



**HAL**  
open science

## Abstract closed patterns beyond lattices

Henry Soldano

► **To cite this version:**

Henry Soldano. Abstract closed patterns beyond lattices. *Reconnaissance de Formes et Intelligence Artificielle (RFIA) 2014*, Jun 2014, France. hal-00989204

**HAL Id: hal-00989204**

**<https://hal.science/hal-00989204>**

Submitted on 9 May 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Abstract closed patterns beyond lattices

Henry Soldano

Universite Paris 13, Sorbonne Paris Cite, L.I.P.N UMR-CNRS 7030  
F-93430, Villetaneuse, France

## Résumé

*La recherche en fouille de motifs a porté ces dernières années en particulier sur les opérateurs de fermeture sur des langages partiellement ordonnés, isomorphes à un sous-ensemble d'un ensemble d'attributs, ne formant pas nécessairement un treillis. Un résultat de M. Boley et co-auteurs définit une propriété qui garantit qu'un opérateur de fermeture existe quel que soit l'ensemble d'objets dans lequel on cherche le support des motifs. Nous relient ce travail au cadre classique de l'analyse de Galois et des concepts formels, détaillons la structure des ensembles de fermés, ainsi que les implications associées, et montrons que la simplification par abstraction extensionnelle reste applicable dans ce cas.*

## Mots Clef

Analyse Formelle de concepts, Treillis de Galois, abstraction, motifs fermés

## Abstract

*Recently pattern mining has investigated closure operators in families of subsets of an attribute set that are not lattices. A result due to M. Boley and coauthors defines a property of such a family, denoted as confluence, that guarantees that a support closure operator exists whatever is the set of objects on which supports are computed. We investigate this pattern mining framework and relate it to FCA. We give results on closure operators outside lattices and discuss the structure of closed pattern sets together with the related set of implications, and show that simplifying the closed set using extensional abstractions hold for this new structures.*

## Keywords

Formal Concept Analysis, Galois lattice, abstraction, closed patterns

## 1 Introduction

Until recently searching for *closed* motifs or patterns when exploring data was restricted to lattices as pattern languages. A pattern in some language  $\mathcal{L}$  is said closed whenever it can be obtained by applying a *closure operator* to some pattern. This subject has been thoroughly explored in Formal Concept Analysis, Galois analysis and Data Mining when considering *support-closed* patterns where we have a set of objects  $O$  and a motif has a support, i.e. it occurs, in a

set of object. The language is a lattice, and the motif occurs in an object whenever the motif is more general than the object description. Motifs that can't be specialized without losing some object in their support are said *support-closed*. Considering only equivalence classes of motifs sharing the same support helps investigating the data. Support-closed motifs are searched for as representatives of such equivalence classes. This can be performed efficiently because there exists a closure operator on the lattice that returns as a closed pattern the unique support-closed pattern of the corresponding equivalence class.

The most investigated pattern language is the power set  $2^X$  of some attribute set  $X$ , ordered following the set-theoretic inclusion order. Formal Concept Analysis [5] as well as Galois analysis [4] relies on the relation between objects and attributes. In data mining, these ideas have been investigated under the name of *itemsets mining* and also rely on the same relation[8].

Recently, pattern mining has gone beyond this general framework in two directions. First, various mining problems have been investigated that comes down to searching for closed motifs which can't be considered, strictly speaking, as support-closed motifs, as for instance, convex hull of subsets of a given set of points, or sequential motifs with wild-cards [1]. To characterize such closure operators, the authors make use of the well-known one-to-one correspondence between families closed under the *meet* operator and the closure operators. Second, various mining problems have been addressed in which the language is a partial order but not a lattice, but still there is a support-closure operators. A general framework has been proposed for that purpose, in which the language is a family  $F$  included in a host lattice  $2^X$ . For instance, consider the set of the subgraphs generated by a subset of the set  $X$  of the edges of a given graph  $(V, X)$ . Such a subgraph can be represented as a subset of  $X$ , however the family  $F$  of connected subgraphs is not a lattice<sup>1</sup>. Still there is a closure operator that relates a connected subgraph to a support-closed connected subgraph. In their paper, [3], M. Boley and coauthors state in particular the necessary and sufficient conditions that have to fulfill a set system the family  $F$  of a set system  $(F, X)$ , in order to guarantee, that whatever, with some mild restriction, is the dataset  $O$  of objects we consider. The corresponding property of *confluence* mainly consists in requiring a

1. the intersection of two such connected subgraphs is not necessarily connected

kind of *local union closure*. and we will further denote as *confluence\** a slightly stronger property that eliminate the mild restriction above.

