bayardo, Efficiently mining long patterns from databases, *Proc. the 1998 ACM-SIGMOD International Conference on Management of Data*, pp. 85-93.

[5] P. Bertolazzi and A. Sassano, An $O(mn)$ time algorithm for regular set-covering problems, *Theoretical Computer Science* 54 (1987) 237-247.

[6] J. C. Bioch and T. Ibaraki, Complexity of identification and dualization of positive Boolean functions, *Information and Computation* 123 (1995) 50-63.

[7] E. Boros, V. Gurvich, and P.L. Hammer, Dual subimplicants of positive Boolean functions, *Optimization Methods and Software*, 10 (1998) 147-156.

[8] E. Boros, K. Elbassioni, V. Gurvich and L. Khachiyan, An efficient incremental algorithm for generating all maximal independent sets in hypergraphs of bounded dimension, *Parallel Processing Letters*, 10 (2000) 253-266.

[9] E. Boros, K. Elbassioni, V. Gurvich, L. Khachiyan and K. Makino, Generating all minimal integer solutions to a monotone system of linear inequalities, *Automata, Languages and Programming, 28-th International Colloquium, ICALP 2001* (F. Orejas, P.G. Spirakis and Jan van Leeuwen, eds.), *Lecture Notes in Computer Science* 2076, pp. 92-103, 2001.

[10] E. Boros, K. Elbassioni, V. Gurvich and L. Khachiyan, An inequality for polymatroid functions and its applications, to appear in *Discrete Applied Mathematics*, 2002. (DIMACS Technical Report 2001-14, Rutgers University.)

[11] E. Boros, V. Gurvich, L. Khachiyan and K. Makino, Dual-bounded generating problems: partial and multiple transversals of a hypergraph, *SIAM J. Comput.*, 30 (2001) 2036-2050.

[12] E. Boros, V. Gurvich, L. Khachiyan and K. Makino, On the complexity of generating maximal frequent and minimal infrequent sets, to appear in *19th International Symposium on Theoretical Aspects of Computer Science, (STACS 2002)* (H. Alt and A. Ferreira, eds.), 2002.

[13] E. Boros, V. Gurvich, L. Khachiyan and K. Makino, Generating weighted transversals of a hypergraph, in *Proc. 2nd Japanese-Hungarian Symposium on Discrete Mathematics and Its Applications*, pp. 13-22. Hungarian Academy of Sciences, Budapest, Hungary, April 20-23, 2001.

[14] E. Boros, P.L. Hammer and J.N. Hooker, Predicting cause-effect relationships from incomplete discrete observations, *SIAM J. Discrete Math.*, 7 (1994), 481-491.

[15] E. Boros, P. L. Hammer, T. Ibaraki and K. Kawakami, Polynomial time recognition of 2-monotonic positive Boolean functions given by an oracle, *SIAM J. Comput.*, 26 (1997) 93-109.

[16] S. Brin, R. Motwani, and C. Silverstein, Beyond market basket: Generalizing association rules to correlations, *Proc. the 1997 ACM-SIGMOD Conference on Management of Data*, pp. 265-276.

[17] S. Brin, R. Motwani, J. Ullman, and S. Tsur, Dynamic itemset counting and implication rules for market basket data, *Proc. the 1997 ACM-SIGMOD Conference on Management of Data*, pp. 255-264.

[18] Y. Crama, Dualization of regular Boolean functions, *Discrete Applied Mathematics*, 16 (1987) 79-85.

[19] C. Domingo, N. Mishra, and L. Pitt, Efficient read-k monotone CNF/DNF dualization using a learning with membership queries approach, *Machine Learning*, 37 (1999) 89-110.

[20] G. Dong and J. Li, Efficient mining of emerging patterns, *Proc. the 1999 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 43-52.

[21] T. Eiter and G. Gottlob, Identifying the minimal transversals of a hypergraph and related problems, *SIAM J. Comput.*, 24 (1995) 1278-1304.

[22] T. Eiter, G. Gottlob, K. Makino, New results on monotone dualization and generating hypergraph transversals, to appear in *the 34th Annual ACM Symposium on Theory of Computing (STOC 2002).*

[23] K. Elbassioni, On dualization in products of forests, to appear in the *19th International Symposium on Theoretical Aspects of Computer Science (STACS 2002)*, (H. Alt and A. Ferreira, eds.), March 14-16, 2002, Antibes, Juan les pins, France.

[24] K. Elbassioni, Dualization in products of bunded-width lattices, Manuscript.

[25] M. L. Fredman and L. Khachiyan, On the complexity of dualization of monotone disjunctive normal forms, *Journal of Algorithms*, 21 (1996) 618-628.

[26] D. Gunopulos, R. Khardon, H. Mannila, and H. Toivonen, Data mining, hypergraph transversals and machine learning, *Proc. the 16th ACM-SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, (1997) pp. 12-15.

[27] V. Gurvich and L. Khachiyan, On generating the irredundant conjunctive and disjunctive normal forms of monotone Boolean functions, *Discrete Applied Mathematics*, 96-97 (1999) 363-373.

