ECAI 2012
Luc De Raedt et al. (Eds.)

doi:10.3233/978-1-61499-098-7-630

Extending Set-based Dualization: Application to Pattern Mining

Lhouari Nourine[1] and **Jean-Marc Petit**[2]

Abstract. Dualization problems have been intensively studied in combinatorics, AI and pattern mining for years. Roughly speaking, for a partial order $(P, \preceq)$ and some monotonic predicate Q over P, the dualization consists in identifying all maximal elements of P verifying Q from all minimal elements of P not verifying Q, and vice versa. The dualization is equivalent to the enumeration of minimal transversal of hypergraphs whenever $(P, \preceq)$ is a boolean lattice. In the setting of interesting pattern mining in databases, P represents a set of patterns and whenever P is isomorphic to a boolean lattice, the pattern mining problem is said to be *representable as sets*. The class of such problems is denoted by $\mathcal{RAS}$.
In this paper, we introduce a *weak representation as sets* for pattern mining problems which extends the $\mathcal{RAS}$ class to a wider and significantly larger class of problems, called $\mathcal{WRAS}$. We also identify $\mathcal{EWRAS}$, an *efficient* subclass of $\mathcal{WRAS}$ for which the dualization problem is still quasi-polynomial. Finally, we point out that one representative pattern mining problem known not to be in $\mathcal{RAS}$, namely *frequent rigid sequences with wildcard*, belongs to $\mathcal{EWRAS}$. These new classes might prove to have large impact in unifying existing pattern mining approaches.

1 Introduction

Dualization problems have been intensively studied in combinatorics, IA and pattern mining for years [4, 8, 13, 9, 10, 14, 18]. Many applications exist in Logic, Artificial Intelligence [8], databases and data mining [19] such as satisfiability checking, dualization of boolean functions, inclusion dependencies and maximal frequent itemsets. Roughly speaking, for a partial order $(P, \preceq)$ and some monotonic predicate Q over P, the dualization consists in identifying all maximal elements of P verifying Q from all minimal elements of P not verifying Q, and vice versa.

In the setting of pattern mining in databases, many problems have been studied over the last decade, from (maximal, closed, non-redundant) frequent itemsets in transactional databases to frequent sub-graphs in a collection of graphs or satisfied inclusion dependencies in databases to mention a few. From a theoretical point of view, we argue that the domain has not been deeply investigated and much work remains to be done to better understand the domain of interesting pattern mining itself [19, 6, 16, 2, 5]. For instance, many problems appear to be quite different in appearance but turn out to be equivalent from an algorithmic point of view.

The seminal work of [19] proposes a general framework for interesting pattern mining problems in databases. One of their results was to identify the class $\mathcal{RAS}$ of pattern mining problems *representable as sets*, i.e. those pattern mining problems isomorphic to some boolean lattices. As a consequence, the dualization problem for this class of problems is well understood: it turns out to be equivalent to the minimal transversal problem of a hypergraph [13]. Many complex pattern mining problems belong to $\mathcal{RAS}$ which means that they essentially reduce to itemset mining.

Paper contribution In this paper, we introduce a *weak representation as sets* for pattern mining problems which extends the $\mathcal{RAS}$ class to a wider and significantly larger class of problems, called $\mathcal{WRAS}$. The key features of such a weak representation are: (1) the encoding f of patterns into sets and the decoding g of sets into patterns are two different functions such that $g(f(\theta)) = \theta$ for all pattern θ ; (2) the encoding preserves the incomparability of patterns, i.e. $\theta \npreceq \varphi \Rightarrow f(\theta) \nsubseteq f(\varphi)$.
We also identify an *efficient* subclass $\mathcal{EWRAS}$ of $\mathcal{WRAS}$ for which the dualization problem is still quasi-polynomial.
Finally, we point out that one representative pattern mining problem – known not to be in $\mathcal{RAS}$– namely *frequent rigid sequences with wildcard*, belongs to $\mathcal{EWRAS}$. Interestingly and to the best of our knowledge, it is the first time that an incremental quasi-polynomial algorithm for the dualization of rigid sequences with wildcard is proposed. These new classes might prove to have large impact in unifying existing pattern mining approaches.

Related work Some theoretical frameworks for pattern mining have been proposed in the past, e.g. [19, 6, 16, 2, 5, 15]. The theoretical framework of [19, 16] proposes both complexity results and algorithms for pattern mining problems. They defined the class of representable as sets problems reused (and extended) in this paper. An extension of the tuple relational calculus for pattern mining problems has been proposed in [6]. Their objective was to characterize data mining queries that reduce to a levelwise search strategy. They obtain negative results pointing out that their class of queries was too expressive. Efficient closed pattern mining has been studied in [2] in a general framework based on *accessible set systems*. The relationship between accessible set systems and pattern mining problems has been pointed out in [5]. The dualization problem has been studied by many researchers, among which we quote [8, 13, 10]. Moreover, a large bunch of techniques developed for dualization in combinatorics and pattern mining could be re-used – almost for free – for the classes of pattern mining problems introduced in this paper [7, 12] .

[1] Université Blaise Pascal, CNRS, LIMOS, France
[2] Université de Lyon, CNRS, LIRIS, INSA Lyon, France

Paper organization: Section 2 introduces the necessary material for the rest of the paper. Weak representation as sets of pattern mining problems is given in Section 3 to extend the dualization problem to pattern sets not isomorphic to a boolean lattice. Efficient weak representations of sets are discussed in Section 4. In Section 5, we show that the frequent rigid sequences with wildcard problem belongs to $\mathcal{EWRAS}$. Discussion and concluding remarks are made in Section 6.

2 Preliminaries

We first recall the framework of Mannila and Toivonen [19] for pattern mining problems. Then we point out that the dualization plays a crucial role for pattern mining and give some known results on arbitrary structures. We finally give the class of problems representable as sets for which the dualization can be reduced to itemset mining and therefore to hypergraph transversal.

2.1 A framework for pattern mining problems

Given a database $\mathcal{D}$, a finite language $\mathcal{L}$ of patterns, and a predicate Q for evaluating whether a pattern $\varphi \in \mathcal{L}$ is true or "interesting" in $\mathcal{D}$, the discovery task is to find the theory of $\mathcal{D}$ with respect to $\mathcal{L}$ and Q, i.e. the set $Th(\mathcal{L}, \mathcal{D}, Q) = \{\varphi \in \mathcal{L} | Q(\mathcal{D}, \varphi) \text{ is true}\}$.
We suppose that the set of patterns $\mathcal{L}$ is structured with a partial order $\preceq$, known as a specialization/generalization relation between patterns of $\mathcal{L}$. We note $\theta \prec \varphi$ if $\theta \preceq \varphi$ and $\theta \neq \varphi$.
The predicate Q is monotonic wrt $\preceq$ if for all $\theta, \varphi \in \mathcal{L}, \varphi \preceq \theta$, we have $Q(\mathcal{D}, \theta) \Rightarrow Q(\mathcal{D}, \varphi)$.