Our contribution concerns the two directions. First, we state sufficient conditions to obtain closed patterns for structures weaker than lattices. These conditions, are denoted here as the *pre-confluence* property. The main condition we require is that given three elements  $t, t_1, t_2$  of  $F$ , if  $t_1$  and  $t_2$  belongs to the up set  $\uparrow t$  then there exists a greatest lower bound of  $t_1$  and  $t_2$  in the upset  $\uparrow t$  of  $F$ , and this *local meet* element is denoted as  $t_1 \wedge_t t_2$ . In Figure 1, the pre-confluent family  $F$  where  $a, b, c, d$  are the edges of a graph is represented on the left.

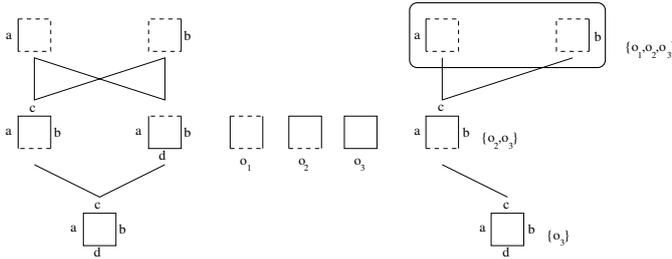


FIGURE 1 – On the left, the Hass diagram of a family of connected subgraphs each generated by a subset (represented by a word) of the edges  $\{a, b, c, d\}$  of the original graph. The subgraphs generated by  $a$  and by  $b$  are the minimal elements.  $F$  is a pre-confluent family in which, for instance,  $\{abc, abd\}$  have two maximal lower bounds,  $a = abc \wedge_a abd$  greater than  $a$ , and the other,  $b = abc \wedge_b abd$  greater than  $b$ . Furthermore, as for all pair of elements greater than some other element in  $F$  the union of these elements also belongs to  $F$ ,  $F$  also is a confluent\* family.  $F$  is such that for any element  $x$ , the up set  $\uparrow x$  is a lattice. In particular  $\uparrow a = \{a, abc, abd, abcd\}$  is a lattice. On the right, the Hass diagram of the support closed connected subgraphs pre-confluence  $f[F]$  with respect to the set of subgraphs  $O = \{o_1, o_2, o_3\}$  (on the middle part of the figure). The thick box around closed patterns  $a$  and  $b$  indicates that both patterns have the same extension  $\{o_1, o_2, o_3\}$ .

Second, we show that, when the pre-confluence  $F$  is a subset of some lattice  $T$  and follows some simple conditions leading to the notion of *confluence\**, there exists a closure operator returning the support-closed elements of  $F$  with respect to any object set  $O$  whose objects are represented as elements of the lattice  $T$ . In left part of Figure 1 we also note that for any element  $x$  of the pre-confluence  $F$ , the up set  $\uparrow x$  is a lattice. This is a general and straightforward result, that allow to link closure operators in pre-confluences to closure operators on lattices, and therefore to FCA, and help defining implication basis on confluences\*.

Finally, a last contribution consists in noticing that, when  $F$  is a *confluence\**, applying projections to the extensional space  $2^O$  preserves closure operators existence and there-

fore, abstract Galois lattices, as alpha Galois lattices [10] are extended to Galois pre-confluences.

## 2 Closure subsets of a partial order

We are interested here with closed elements of an ordered set. When this ordered set refers to a language for pattern mining, we call patterns the elements of the ordered set.

### 2.1 Preliminaries

We first recall definitions of closure and dual closure operators :

**Definition 1** Let  $E$  be an ordered set and  $f : E \rightarrow E$  be an automorphism such that for any  $x, y \in E$ ,  $f$  is monotone, i.e.  $x \leq y \implies f(x) \leq f(y)$  and idempotent, i.e.  $f(f(x)) = f(x)$ , then :

- if  $f$  is extensive, i.e.  $f(x) \geq x$ ,  $f$  is called a closure operator
- if  $f$  is intensive, i.e.  $f(x) \leq x$ ,  $f$  is called a dual closure operator or an interior operator, or also a projection.

In the first case, an element such that  $x = f(x)$  is called a closed element.