[28] J. Han, Y. Cai and N. Cercone, Data driven discovery of quantitative rules in relational databases, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 5 No. 1, (1993) pp. 29-40.

[29] J. Han and Y. Fu, Discovery of multiple-level association rules from large databases, *Proc. 21st International Conference on Very Large Data Bases*, pp. 420-431, 1995.

[30] J. Han, J. Pei, and Y. Yin, Mining frequent patterns without candidate generation, *Proc. the 2000 ACM-SIGMOD Conference on Management of Data*, pp. 1-12.

[31] T. Helgason, Aspects of the theory of hypermatroids, in Hypergraph Seminar, eds C. Berge and D.K. Ray-Chaudhuri), Lecture Notes in Math. 411 (1975) Springer, 191-214.

[32] D. S. Johnson, M. Yannakakis and C. H. Papadimitriou, On generating all maximal independent sets, *Information Processing Letters*, 27 (1988) 119-123.

[33] E. Lawler, J. K. Lenstra and A. H. G. Rinnooy Kan, Generating all maximal independent sets: NP-hardness and polynomial-time algorithms, *SIAM J. Comput.*, 9 (1980) 558-565.

[34] D. Lin and Z.M. Kedem, Pincer-search: a new algorithm for discovering the maximum frequent set, *Proc. 6th European Conference on Extending Database Technology*, to appear.

[35] L. Lovász, Submodular functions and convexity, in *Mathematical Programming: The State of the Art*, Bonn 1982, pp. 235-257, (Springer Verlag, 1983).

[36] K. Makino, Efficient dualization of $O(\log n)$-term monotone disjunctive normal forms, to appear in *Discrete Applied Mathematics*. (Technical Report 00-07, Discrete Mathematics and Systems Science, Osaka University, 2000.)

[37] K. Makino and T. Ibaraki, Interior and exterior functions of Boolean functions, *Discrete Applied Mathematics*, 69 (1996) 209-231.

[38] K. Makino and T. Ibaraki, Inner-core and outer-core functions of partially defined Boolean functions, *Discrete Applied Mathematics*, 96-97, (1999), 307-326.

[39] K. Makino and T. Ibaraki. The maximum latency and identification of positive Boolean functions. *SIAM J. Comput.*, 26 (1997) 1363–1383.

[40] K. Makino and T. Ibaraki, A fast and simple algorithm for identifying 2-monotonic positive Boolean functions, *Journal of Algorithms*, 26 (1998) 291-305.

[41] H. Mannila and H. Toivonen, Multiple uses of frequent sets and condensed representations, *Proc. 2nd International Conference on Knowledge Discovery and Data Mining*, (1996) pp. 189-194.

[42] H. Mannila and H. Toivonen, Levelwise search and borders of theories in knowledge discovery. Series of Publications C C-1997-8, University of Helsinki, Department of Computer Science (1997).

[43] H. Mannila, H. Toivonen, and A. I. Verkamo. Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, 1 (1997), 259-289.

[44] C.G.H. McDiarmid, Rado's theorem for polymatroids, *Math. Proc. Cambridge Phil. Soc.* 78 (1975) 263-281.

[45] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, Discovering frequent closed itemsets for association rules. *Proc. of the 7th ICDT Conference*, Jerusalem, Israel, January 10-12, 1999; *Lecture Notes in Computer Science*, 1540, pp. 398-416, Springer Verlag, 1999.

[46] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, Closed set based discovery of small covers for association rules, *Proc. 15emes Journees Bases de Donnees Avancees, BDA*, pp. 361-381, 1999.

[47] U. N. Peled and B. Simeone, Polynomial-time algorithm for regular set-covering and threshold synthesis, *Discrete Applied Mathematics* 12 (1985) 57-69.

[48] U. N. Peled and B. Simeone, An $O(nm)$-time algorithm for computing the dual of a regular Boolean function, *Discrete Applied Mathematics* 49 (1994) 309-323.

[49] R. H. Sloan, K. Takata, G. Turan, On frequent sets of Boolean matrices, *Annals of Mathematics and Artificial Intelligence* 24 (1998) 1-4.

[50] R. Srikant and R. Agrawal, Mining generalized association rules, *Proc. 21st International Conference on Very Large Data Bases*, pp. 407-419, 1995.

[51] R. Srikant and R. Agrawal, Mining quantitative association rules in large relational tables, *Proc. of the ACM-SIGMOD 1996 Conference on Management of Data*, pp. 1-12, 1996.

[52] S. Tsukiyama, M. Ide, H. Ariyoshi and I. Shirakawa, A new algorithm for generating all maximal independent sets, *SIAM J. Comput.*, 6 (1977) 505-517.

[53] D.J.A. Welsh, Matroid Theory (Academic Press, London, New York, San Francisco 1976).

[54] M.J. Zaki and M. Ogihara, Theoretical foundations of association rules, *3rd SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, June 1998.