In the sequel and when clear from context, we will use $\mathcal{L}$ instead of $(\mathcal{L}, \preceq)$, Q will be supposed to be a monotonic predicate and a pattern mining problem will be denoted by a triple $(\mathcal{L}, \mathcal{D}, Q)$. Given $(\mathcal{L}, \mathcal{D}, Q)$, the set $Th(\mathcal{L}, \mathcal{D}, Q)$ can now be represented by its maximal elements only.

A set $S \subseteq \mathcal{L}$ of patterns is downward (resp. upward) closed under the relation $\preceq$ if for all $\phi \preceq \theta, \theta \in S \Rightarrow \phi \in S$ (resp. $\phi \preceq \theta, \phi \in S \Rightarrow \theta \in S$)[3]. For any set of patterns S, we shall denote by $\downarrow S$ (resp. $\uparrow S$) the downward (resp. upward) closed set of S under the relation $\preceq$.

A downward closed set S can be represented by two borders: the *positive border* of S, denoted by $\mathcal{B}d^+(S)$, and the *negative border* of S, denoted by $\mathcal{B}d^-(S)$. They are defined as follows:

$$\mathcal{B}d^+(S) = \{\sigma \in S \mid \nexists \varphi \in S, \sigma \prec \varphi\}$$

$$\mathcal{B}d^-(S) = \{\sigma \in \mathcal{L} \setminus S \mid \nexists \varphi \in \mathcal{L} \setminus S, \varphi \prec \sigma\}$$

$\mathcal{B}d^+(S)$ and $\mathcal{B}d^-(S)$ are antichains[4]. Since Q is monotonic, $Th(\mathcal{L}, \mathcal{D}, Q) = \downarrow \mathcal{B}d^+(Th(\mathcal{L}, \mathcal{D}, Q))$.

2.2 Dualization problems

The dualization problem has been studied in theoretical computer sciences for years. In our setting, dualization concerns the relationship between the positive border and the negative border. We recall the associated enumeration and decision problems.

Let us consider a pattern mining problem $(\mathcal{L}, \mathcal{D}, Q)$ and its associated theory $Th(\mathcal{L}, \mathcal{D}, Q)$.

[3] S is an ideal (resp. a filter) of the poset $(\mathcal{L}, \preceq)$

[4] $T \subseteq \mathcal{L}$ in an antichain if for all $X, Y \in T$, $X \npreceq Y$ and $Y \npreceq X$

Dualization (DualEnum)
Input: $\mathcal{D}$, $\mathcal{B}d^+(Th(\mathcal{L}, \mathcal{D}, Q))$
Question: Enumerate $\mathcal{B}d^-(Th(\mathcal{L}, \mathcal{D}, Q))$.

Dualization (DualDecision)
Input: $\mathcal{D}$, $\mathcal{B}d^+(Th(\mathcal{L}, \mathcal{D}, Q))$ and $\mathcal{F} \subseteq \mathcal{B}d^-(Th(\mathcal{L}, \mathcal{D}, Q))$
Question: Is $\mathcal{B}d^-(Th(\mathcal{L}, \mathcal{D}, Q)) = \mathcal{F}$? Otherwise find $\theta \in \mathcal{B}d^-(Th(\mathcal{L}, \mathcal{D}, Q)) \setminus \mathcal{F}$.

The complexity depends on the structural properties of the poset $(\mathcal{L}, \preceq)$. Some known results about dualization exist and are given below:

- $(\mathcal{L}, \preceq)$ is isomorphic to a boolean lattice: **DualDecision** is quasi-polynomial [13].
- $(\mathcal{L}, \preceq)$ is isomorphic to a product of chains: **DualDecision** is quasi-polynomial [11]
- $(\mathcal{L}, \preceq)$ is isomorphic to a lattice: **DualDecision** is $coNP$-complete [3].

Note that if the decision problem has a quasi-polynomial time complexity, then there exists an incremental quasi-polynomial time algorithm to enumerate each element in $\mathcal{B}d^-(Th(\mathcal{L}, \mathcal{D}, Q))$. Indeed, we start with an empty set $\mathcal{F}$ and at each step either we add an element to $\mathcal{F}$ and apply the decision problem or the algorithm stops.
In a data mining setting, we suppose that the predicate Q is a function from patterns and the database to $\{true, false\}$ which is computable in polynomial time. Moreover, we suppose that checking comparability in $\mathcal{L}$ can be done in polynomial time.

2.3 Set-oriented dualization

In this section, we describe the $\mathcal{RAS}$ class, i.e. pattern mining problems which are *representable as sets* [19].

Definition 1 Let $(\mathcal{L}, \mathcal{D}, Q)$ be a pattern mining problem. A finite set R and a total function f with $f : \mathcal{L} \rightarrow \mathcal{P}(\mathcal{R})$[5], denoted by the pair (R, f), is said to be a *representation as sets* of $(\mathcal{L}, \mathcal{D}, Q)$ if:

- f and f^{-1} are polynomially computable,
- f is bijective and
- for all $\theta, \varphi, \theta \preceq \varphi$ iff $f(\theta) \subseteq f(\varphi)$.

By extension, for any set $S \subseteq \mathcal{L}$, we denote by $f(S)$ the set $\cup_{\theta \in S}\{f(\theta)\}$. The class of pattern mining problems for which a representation as sets exists will be referred to as $\mathcal{RAS}$ in the sequel.

In this setting, we have a relationship between the positive and negative borders through the notion of minimal transversal of hypergraphs.

Let $\mathcal{H} \subseteq \mathcal{P}(R)$ be a hypergraph on a set R. We denote by $Max(\mathcal{H})$ (resp. $Min(\mathcal{H})$) the set of maximal (resp. minimal) hyperedges of $\mathcal{H}$ with respect to set inclusion. $\mathcal{H}$ is said to be *simple* if $\mathcal{H} = Min(\mathcal{H}) = Max(\mathcal{H})$. A minimal transversal of $\mathcal{H}$ is a set of elements $X \subseteq R$ such that (1) X has a non empty intersection with every hyperedge of $\mathcal{H}$ and (2) X is minimal w.r.t. this property. We denote by $\mathbf{TrMin}(\mathcal{H})$ the set of minimal transversals of $\mathcal{H}$ and $\overline{\mathcal{H}} = \{R \setminus E \mid E \in \mathcal{H}\}$ the complement of $\mathcal{H}$.