We define hereunder a closure subset of an ordered set  $E$  as the range  $f[E]$  of a closure operator on  $E$ . We give then a necessary and sufficient condition for a subset of  $E$  to be a closure subset. This condition answers the general question of which subsets of some pattern language are sets of closed patterns. The set of upper bounds of some element  $x$  in  $E$  is denoted as the up set  $\uparrow x = \{y \mid y \geq x\}$  also denoted as  $E^x$  when more than one partial order is concerned. In the same way, the set of lower bounds of  $x$  is denoted as the down set  $\downarrow x = \{y \mid y \leq x\}$  also denoted as  $E_x$ .

**Definition 2 (T.S. Blyth [2])** A subset  $C$  of an ordered set  $E$  is called a closure subset if there is a closure  $f : E \rightarrow E$  such that  $C = f[E]$ .

**Proposition 1 (T.S. Blyth [2])** A subset  $C$  of an ordered set  $E$  is a closure subset of  $E$  if and only if for every  $x \in E$  the set  $\uparrow x \cap C$  has a bottom element  $x_*$ . The closure  $f : E \rightarrow E$  is then unique and defined as  $f(x) = x_*$ .

However this property does not give a direct information in which pattern languages closed patterns are to be found and in which conditions closure operators exist. A direct information is provided by a well known result on closure subsets of complete  $\wedge$ -semilattices [5]. This result states that in such a pattern language, the closure subsets are the subsets closed by the meet operator  $\wedge$ . When the language is the power set of some set  $X$ , the meet operator simply is the intersection operator  $\cap$ .

**Proposition 2** Let  $T$  be a lattice. A subset  $C$  of  $T$  is a closure subset if and only if  $C$  is closed under meet. The closure  $f : T \rightarrow T$  is then unique and defined as  $f(x) = \wedge_{\{c \in C \cap \uparrow x\}} c$  and  $C$  is a lattice.

All ordered sets considered here are finite, and as all lattices are finite lattices they are also complete lattices : any subset of a lattice  $T$  is then closed under arbitrary meet and arbitrary join. Note that when saying that  $C$  is closed under meets we intend here that the meet of  $\emptyset$  also belongs to  $C$ . Therefore  $\top = \bigwedge_{\emptyset} c$  belongs to  $C$ .

We will also further need the dual proposition which states that a subset  $A$  of  $T$  is a *dual closure subset*, also denoted as an *abstraction*, whenever  $A$  is closed under joins. The interior operator  $p : T \rightarrow T$  is then defined as  $p(x) = \bigvee_{\{a \in A \mid a \leq x\}} a$ ,  $A$  is a lattice and  $\perp$  belongs to  $A$ . In particular when  $T$  is a powerset  $2^K$ ,  $p(x) = \bigcup_{\{a \in A \mid a \subseteq x\}} a$ . We are interested now in pre-confluences which are structures weaker than lattices.

## 2.2 Closure subsets in pre-confluences

**Definition 3** Let  $F$  be an ordered set such that for any  $t \in F$ ,  $\uparrow t$  is a  $\wedge$ -semilattice and has a top element.  $F$  is called a *pre-confluence*,  $x \wedge_t y$  is a *local infimum* or *local meet*, and  $\top_t$  a *local top*.

**Lemma 1** Let  $F$  be a pre-confluence, then for any  $t$  in  $F$  and  $x, y \in F \cap \uparrow t$

1.  $\uparrow t$  is a lattice with as join, denoted as  $x \vee_F y$ , the least element of  $\uparrow x \cap \uparrow y$
2. Let  $t' \geq t$  then  $\uparrow t'$  is a sublattice of  $\uparrow t$ .

**Proof**

1. As  $F$  is a pre-confluence,  $\uparrow t$  is a finite  $\wedge$ -semilattice (with meet  $x \wedge_t y$ ) and has a top element ( $\top_t$ ). As a consequence of a well known result on lattice theory,  $\uparrow t$  is lattice. The join  $x \vee_t y$  is the least upper bound of  $\{x, y\}$  in  $\uparrow t$ , i.e. the least element of  $\uparrow t \cap \uparrow x \cap \uparrow y$  which is also  $\uparrow x \cap \uparrow y$ , as both  $x$  and  $y$  are greater than or equal to  $t$ . As it does not depend on  $t$  we simply denote it as  $x \vee_F y$ .
2. For any  $t' \geq t$  and  $x, y$  in  $\uparrow t'$ ,  $x, y$  also belong to  $\uparrow t$ . As a consequence,  $x \wedge_{t'} y$  is also a lower bound of  $\{x, y\}$  in  $\uparrow t$ , and therefore  $t' \leq x \wedge_{t'} y \leq x \wedge_t y$ . But this means that  $x \wedge_t y$  belongs to  $\uparrow t'$  and therefore is also smaller than or equal to  $x \wedge_{t'} y$ . As a consequence we have that  $x \wedge_{t'} y = x \wedge_t y$ . As  $\uparrow t'$  has same meet and join as  $\uparrow t$ , it is a sublattice of  $\uparrow t$ .  $\square$