The relationship between the positive and negative border is given as follows [19]:

[5] $\mathcal{P}(R)$ denotes the powerset of R.

Theorem 1 Let $(\mathcal{L}, \mathcal{D}, Q) \in \mathcal{RAS}$, $S \subseteq \mathcal{L}$ and (R, f) a representation as sets of $(\mathcal{L}, \mathcal{D}, Q)$.
Then $\mathcal{B}d^-(\downarrow S) = f^{-1}(\mathbf{TrMin}(\overline{f(\mathcal{B}d^+(\downarrow S))}))$

Equivalently, by noticing that whenever $\mathcal{H}$ is a simple hypergraph, we have $\mathcal{H} = \mathbf{TrMin}(\overline{\mathbf{TrMin}(\mathcal{H})})$ [4], and therefore: $\mathcal{B}d^+(\downarrow S) = f^{-1}(\overline{\mathbf{TrMin}(f(\mathcal{B}d^-(\downarrow S)))})$.

It is worth noting that for any pattern mining problems belonging to the class $\mathcal{RAS}$, there exists incremental quasi-polynomial time algorithms [16] to enumerate the positive borders.

We give now a classical example of $\mathcal{RAS}$, i.e. frequent itemsets over a transactional database (see [1] for notations).

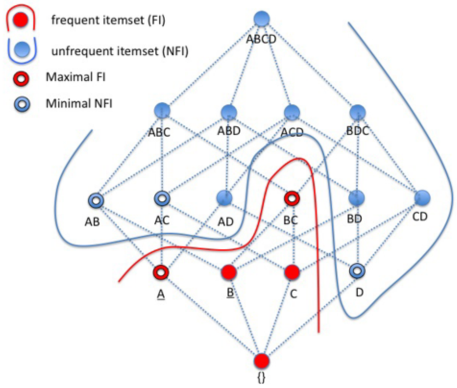

Figure 1. Example of set-oriented dualization

Example 1 Let $td_0 = \{A, ABC, BD, BC, BCD, A\}$ be a transactional database over $I = \{A, B, C, D\}$ and a minimum threshold value of 3. Let us denote the frequency constraint by $Freq_c$. We have: $\mathcal{B}d^+(Th(\mathcal{P}(I), td_0, Freq_c)) = \{A, BC\}$ and $\mathcal{B}d^-(Th(\mathcal{P}(I), td_0, Freq_c)) = \{D, AB, AC\}$.
It is easy to verify that the dualization of $\{A, BC\}$ gives $\{D, AB, AC\}$ (cf Figure 1). ◇

3 Weak representation as sets

Like [16], we note that the class $\mathcal{RAS}$ is quite restrictive. The encoding f has to be surjective, and in particular the number of patterns must be a power of two. This is indeed a very strong constraint, i.e. a large number of pattern mining problems do not have such a property, even for simple patterns as the following example shows.

Example 2 Let us consider a very simple case of sequence: suppose an alphabet with 2 letters (e.g. a and b) and an input sequence S of size 2. The set of all sub-sequences is made up of 7 elements (i.e. $\{\epsilon, a, b, aa, ab, ba, bb\}$). ◇

The previous example tells us that the number of possible subsequences of a given sequence is not a power of two in general, giving no hope to identify a representation as sets for sequences (as notified in [16] for episodes).

The following definition proposes weaker conditions, which we call *weak representation as sets* of pattern mining problems. As for $\mathcal{RAS}$, all patterns should have a set representation but the intuition is to dissociate the encoding and the decoding while preserving essential properties, mainly (1) the composition of the decoding of the encoding of a pattern is the identity, (2) the encoding preserves incomparability of patterns.

Definition 2 Let $(\mathcal{L}, \mathcal{D}, Q)$ be a pattern mining problem and $\perp$ a special pattern, $\perp \notin \mathcal{L}$. A finite set R and a pair of total functions (f, g) with $f : \mathcal{L} \rightarrow \mathcal{P}(R)$ and $g : \mathcal{P}(R) \rightarrow \mathcal{L} \cup \perp$, denoted by the triple (R, f, g), is said to be a *weak representation as sets* of $(\mathcal{L}, \mathcal{D}, Q)$ if

1. f and g are polynomially computable
2. for all $\theta \in \mathcal{L}$, $g(f(\theta)) = \theta$
3. for all $\theta, \varphi \in \mathcal{L}$, $f(\theta) \subseteq f(\varphi) \Rightarrow \theta \preceq \varphi$

The class of such problems will be referred to as $\mathcal{WRAS}$ in the sequel. Several observations can be made:

- From these conditions, f is injective and g is surjective.
- The condition 3 guarantees that the encoding preserves the incomparability of patterns, and then is "borders-preserving".

With respect to the class $\mathcal{RAS}$, our assumptions are clearly weaker:

- g is not required to be injective, but $g \circ f$ is the identity.
- f is no longer required to be neither surjective nor monotonic with respect to $\preceq$.

We now give some key technical notions underlying the dualization in a set-theoretical framework.

Definition 3 Let R be a finite set and $\mathcal{E}_1, \mathcal{E}_2$ two hypergraphs on R. $\mathcal{E}_1$ and $\mathcal{E}_2$ are said to be *dual* in $\mathcal{P}(R)$ if $\downarrow \mathcal{E}_1 \cup \uparrow \mathcal{E}_2 = \mathcal{P}(R)$ and $\downarrow \mathcal{E}_1 \cap \uparrow \mathcal{E}_2 = \emptyset$. Note that $Max(\mathcal{E}_1)$ and $Min(\mathcal{E}_2)$ are simple and dual.

Lemma 1 [4] Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be two simple dual hypergraphs on R. We have:

$$\mathcal{E}_1 = \overline{\mathbf{TrMin}(\mathcal{E}_2)}$$
$$\mathcal{E}_2 = \mathbf{TrMin}(\overline{\mathcal{E}_1})$$
$$\mathbf{TrMin}(\mathbf{TrMin}(\mathcal{E}_1)) = \mathcal{E}_1$$

Lemma 2 Let R be a finite set and $\mathcal{E}, \mathcal{E}_1, \mathcal{E}_2$ pairwise disjoints sets on $\mathcal{P}(R)$, such that $\mathcal{E}_1 = \downarrow \mathcal{E}_1$ and $\mathcal{E}_2 = \uparrow \mathcal{E}_2$. Then $\mathcal{P}(R) = \mathcal{E} \cup \mathcal{E}_1 \cup \mathcal{E}_2$ implies

(1) $\mathcal{E} \cup \downarrow \mathcal{E}_1 = \downarrow (\mathcal{E} \cup \mathcal{E}_1)$
(2) $\mathcal{E} \cup \uparrow \mathcal{E}_2 = \uparrow (\mathcal{E} \cup \mathcal{E}_2)$

Proof It suffices to note that $\mathcal{E} \cap \downarrow \mathcal{E}_1 = \emptyset$ since $\mathcal{E}_2$ is upward closed and disjoint with $\mathcal{E}$. The same result holds for the second equality. □

Clearly, $f(\mathcal{B}d^+(S))$ and $f(\mathcal{B}d^-(S))$ are not always dual since f is not surjective. So we need to identify those elements in $\mathcal{P}(R)$ that do not belong to $\downarrow f(\mathcal{B}d^+(S)) \cup \uparrow f(\mathcal{B}d^-(S))$.

Notation 1 In the rest of the paper, we shall denote by $\mathcal{E}$ the set of *extra elements* defined by $\mathcal{E} = \mathcal{P}(R) \setminus (\downarrow f(\mathcal{B}d^+(S)) \cup \uparrow f(\mathcal{B}d^-(S)))$. In Example 1, $\mathcal{E} = \emptyset$.

Proposition 1 $(\mathcal{L}, \mathcal{D}, Q) \in \mathcal{RAS}$ implies $\mathcal{E} = \emptyset$

Proof Suppose that $(\mathcal{L}, \mathcal{D}, Q) \in \mathcal{RAS}$. Then there exists an encoding f satisfying the Definition 1. Clearly for any $S \subseteq \mathcal{L}$, $\mathcal{B}d^+(\downarrow S)$ and $\mathcal{B}d^-(\downarrow S)$ are dual in $\mathcal{L}$. Moreover, the poset $\mathcal{L}$ is isomorphic to $\mathcal{P}(R)$ where R is the set of the encoding. So $f(\mathcal{B}d^+(\downarrow S))$ and $f(\mathcal{B}d^-(\downarrow S))$ are dual in $\mathcal{P}(R)$, and therefore $\mathcal{E} = \emptyset$. □

Whenever $\mathcal{E} \neq \emptyset$, the idea is then to "push" those extra elements either towards the negative border or towards the positive borders in order to get new dual sets.

Proposition 2 Let $(\mathcal{L}, \mathcal{D}, Q)$ be a pattern mining problem, $S \subseteq \mathcal{L}$ a downward closed set and (R, f, g) a weak representation as sets of $(\mathcal{L}, \mathcal{D}, Q)$. Then

(1) $f(\mathcal{B}d^+(S)) = \overline{\mathbf{TrMin}(Min(\mathcal{E} \cup f(\mathcal{B}d^-(S))))}$
(2) $f(\mathcal{B}d^-(S)) = \mathbf{TrMin}(\overline{Max(\mathcal{E} \cup f(\mathcal{B}d^+(S)))})$

Proof Clearly, by definition of $\mathcal{E}$, we have $\mathcal{P}(R) = \mathcal{E} \cup \uparrow f(\mathcal{B}d^-(S)) \cup \downarrow f(\mathcal{B}d^+(S))$. Moreover we have $\mathcal{E} \cap \uparrow f(\mathcal{B}d^-(S)) = \emptyset$, $\mathcal{E} \cap \downarrow f(\mathcal{B}d^+(S)) = \emptyset$. By the condition 3 of the Definition 2 we have $\downarrow f(\mathcal{B}d^+(S)) \cap \uparrow f(\mathcal{B}d^-(S)) = \emptyset$.

From Lemma 2 we deduce that $f(\mathcal{B}d^+(S))$ and $\mathcal{E} \cup f(\mathcal{B}d^-(S))$ are dual in $\mathcal{P}(R)$ and then the Lemma 1 applies. The same proof holds for the second equality. □

Remark 1 Note that in Proposition 2, we can omit Min in (1) since for any hypergraph $\mathcal{H}$, we have $\mathbf{TrMin}(\mathcal{H}) = \mathbf{TrMin}(Min(\mathcal{H}))$. For the same reason we can omit Max in (2) using the complementary.

Theorem 2 Let $(\mathcal{L}, \mathcal{D}, Q)$ be a pattern mining problem, $S \subseteq \mathcal{L}$ a downward closed set and (R, f, g) a weak representation as sets of $(\mathcal{L}, \mathcal{D}, Q)$. Then

(1) $\mathcal{B}d^+(S) = g(\overline{\mathbf{TrMin}(Min(\mathcal{E} \cup f(\mathcal{B}d^-(S))))})$
(2) $\mathcal{B}d^-(S) = g(\mathbf{TrMin}(\overline{Max(\mathcal{E} \cup f(\mathcal{B}d^+(S)))}))$

Proof We only show case (1). The other case is similar. From Proposition 2, we have $f(\mathcal{B}d^+(S)) = \overline{\mathbf{TrMin}(\mathcal{E} \cup f(\mathcal{B}d^-(S)))}$. Then $g(f(\mathcal{B}d^+(S))) = \mathcal{B}d^+(S)$ since for all $\theta \in \mathcal{L}$, $g(f(\theta)) = \theta$ (cf. condition 2 of Definition 2). □

For any $\mathcal{WRAS}$ problem, Theorem 2 allows to dualize and reuse well known existing techniques. Nevertheless, the overhead can be on the cost to compute the set $\mathcal{E}$.

4 Efficient weak representation as sets

Among weak representations as sets, the notion of *efficient* weak representation as sets characterizes those for which the dualization can be done in incremental quasi-polynomial time.

Definition 4 Let $\mathcal{E}^+, \mathcal{E}^- \subseteq \mathcal{E}$. $(\mathcal{E}^+, \mathcal{E}^-)$ is said to be a *separating pair* of $\mathcal{E}$ if $\mathcal{E}^+ \cap \mathcal{E}^- = \emptyset$, $\mathcal{E} \subseteq \downarrow \mathcal{E}^+ \cup \uparrow \mathcal{E}^-$, and $f(\mathcal{B}d^+(S)) \cup \mathcal{E}^+$ and $f(\mathcal{B}d^-(S)) \cup \mathcal{E}^-$ are antichains.