Furthermore we only need minimal elements of  $F$  to check whether  $F$  is a pre confluence : whenever there is a local meet and a local top on the up set of minimal elements, there is also a local meet and a top element in the up set of any element of  $F$ .

**Lemma 2**  $F$  is a pre-confluence if and only if for any  $m \in \min(F)$ ,  $\uparrow m$  is a  $\wedge$ -semilattice and has a top element.

**Proof** if  $F$  is a pre-confluence, as  $M \subseteq F$  obviously all  $\uparrow m$  are  $\wedge$ -semilattices and have a Top element. Now

suppose that all elements  $m$  of  $M$  are such that  $\uparrow m$  is a  $\wedge$ -semilattice and has a Top element, then consider some  $t \geq m$  and two elements  $t_1, t_2 \in \uparrow t$ , we have then that  $t_1, t_2 \in \uparrow m$ . We know that  $t_1 \wedge_m t_2$  is the greatest lower bound of  $\{t_1, t_2\}$  in  $\uparrow m$  and as  $t$  is a lower bound of  $\{t_1, t_2\}$  and  $t \in \uparrow m$ , we have that  $t_1 \wedge_m t_2 \in \uparrow t$ . As a consequence  $t_1 \wedge_m t_2$  is also the greatest lower bound of  $\{t_1, t_2\} \in \uparrow t$  and so  $t_1 \wedge_t t_2$  exists and this means that  $\uparrow t$  is a  $\wedge$ -semilattice. Furthermore,  $\top_m$  also belongs to  $\uparrow t$  and therefore  $\uparrow t$  also has a greatest element. As for any  $t \in F$  there exists some  $m \in M$  such that  $t \geq m$ , then  $F$  is a pre-confluence.  $\square$

**Definition 4** A subset  $C$  of a pre-confluence  $F$  is called *closed under local meet* whenever for any element  $t$  and any  $C' \subseteq C \cap \uparrow t$  we have

$$\bigwedge_{t \in C'} c \text{ belongs to } C.$$

This means in particular that  $\top_t = \bigwedge_{t \in \emptyset} c$  belongs to any subset which is closed under local meet and then, by definition,  $C$  is also a pre-confluence. The following theorem extends Proposition 2 to pre-confluences :

**Theorem 1** Let  $F$  be a pre-confluence. A subset  $C$  of  $F$  is a closure subset if and only if  $C$  is closed under local meet. The closure  $f : F \rightarrow F$  is then defined as  $f(t) = \bigwedge_{t \in C \cap \uparrow t} c$  and  $C = f[F]$  is a pre-confluence.

**Proof**

We use Proposition 2 and the fact that  $\uparrow t$  in a pre-confluence is a lattice.

—  $\Rightarrow$   $C$  is a closure subset of  $F$  means that there exists a closure operator  $f : F \rightarrow F$  such that  $f[F] = C$ . As  $F$  is a pre-confluence, for any  $t \in F$ ,  $C^t = \uparrow t \cap C$  is a lattice with meet operator  $\wedge_t$ . Furthermore, for any  $x \in \uparrow t$ , we have that  $f(x) \in \uparrow t$  (extensivity of  $f$ ). We can then define  $f_t : \uparrow t \rightarrow \uparrow t$  such that for any  $x \in \uparrow t$ ,  $f_t(x) = f(x)$ . It is straightforward that  $f_t$  is a closure on  $\uparrow t$  as  $f$  is a closure on  $F$ .

As a result, from Proposition 2 we have that  $C^t = f_t[\uparrow t]$  is closed under the meet operator  $\wedge_t$  of  $\uparrow t$ . But, as this is true for any  $t$  in  $F$ , this also means that  $C = \bigcup_{t \in F} C^t$  is by definition closed under local meet.