Corollary 1 Let $(\mathcal{L}, \mathcal{D}, Q)$ be a pattern mining problem, $S \subseteq \mathcal{L}$ a downward closed set, (R, f, g) a weak representation as sets of $(\mathcal{L}, \mathcal{D}, Q)$ and $(\mathcal{E}^+, \mathcal{E}^-)$ a separating pair of $\mathcal{E}$.
Then

(1) $\mathcal{B}d^+(S) = Max(g(\overline{\mathbf{TrMin}(\mathcal{E}^- \cup f(\mathcal{B}d^-(S)))}))$
(2) $\mathcal{B}d^-(S) = Min(g(\mathbf{TrMin}(\overline{\mathcal{E}^+ \cup f(\mathcal{B}d^+(S))})))$

Proof Using Definition 4, the sets $f(\mathcal{B}d^+(S)) \cup \mathcal{E}^+$ and $f(\mathcal{B}d^-(S)) \cup \mathcal{E}^-$ are dual.
Now consider the equality (1). Clearly $\overline{\mathbf{TrMin}(\mathcal{E}^- \cup f(\mathcal{B}d^-(S)))}$ contains $f(\mathcal{B}d^+(S))$ since $f(\mathcal{B}d^+(S)) \cup \mathcal{E}^+$ is an antichain. Moreover , the elements of $\mathcal{E}^+$ which have images by g are smallest to at least one element of $\mathcal{B}d^+(S)$. So $\mathcal{B}d^+(S) = Max(g(\overline{\mathbf{TrMin}(\mathcal{E}^- \cup f(\mathcal{B}d^-(S)))}))$. The same proof holds for the second equality. □

The difficulties to have a separating pair come from the following property: a separating pair may not exist whenever $f(\mathcal{L})$ is not a convex set, i.e. there exist $\theta, \varphi \in \mathcal{L}$ and $X \subseteq R$ such that $f(\varphi) \subset X \subset f(\theta)$ and $X \neq f(\phi)$ for any $\phi \in \mathcal{L}$. This may happen since f is not surjective.

Theorem 3 Let $(\mathcal{L}, \mathcal{D}, Q)$ be a pattern mining problem, $S \subseteq \mathcal{L}$ a downward closed set, (R, f, g) a weak representation as sets of $(\mathcal{L}, \mathcal{D}, Q)$ such that $f(\mathcal{L})$ is convex. Then there exists a separating pair of $\mathcal{E}$.

Proof $f(\mathcal{L})$ is convex, i.e. for any $\theta, \varphi \in \mathcal{L}$ and $X \subseteq R$ such that $f(\varphi) \subset X \subset f(\theta)$ we have $X = f(\phi)$ for some $\phi \in \mathcal{L}$. Consider the sets defined as follows:

$$\mathcal{E}^- = Min(\{X \in \mathcal{E} | \exists Y \in f(\mathcal{B}d^+(S)), Y \subset X\})$$

$$\mathcal{E}^+ = Max(\mathcal{E} \setminus \uparrow \mathcal{E}^-)$$

Clearly $\mathcal{E}^+ \cap \mathcal{E}^- = \emptyset$ and $\mathcal{E} \subseteq \downarrow \mathcal{E}^+ \cup \uparrow \mathcal{E}^-$. Moreover $f(\mathcal{B}d^+(S)) \cup \mathcal{E}^+$ and $f(\mathcal{B}d^-(S)) \cup \mathcal{E}^-$ are antichains, since $\mathcal{E} \cap \uparrow f(\mathcal{B}d^-(S)) = \emptyset$ and $\mathcal{E} \cap \downarrow f(\mathcal{B}d^+(S)) = \emptyset$. Then $(\mathcal{E}^-, \mathcal{E}^+)$ is a separating pair of $\mathcal{E}$. □

Definition 5 Let $(\mathcal{E}^+, \mathcal{E}^-)$ be a separating pair of $\mathcal{E}$. $(\mathcal{E}^+, \mathcal{E}^-)$ is said to be a *efficient* if $|\mathcal{E}^+|$ and $|\mathcal{E}^-|$ are bounded by a polynom in the size of the borders of *Th*$(\mathcal{L}, \mathcal{D}, Q)$.

The class of efficient $\mathcal{WRAS}$ problems, denoted by $\mathcal{EWRAS}$, consists of $\mathcal{WRAS}$ problems for which there exists at least one *efficient* separating pair.

Theorem 4 The dualization problem of any $\mathcal{EWRAS}$ problem can be polynomially reduced to hypergraph transversal problem.

Proof Suppose $(\mathcal{L}, \mathcal{D}, Q) \in \mathcal{EWRAS}$, (R, f, g) its weak representation as sets and $S \subseteq \mathcal{L}$ a downward closed set. Let $\mathcal{F} = f(\mathcal{B}d^+(S)) \cup \mathcal{E}^+$ where $(\mathcal{E}^+, \mathcal{E}^-)$ is a separating pair. Note that $\mathcal{F}$ can be computed in polynomial time in the size of S, since $(\mathcal{L}, \mathcal{D}, Q) \in \mathcal{EWRAS}$. Moreover $f(\mathcal{B}d^+(S)) \cup \mathcal{E}^+$ and $f(\mathcal{B}d^-(S)) \cup \mathcal{E}^-$ are dual in $\mathcal{P}(R)$. According to Corollary 1, $\mathcal{B}d^-(S) = Min(g(\mathbf{TrMin}(\overline{\mathcal{E}^+ \cup f(\mathcal{B}d^+(S))})))$ which is computable in polynomial time, since the decoding g and comparability checking are in polynomial time. □

5 Applications to rigid sequences with wildcards

We now study a simple but representative pattern mining problems not representable as sets, namely frequent rigid sequences with wildcards. The study of other pattern mining problems such as those involving trees or episodes is out of the scope of this paper.

5.1 Problem statement

Let Σ be an alphabet and $\star \notin \Sigma$ be a wildcard. Let us fix a finite string, called an *input sequence*, $S = S[1] \cdots S[n] \in \Sigma^*$ of length $n \geq 0$. A rigid motif (or motif) is a string $P = P[1] \cdots P[m] \in (\Sigma \cup \{\star\})^*$ of length $m \leq n$ such that $P[1] \neq \star$ and $P[m] \neq \star$. For motifs $P[1..m]$ and $Q[1..n]$, we say that P *occurs* in Q at position $p \in [1..n]$, denoted by $P \sqsubseteq_p Q$, if for every $i \in [1..m]$ we have either $P[i] = Q[p + i - 1]$ or $P[i] = \star$. The location list for P in S is the set $L_S(P) = \{p \in [1..n] \mid P \sqsubseteq_p S\}$. The *frequence* of P in S is defined by: $freq(P, S) = |L_S(P)|$. A motif P is said to be k-frequent in S if $freq(P, S) \geq k$. The frequency predicate for sequences is denoted by $Freq_s$.

We will say that P occurs in Q, denoted by $P \sqsubseteq Q$, if there exists at least one position p such that $P \sqsubseteq_p Q$ (i.e. $L_Q(P) \neq \emptyset$).

Let $\mathcal{L}_S$ be the set of all rigid motifs over $\Sigma \cup \star$. $(\mathcal{L}_S, \sqsubseteq)$ is a partial order.

In the sequel, $Th(\mathcal{L}_S, S, Freq_s)$, the theory of the frequent rigid sequences with wildcard problem, will be referred to as $Th(S)$ to simplify the notations.

The dualization of frequent rigid sequences with wildcard over $(\mathcal{L}_S, S, Freq_s)$ can now be stated as follows.

Problem statement SEQ (DualEnum of sequences)
Input: $S, \mathcal{B}d^+(Th(S))$
Question: Enumerate $\mathcal{B}d^-(Th(S))$.

Example 3 Let us consider a simple sequence $S = abab$ defined over $\Sigma = \{a, b\}$ and a minimum threshold value of 2. We get:
$\mathcal{B}d^+(\{ab, a, b\}) = \{ab\}$,
$\mathcal{B}d^-(\{ab, a, b\}) = \{aa, bb, ba, a\star a, a\star b, b\star a, b\star b\}$ ◇

SEQ does not belong to the class $\mathcal{RAS}$ as shown in Example 2.

5.2 Weak representation as sets for SEQ

Let $S = S[1]\cdots S[n]$ be a sequence. First, we define a classical encoding of sequences into sets: Let $R = \{(i, x) | i \in [1..n], x \in \Sigma\}$.

Second, we define a set encoding function f from $\mathcal{L}_S$ to $\mathcal{P}(R)$ as follows: Let $P[1..m] \in \mathcal{L}_S$. We define f as:

$$f(P) = \{(i, P[i]) \mid i \in [1..m], P[i] \neq \star\}$$

Example 4 Continuing Example 3, we have $R = \{(1, a), (2, a), (3, a), (4, a), (1, b), (2, b), (3, b), (4, b)\}$ and $f(abab) = \{(1, a), (2, b), (3, a), (4, b)\}$, $f(ab\star\star b) = \{(1, a), (2, b), (5, b)\}$. ◇

Moreover, as the following example shows, f is neither surjective nor monotonic.

Example 5 Let $X = \{(1, a), (2, b)\}$ and $X' = \{(2, a), (3, b)\}$. The sequence ab corresponds to X by f while X' is not an image by f; hence f is not surjective.
Let bb and abb be two patterns. We have $f(bb) \not\subseteq f(abb)$ whereas $bb \sqsubseteq abb$; hence f is not monotonic. ◇

Let us now characterize elements of $\mathcal{P}(R)$ which are not images of f. Given a pattern $P \in \mathcal{L}_S$, two remarks can be done:

- the image $f(P)$ must contain a unique symbol in each index; and
- $f(P)$ also contains $(1, x)$ for some symbol $x \in \Sigma$.

As a consequence, two sets $\mathcal{F}^+$ and $\mathcal{F}^-$ can be identified as follows:

$\mathcal{F}^+ = \{\{(i, x) \mid x \in \Sigma, i \in [2..n]\}\}$

$\mathcal{F}^- = \{\{(1, x), (1, y) \mid x, y \in \Sigma, x \neq y\}\} \cup \{\{(1, x), (i, y), (i, z)\} \mid x, y, z \in \Sigma, y \neq z, i \in [2..n]\}$

The sets $\downarrow \mathcal{F}^+$ and $\uparrow \mathcal{F}^-$ characterize useless elements of $\mathcal{P}(R)$, i.e. those elements which do not have an image by the encoding f.

Example 6 Continuing Example 3, we have:
$\mathcal{F}^+ = \{\{(2, a), (3, a), (4, a), (2, b), (3, b), (4, b)\}\}$
$\mathcal{F}^- = \{\{(1, a), (1, b)\}, \{(1, a), (2, a), (2, b)\}, \{(1, b), (2, a), (2, b)\}, \{(1, a), (3, a), (3, b)\}, \{(1, b), (3, a), (3, b)\}, \{(1, a), (4, a), (4, b)\}, \{(1, b), (4, a), (4, b)\}\}$ ◇

Consequently, the decoding function g is defined as follows:

$$g(X) = \begin{cases} \theta & \text{if} \quad X = f(\theta) \\ \perp & otherwise \quad (i.e. \quad X \in \uparrow \mathcal{F}^- \cup \downarrow \mathcal{F}^+) \end{cases}$$

Lemma 3 Let $A \in \mathcal{P}(R)$ such that $A \notin\ \uparrow \mathcal{F}^-$ and $A \notin\ \downarrow \mathcal{F}^+$. Then $f(g(A)) = A$.

Proof Clearly, if $(i, x) \in A$ and $(j, y) \in A$ implies $i \neq j$. Moreover there is at most one pair $(1, x) \in A$ with $x \in \Sigma$. We conclude that there is $\theta \in \mathcal{L}_S$ such that $g(A) = \theta$ and therefore $f(g(A)) = A$. □

In other words, there is a bijection between $\mathcal{L}_S$ and $\mathcal{P}(R) \setminus (\uparrow \mathcal{F}^- \cup \downarrow \mathcal{F}^+)$.

Theorem 5 **SEQ** belongs to $\mathcal{WRAS}$.

Proof Let us show that (R, f, g) satisfies the conditions of the Definition 2. Let $\theta \in \mathcal{L}_S$. We have $(1, \theta[1]) \in f(\theta)$ and for all $i \in [2..n]$, $\{(1, z), (i, x), (i, y)\} \not\subseteq f(\theta)$ and $\{(1, x), (1, y)\} \not\subseteq f(\theta)$. Then from Lemma 3, we conclude that $f(g(f(\theta))) = f(\theta)$. Or f is injective, then $g(f(\theta)) = \theta$
Let $\theta, \varphi \in \mathcal{L}_S$ such that $f(\varphi) \subseteq f(\theta)$. Then we have $g(f(\theta)) = \theta$ and $g(f(\varphi)) = \varphi$ and $\theta[1] = \varphi[1]$ since there is $(1, x) \in f(\varphi)$ and it must be in $f(\theta)$. This concludes that $\varphi \sqsubseteq \theta$ since $\star \sqsubseteq x$ for $x \in \Sigma$. □

5.3 Efficient weak representation as sets for SEQ

In this section we show that the set $\mathcal{E}$ can be partitioned into sets $\mathcal{E}^+$ and $\mathcal{E}^-$ and have polynomial size in the size of borders of $(\mathcal{L}_S, S, Freq_s)$. First we give some properties for the encoding f and the decoding g.

Proposition 3 $f(\mathcal{L}_S)$ is convex.