—  $\Leftarrow$  Let  $C$  be a subset of  $F$  closed under local meet, and let for any  $t$  in  $F$ ,  $C^t = \uparrow t \cap C$ . By hypothesis, for any  $x, y \in \uparrow t$ ,  $x \wedge_t y$  belongs to  $C$ , and as  $x \wedge_t y$  is the greatest lower bound of  $x$  and  $y$  in  $\uparrow t$ , we have that  $x \wedge_t y$  belongs to  $C^t$ . This means that  $C^t$  is a subset of the lattice  $\uparrow t$  and is closed under the meet operator. As a result of Proposition 2 we have then that there exists a closure  $f_t : \uparrow t \rightarrow \uparrow t$  which is such that for any  $x \in \uparrow t$ ,  $f_t(x) = \bigwedge_{t \in \uparrow t \cap C^t} c$ . Furthermore, as  $x \in \uparrow t$ , we have that  $\uparrow x \cap C^t = \uparrow x \cap C$  and therefore  $f_t(x) = \bigwedge_{t \in \uparrow t \cap C^t} c$  and also as  $\uparrow x$  is a sublattice of  $\uparrow t$ ,  $f_t(x) = f_x(x) = \bigwedge_{x \in \uparrow t \cap C^t} c$ .

Let then define  $f : F \rightarrow F$  as  $f(x) = f_x(x)$ . It is straightforward that  $f$  is a closure :

- $f(x) = f_t(x)$  for any  $t \leq x$ , therefore as  $f_t$  is a closure,  $f_t(x) \geq x$ . As there always exists such a  $t$ , then  $f(x) \geq x$
- if  $x \geq y$  we have some  $t$  such that  $x, y \in \uparrow t$ , therefore  $f(x) = f_t(x)$  and  $f(y) = f_t(y)$  and therefore  $f(x) \geq f(y)$ .
- We have that  $f(x) \geq x$  and there is some  $t$  in  $F$  such that  $f(x), x$  both belong to  $\uparrow t$ , therefore  $f(f(x)) = f_t(f_t(x)) = f_t(x) = f(x)$ .

□

As a summary, we have a generalization of the meet operator which is the basis of most work on closed patterns in data mining, as well as all work on formal concept analysis. This generalization, denoted as local meet operator ensures the existence of closure operators whose ranges are subsets closed with respect to the local meet operator. Whenever we consider a pre-confluence as a subset of a finite powerset  $2^X$  we call  $F$  also a pre-confluent family. A typical example of such a structure is the set of subgraphs generated by the vertices (or edges) of a given graph. We consider here the family  $F = \{a, b, abc, abd, abcd\}$  which diagram is represented in the leftmost part of Figure 1. Here we have that  $abc \wedge_a abd = a$  and  $abc \wedge_b abd = b$  i.e. there are two maximal lower bounds of  $abc$  and  $abd$  in  $F$  because  $ab$  does not belong to  $F$ . Note that the up sets  $F^a$  and  $F^b$  are lattices, and share the same join operator, which in this case is the union operator.

### 3 Support closed patterns

#### 3.1 Support closures in lattices

The standard case in which closed patterns are searched for is when the language is a lattice and that closure of a pattern relies on the occurrences of the pattern in a set of objects. In data mining the set of occurrences is known as the support of the pattern whereas in Formal concept analysis the set of occurrences defines the extension of the pattern and the extent of the corresponding concept.

We give hereunder a general notion of occurrence

**Definition 5** Let  $F$  be a partial order and  $O$  a set of objects, a relation of occurrence on  $F \times O$  is such that if  $t_1 \geq t_2$  and  $t_1$  occurs in  $o$  then  $t_2$  occurs in  $o$ .

The extension of  $t$  in  $O$  is defined as  $\text{ext}(t) = \{o \in O \mid t \text{ occurs in } o\}$ .

The cover of  $o$  is defined as the part of  $F$  whose elements occur in the object  $o$ , i.e.  $S(o) = \{t \in F \mid t \text{ occurs in } o\}$ .

The cover of a subset  $e$  of objects is defined as the part of  $F$  whose elements occur in all objects of  $e$ , i.e.  $S(e) = \bigcap_{\{o \in e\}} S(o)$ .

We will say hereafter indifferently that  $t$  belongs to the cover of  $o$ , or that  $t$  occurs in  $o$ . We consider the standard case where we start from a lattice  $T$  in which each object  $o$  of

$O$  has a description  $d(i)$ , and we further consider that any element of  $T$  can be such a description. We are then interested in which subsets  $F$  of  $T$  have support-closures with respect to any  $O$ . We connect here to the seminal result of M. Boley and collaborators [3] on confluent systems. To avoid confusion, up sets and down sets of a partial order  $E$  starting from an element  $x$  will be denoted respectively as  $E^x$  and  $E_x$ .

We will need the following lemma to characterize how an object, as an element  $x$  of  $T$ , can be represented in  $F$ .

**Lemma 3** Let  $F$  be a subset of a lattice  $T$ .

If for any  $t \in F$  and any  $x \in T^t$ , there exists a greatest element  $p_t(x)$  in  $F^t \cap T_x$ , then the mapping  $p_t : T^t \rightarrow T^t$  is a projection on the lattice  $T^t$  and  $p_t(T^t) = F^t$

$p_t(x)$  is the local description of  $x$  in  $F^t$ .

**Proposition 3** Let  $F$  be a subset of a lattice  $T$ , the three following properties are equivalent :

1. For any  $t \in F$  and any  $x \in T^t$ , there exists a greatest element  $p_t(x)$  in  $F^t \cap T_x$
2. For any  $x, y, t$  in  $F$  with  $x \geq t$  and  $y \geq t$ , we have that  $x \vee y$  belongs to  $F$
3.  $F$  is a pre-confluence with join  $\vee_F = \vee$

$F$  is then denoted as a confluence\* on  $T$  and we have that  $p_t(x) = \bigvee_{q \in F^t \cap T_x} q$

**Proposition 4** Let  $F$  be a confluence\* of a lattice  $T$  and  $O$  a set whose objects are described as elements of  $T$ , then : Let  $p_t$  denote the local description operators on  $F$ , we have that

$f(t) = p_t \circ \text{int} \circ \text{ext}(t)$  where  $(\text{int}, \text{ext})$  is a Galois connection on  $(T, O)$ , is a support closure operator on  $F$  with respect to  $O$ .

Conversely, in order to guarantee that such a support closure operator exists for any set of objects  $O$  described in  $T$ , a subset of  $T$  has to be a confluence\* :

**Proposition 5** Let  $F$  be a subset of the lattice  $T$ , then the support closure operator on  $F$  with respect to any set  $O$  whose objects are described as elements of  $T$  exists if and only if  $F$  is a confluence\*.

In Boley and collaborators, the lattice  $T$  is a powerset  $2^X$  and a confluent system  $S$  is similar to the latter definition of confluences\* except that  $\perp = \emptyset$  belongs to  $S$  but  $x \cup y$  is only required to belong to  $F$  when  $x \supseteq t$  and  $y \supseteq t$  for any  $t! = \emptyset$ . Proposition 5 is a straightforward adaptation and rewriting of the theorem of Boley and collaborators in the case in which  $T = 2^X$ , where confluent systems replaces confluences\*, and which prohibits to have any attribute common to all objects in  $O$  in order to ensure a greatest element in the cover of  $\emptyset$ .

A useful Lemma is the following :

**Lemma 4** *If  $F$  is a confluence\*, then if  $q \leq t$ , and  $x \in T^q$ , then  $p_t(x) = p_q(x)$*

This means that to compute the support closure of some  $t$  we only need  $p_m$  where  $m \in \min(F)$ . Implicitly this also means that whether  $t$  is greater than two minimal elements  $m$  and  $m'$  then  $p_m(t) = p_{m'}(t)$ . This is interesting as, these minimal elements are in general well known and the corresponding projection easy to define.

To summarize, the support closure set  $f(F)$  of a confluence\*  $F$  on some lattice  $T$ , forms a pre-confluence of  $T$ , made of projected Galois lattices and we only need the minimal elements of  $F$  to characterize the pre-confluence  $f[F]$ . When considering  $T = 2^X$ ,  $T^t$  is  $2^{X \setminus t}$  and  $p_t$  is a projection on  $2^{X \setminus t}$ .