Proof We need to show that $\varphi, \theta \in \mathcal{L}_S$, $X \subseteq R$ such that $f(\varphi) \subseteq X \subseteq f(\theta)$. Then there exists $\alpha \in \mathcal{L}_S$ with $f(\alpha) = X$ and $\varphi \sqsubseteq \alpha \sqsubseteq \theta$. Suppose $f(\varphi) \subseteq X \subseteq f(\theta)$. Then exists $(1, x), x \in \Sigma$ such that $(1, x) \in X$ since $f(\varphi) \subseteq X$. On the other side, for any $i \in [2..n], x, y, z \in \Sigma$ we have $\{(1, x), (1, y)\} \not\subseteq X$ and $\{(1, x), (i, y), (i, z)\} \not\subseteq X$, since $X \subseteq f(\theta)$. Thus there exists α such that $f(\alpha) = X$. □

From Theorem 3, we deduce that the set $\mathcal{E} = \mathcal{P}(R) \setminus (\uparrow f(\mathcal{B}d^-(\mathcal{L}_S, , S, Freq_s)) \cup \downarrow f(\mathcal{B}d^+(Th(\mathcal{L}, \mathcal{D}, Q)\mathcal{S})))$ has a separating pair. Now the question is how to obtain an efficient partition. The difficulties come from the facts that (1) the set $\mathcal{E}$ may contain images of the encoding f, and (2) the computation of the sets $\mathcal{E}^+$ and $\mathcal{E}^-$.

Therefore, we characterize elements of $\mathcal{E}$ that are images of the encoding f; such elements are images of patterns that are either greater than an element of the negative border or smaller than an element of the positive border.
Second, we define the successor and predecessor as follows:
Let $\theta \in \mathcal{E}$. The set $Succ(\theta) = \{x\star^i\theta \mid i \in [0..n - |\theta| - 1], x \in \Sigma\}$.
Dually we define $Pred(\theta) = \{\theta[i..|\theta|] \mid i \in [2..|\theta|]\}$.

Example 7 Let $\Sigma = \{a, b\}, \theta = bb, \theta' = babba$ and $n = 5$. Then $Succ(\theta) = \{abb, bbb, a\star bb, b\star bb, a\star\star bb, b\star\star bb\}$ and $Pred(\theta') = \{abba, bba, ba, a\}$. Clearly for all $\varphi \in Succ(\theta)$, $\theta \sqsubseteq \varphi$ and for all $\varphi \in Pred(\theta')$, $\varphi \sqsubseteq \theta'$ ◇

Lemma 4 Let $\varphi, \theta \in \mathcal{L}_S$ such that $\varphi \sqsubseteq \theta$. Then
(1) there exist $\alpha \in \{\varphi\} \cup Succ(\varphi)$ such that $f(\alpha) \subseteq f(\theta)$, and
(2) there exist $\phi \in \{\theta\} \cup Pred(\theta)$ such that $f(\varphi) \subseteq f(\phi)$,

Proof Let $\theta, \varphi \in \mathcal{L}_S$ and $\varphi \sqsubseteq \theta$. Then there exists an index $i \in [1..|\theta|]$ such that for all $j \in [1..|\varphi|]$, $\varphi[j] = \theta[i + j - 1]$ or $\varphi[j] = \star$.
(1) Clearly if $i = 1$ then $\alpha = \varphi$. Now suppose $i > 1$. Let $\alpha[1] = \theta[1]$, $\alpha[j] = \star, j \in [2..i - 1]$ and $\alpha[i + k - 1] = \varphi[k], k \in [1..|\varphi|]$. Then $\alpha \in Succ(\varphi)$ and $f(\alpha) \subseteq f(\theta)$ by construction.

(2) Let $\phi = \theta[i..|\theta|]$. Then $\phi \in \{\theta\} \cup Pred(\theta)$, and $f(\varphi) \subseteq f(\phi)$. □

We show how to extend the borders $f(\mathcal{B}d^+(Th(S)))$ and $f(\mathcal{B}d^-(Th(S)))$ to $\mathcal{E}^+ \cup f(\mathcal{B}d^+(Th(S)))$ and $\mathcal{E}^- \cup f(\mathcal{B}d^-(Th(S)))$ such that all images of f are either in $\downarrow (\mathcal{E}^+ \cup f(\mathcal{B}d^+(Th(S)))$ or $\uparrow (\mathcal{E}^- \cup f(\mathcal{B}d^-(Th(S)))$. Consider the sets $\mathcal{F}'^+$ and $\mathcal{F}'^-$ defined as follows:

$$\mathcal{F}'^+ = \bigcup_{\theta \in \mathcal{B}d^+(Th(S))} f(Pred(\theta))$$

$$\mathcal{F}'^- = \bigcup_{\theta \in \mathcal{B}d^-(Th(S))} f(Succ(\theta))$$

Now we are able to give the sets $\mathcal{E}^+$ and $\mathcal{E}^-$:

$$\mathcal{E}^+ = Max(\mathcal{F}^+ \cup \mathcal{F}'^+ \cup f(\mathcal{B}d^+(Th(S)))) \setminus f(\mathcal{B}d^+(Th(S)))$$

$$\mathcal{E}^- = Min(\mathcal{F}^- \cup \mathcal{F}'^- \cup f(\mathcal{B}d^-(Th(S)))) \setminus f(\mathcal{B}d^-(Th(S)))$$

Example 8 Consider the sequence $S = aba$ and a frequency threshold of 2. Then we get $\mathcal{B}d^+(Th(S)) = \{a\}$ and $\mathcal{B}d^-(Th(S)) = \{b, aa, a \star a\}$. $\mathcal{L}_S$ is encoded in $\mathcal{P}(R)$ with $R = \{(1,a),(1,b),(2,a),(2,b),(3,a),(3,b)\}$. We have:
$f(\mathcal{B}d^+(Th(S))) = \{(1,a)\}$,
$f(\mathcal{B}d^-(Th(S))) = \{\{(1,b)\},\{(1,a),(2,a)\},\{(1,a),(3,a)\}\}$,
$\mathcal{F}^+ = \{\{(2,a),(2,b),(3,a),(3,b)\}\}$,
$\mathcal{F}^- = \{\{(1,a),(1,b)\},\{(1,a),(2,a),(2,b)\},\{(1,b),(2,a),(2,b)\},\{(1,a),(3,a),(3,b)\},\{(1,b),(3,a),(3,b)\}\}$,
$\mathcal{F}'^+ = Pred(a) = \emptyset$ and
$\mathcal{F}'^- = f(Succ(b)) \cup f(Succ(aa)) \cup f(Succ(a \star a)) = f(\{ab, bb, a \star b, b \star b\}) \cup f(\{aaa, baa\}) \cup f(\emptyset) = \{\{(1,a),(2,b)\},\{(1,b),(2,b)\},\{(1,a),(3,b)\},\{(1,b),(3,b)\},\{(1,a),(2,a),(3,a)\},\{(1,b),(2,a),(3,a)\}\}$.