### 3.2 Implications

Another question regards the definition and construction of an implication basis whose implications have both left part and right part in  $F$ . An implication  $p \rightarrow q$  holds on  $F$  whenever  $\text{ext}(p) \subseteq \text{ext}(q)$  and a basis of such implications is typically made of implications such that both  $p$  and  $q$  belong to the same equivalence class i.e.  $\text{ext}(p) = \text{ext}(q)$ . Whenever  $F$  is a lattice, the nodes of the concept lattice represents these equivalence classes and  $q$  is a closed pattern i.e. the greatest element of the class, and therefore we have  $p \leq q$ . As an example the *min-max basis* is made of the implications  $p \rightarrow q$  where  $p \neq q$  and  $p$  is a minimal element of the class of  $q$  [8]. Whenever  $F$  is a confluence\*, we have seen that each such equivalence class is associated to several closed patterns  $q_1 \dots q_m$  each being the greatest element of a subclass. We have then in the basis both implications of the form  $p_i \rightarrow q_i$  where  $p_i \leq q_i$  and both belong to subclass  $i$  together with implications of the form  $p_j \rightarrow q_i$  where  $j \neq i$  and therefore  $p_j$  and  $q_j$  are unordered. We extend the idea of the min-max basis to confluences\* as follows :

**Definition 6** *Let  $F$  be a confluence\*, and  $F(e) = \{t \in F \mid \text{ext}(t) = e\}$ , the min-max basis  $B = B_i \cup B_e$  of implications in  $F$  is defined as the set*

$$\{p \rightarrow q \mid \text{ext}(p) = \text{ext}(q), p \neq q, p \in \min(F(e)), q \in f[F(e)]\}$$

*The internal sub basis  $B_i$  is made of the implications of the form  $p_i \rightarrow q_i$  where  $p_i \leq q_i$  and the external sub basis  $B_e$  is made of the implications of the form  $p_j \rightarrow q_i$  where  $\{p_j, q_j\}$  are unordered.*

There are other implication basis such as the minimal Guigue-Duquenne basis [6] that can be as well extended to the case of confluences\*.

### 3.3 Example

We consider here the example displayed in Figure 1. We have  $F = \{a, b, abc, abd, abcd\}$  and  $O = \{ab, abc, abcd\}$ . To compute the closures in  $F$  we take advantage of the fact that  $F$  has two minimal elements  $a$  and  $b$  and that for any

$t \geq a$  (resp.  $t \geq b$ ) we can write  $f(t) = p_a \circ \text{int} \circ \text{ext}(t)$  (resp.  $f(t) = p_b \circ \text{int} \circ \text{ext}(t)$ ). We obtain then :

- $f(a) = p_a \circ \text{int}(\{ab, abc, abcd\}) = p_a(ab) = a$
- $f(b) = p_b \circ \text{int}(\{ab, abc, abcd\}) = p_b(ab) = b$
- $f(abc) = p_a \circ \text{int}(\{abc, abcd\}) = p_a(abc) = abc$   
(we could have used  $p_b$  as  $abc \in T^b$  with the same result  $abc$ )
- $f(abd) = p_a \circ \text{int}(\{abcd\}) = p_a(abcd) = abcd$   
(same remark as above)
- $f(abcd) = p_a \circ \text{int}(\{abcd\}) = p_a(abcd) = abcd$   
(same remark as above)

Note that the confluence\*  $F$  is the union of the two lattices  $F^a = \{a, abc, abd, abcd\}$  and  $F^b = \{b, abc, abd, abcd\}$ . Therefore we have  $f[F] = \{a, b, abc, abcd\}$  which is a pre-confluence whose minimal elements are  $f(a) = a$  and  $f(b) = b$ . We have that  $f[F] = f[F^a] \cup f[F^b]$  where  $f[F^a]$  and  $f[F^b]$  are the sets of closed patterns from the concept lattices built respectively from  $(F^a, O^a)$ , and from  $(F^b, O^b)$ . We have here  $f[F^a] = \{a, abc, abcd\}$  and  $f[F^b] = \{b, abc, abcd\}$ .

Regarding the min-max implication basis we first consider the set of extensions  $\text{ext}[F] = \{e_1 = \{ab, abc, abcd\}, e_2 = \{abc, abcd\}, e_3 = \{abcd\}\}$  together with the corresponding equivalence classes  $F(e_1), F(e_2), F(e_3)$ . Each each equivalence class is divided into subclasses each containing one closed element :

- $F(e_1) = \{a\} + \{b\}$
- $F(e_2) = \{abc\}$
- $F(e_3) = \{abd, abcd\}$

Figure 1 displays on the left the confluence\*  $F$ , on the middle we have the object set  $O$ , and on the right is represented the pre-confluence  $f[F]$  of support closed patterns of  $F$ . The min-max implication basis is made of the internal basis  $B_i = \{abc \rightarrow abcd\}$  (this implication holds both in  $(F^a, O^a)$  and in  $(F^b, O^b)$ ) plus the external basis  $B_e = \{a \rightarrow b, b \rightarrow a\}$ .