Finally, we obtain:
$\mathcal{E}^+ = \{\{(2,a),(2,b),(3,a),(3,b)\}\}$ and
$\mathcal{E}^- = \{\{(1,a),(1,b)\},\{(1,a),(2,a),(2,b)\},\{(1,b),(2,a),(2,b)\},\{(1,a),(3,a),(3,b)\},\{(1,b),(3,a),(3,b)\},\{(1,a),(2,b)\},\{(1,a),(3,b)\}\}\}$ ◇

Proposition 4 $\mathcal{E}^+ \cup f(\mathcal{B}d^+(Th(S)))$ and $\mathcal{E}^- \cup f(\mathcal{B}d^-(Th(S))))$ form a separating pair of $\mathcal{E}$.

Proof First note that $\downarrow \mathcal{F}^+ \cap \uparrow \mathcal{F}'^- = \emptyset$. Moreover, by construction of the sets $\mathcal{F}'^+$ and $\mathcal{F}'^-$ we have $f(\mathcal{B}d^+(Th(S))) \cap \mathcal{E}^+ = \emptyset$ and $f(\mathcal{B}d^-(Th(S))) \cap \mathcal{E}^- = \emptyset$. From Lemma 3 and 4, we conclude that they are dual. □

Theorem 6 **SEQ** belongs to $\mathcal{EWRAS}$.

Proof From Proposition 4, we conclude that Corrollary 1 can be applied. Moreover the size of $\mathcal{E}^+$ is bounded by $n \times |\Sigma| \times |\mathcal{B}d^+(\mathcal{L}_S, S, Q))|$ and the size of $\mathcal{E}^-$ by $n \times (|\Sigma|^3 + |\mathcal{B}d^-(\mathcal{L}_S, S, Q)|)$ □

6 Conclusion and Discussion

The main contribution of this paper is to extend the class of problems representable as sets of [19], i.e. the class of problems for which dualization based on minimal transversal applies, to a wider class of problems. We have proposed the notion of *weak representation as sets* of pattern mining problems from which we have defined an *efficient* subclass: Its main merit is to ensure the existence of an incremental quasi-polynomial algorithm for the dualization problem and hence, for the pattern mining mining problem itself. We have considered in this paper only one reduction to $\mathcal{EWRAS}$ for rigid sequences with wildcards. Many other pattern mining problems have to be revisited in this setting, such as those involving episodes, graphs or queries in database [17].

The new classes of pattern mining problems introduced in this paper should be much more wider and might prove to have large impact in unifying existing pattern mining approaches.

Acknowledgment We would like to thank the anonymous reviewers for their helpful comments and remarks. This work has been done in the DAG project (`http://liris.cnrs.fr/dag`), funded by the French ANR agency under the DEFIS 2009 program.

REFERENCES

[1] R. Agrawal and R. Srikant, 'Fast algorithms for mining association rules in large databases', in *Proceedings of the* 20^{th} *International Conference on Very Large Databases, Santiago de Chile, Chile*, pp. 487–499, (1994).

[2] Hiroki Arimura and Takeaki Uno, 'Polynomial-delay and polynomial-space algorithms for mining closed sequences, graphs, and pictures in accessible set systems', in *SDM*, pp. 1087–1098, (2009).

[3] Mikhail A. Babin and Sergei O. Kuznetsov, 'Enumerating minimal hypotheses and dualizing monotone boolean functions on lattices', in *ICFCA*, pp. 42–48, (2011).

[4] Claude Berge, *Graphs and Hypergraphs*, North-Holland Mathematical Library 6, American Elsevier, 2d rev. ed. edn., 1976.

[5] Mario Boley, Tamás Horváth, Axel Poigné, and Stefan Wrobel, 'Listing closed sets of strongly accessible set systems with applications to data mining', *Theor. Comput. Sci.*, **411**(3), 691–700, (2010).

[6] Toon Calders and Jef Wijsen, 'On monotone data mining languages', in *DBPL*, pp. 119–132, (2001).

[7] Vineet Chaoji, Mohammad Al Hasan, Saeed Salem, and Mohammed Javeed Zaki, 'An integrated, generic approach to pattern mining: data mining template library', *Data Min. Knowl. Discov.*, **17**(3), 457–495, (2008).

[8] Thomas Eiter and Georg Gottlob, 'Identifying the minimal transversals of a hypergraph and related problems.', *SIAM J. Comput.*, **24**(6), 1278–1304, (1995).

[9] Thomas Eiter, Georg Gottlob, and Kazuhisa Makino, 'New results on monotone dualization and generating hypergraph transversals', *SIAM J. Comput.*, **32**, 514–537, (February 2003).

[10] Thomas Eiter, Kazuhisa Makino, and Georg Gottlob, 'Computational aspects of monotone dualization: A brief survey', *Discrete Appl. Math.*, **156**, 2035–2049, (June 2008).

[11] Khaled M. Elbassioni, 'Algorithms for dualization over products of partially ordered sets', *SIAM J. Discrete Math.*, **23**(1), 487–510, (2009).

[12] Frédéric Flouvat, Fabien De Marchi, and Jean-Marc Petit, *New Frontiers in Applied Data Mining*, chapter The iZi Project: Easy Prototyping of Interesting Pattern Mining Algorithms, 1–15, LNAI 5669, Springer-Verlag, February 2010.

[13] Michael L. Fredman and Leonid Khachiyan, 'On the complexity of dualization of monotone disjunctive normal forms', *J. Algorithms*, **21**(3), 618–628, (1996).

[14] Alain Gély, Lhouari Nourine, and Bachir Sadi, 'Enumeration aspects of maximal cliques and bicliques', *Discrete Applied Mathematics*, **157**(7), 1447–1459, (2009).

[15] Arnaud Giacometti, Patrick Marcel, and Arnaud Soulet, 'A relational view of pattern discovery', in *DASFAA*, pp. 153–167, (2011).

[16] Dimitrios Gunopulos, Roni Khardon, Heikki Mannila, Sanjeev Saluja, Hannu Toivonen, and Ram Sewak Sharm, 'Discovering all most specific sentences', *ACM Trans. Database Syst.*, **28**(2), 140–174, (2003).

[17] Tao-Yuan Jen, Dominique Laurent, and Nicolas Spyratos, 'Computing supports of conjunctive queries on relational tables with functional dependencies', *Fundam. Inform.*, **99**(3), 263–292, (2010).

[18] Mamadou Moustapha Kanté, Vincent Limouzy, Arnaud Mary, and Lhouari Nourine, 'Enumeration of minimal dominating sets and variants', FCT'11, pp. 298–309. Springer-Verlag, (2011).

[19] H. Mannila and Hannu Toivonen, 'Levelwise search and borders of theories in knowledge discovery', *Data Mining and Knowledge Discovery*, **1**(3), 241–258, (1997).