## 4 Abstract closed patterns in confluences\*

In this section we consider abstract closed patterns as those obtained in *extensionally abstract Galois lattices*, denoted here as abstract Galois lattices for short, by constraining the space  $2^O$ . The general idea, as proposed in [9] and resulting in Proposition ?? in section 3.1 is that an abstract Galois lattice is obtained by selecting as an extensional space a subset  $A$  of  $2^O$  closed under union i.e. an abstraction (or dual closure subset) and therefore such that  $A = p_A(2^O)$  where  $p_A$  where  $p_A$  is an interior operator on  $2^O$ . The intuitive meaning is that the abstract extension  $\text{ext}_A(t)$  of some pattern  $t$  will then be the union of the elements of  $A$  contained in its (standard) extension, i.e.  $\text{ext}_A = p_A \circ \text{ext}$  and the corresponding *abstract support closure operator with respect to  $A$*  is therefore  $f_A = \text{int} \circ p_A \circ \text{ext}$ . Intuitively, as noticed in [10], this is because the corresponding abstract Galois lattice is isomorphic, and as same support closure subset as the Galois lattice associated to the object set

$O(A)$  each object  $a$  of which is an element of  $A$  and described in  $T$  as  $\text{int}(a)$ <sup>2</sup>. It is then straightforward that we obtain that abstract Galois pre-confluences are simply the Galois pre-confluences obtained through this change on object set.

**Theorem 2** *Let  $F$  be a confluence\* of a lattice  $T$ ,  $O$  a set whose objects are described as elements of  $T$ ,  $A = p_A(O)$  an abstraction of  $A$ , then :*

*Let  $p_t$  denote the local description operators on  $F$ , we have that*

*$f_A(t) = p_t \circ \text{int} \circ p_A \circ \text{ext}(t)$ , where  $(\text{int}, p_A \circ \text{ext})$  is a Galois connection on  $(T, A)$ , is a support closure operator on  $F$  with respect to  $A$  and  $f_A[F]$  is a pre-confluence.*

We continue here the example of section 3.3 by using the abstraction

$A = \{\{o_1, o_2\}, o_1, o_3\} = \{\{ab, abc\}, \{ab, abcd\}\}$ . Recall that  $p_A(e) = \cup_{\{a \in A | a \subseteq e\}} a$ . We obtain then :

- $f_A(a) = p_a \circ \text{int} \circ p_A(\{o_1, o_2, o_3\}) = a$  as  $p_A(\{o_1, o_2, o_3\}) = \{o_1, o_2, o_3\} = \{ab, abc, abcd\}$
- $f_A(b) = p_b \circ \text{int} \circ p_A(\{o_1, o_2, o_3\}) = b$  (same reason as above)
- $f_A(abc) = p_a \circ \text{int} \circ p_A(\{o_2, o_3\}) = \top_a = abcd$  as  $p_A(\{o_2, o_3\}) = \emptyset$  and therefore  $p_a \circ \text{int}(\emptyset) = p_a(\top_a) = \top_a$
- $f_A(abd) = p_a \circ \text{int} \circ p_A(\{o_3\}) = \top_a = abcd$  as  $p_A(\{o_3\}) = \emptyset$  (as above)
- $f_A(abcd) = p_a \circ \text{int} \circ p_A(\{o_3\}) = abcd$  (same as above)

$F$  is represented on the left of Figure 2. The corresponding abstract support closure pre-confluence  $f_A[F]$  is displayed on the right of the figure. What happens here, is that there are only two possible extensions as  $\text{ext}_A[F] = \{\emptyset, O\}$ . As a result the two minimal elements of  $f_A[F]$  share the same abstract extension  $O$  whereas the unique maximal element  $\top_a = \top_b = abcd$  have an empty abstract extension.

## 5 Algorithmics

An algorithm to build closure support on confluent families on  $2^X$  has been proposed in [3] whenever  $F$  is *strongly accessible*. This restriction<sup>3</sup> ensures a polynomial delay in outputting support closed elements. This algorithm has further been implemented as a generic tool and in order to be efficient on multicores architectures particular in PARAMINER [7]. Adapting it to confluences\* is straightforward by avoiding computing the support closure of  $\emptyset$ . Basically, the algorithm performs a depth-first search each step of which consists in adding an attribute  $x$  to the current closed pattern  $t$ , checking whether the resulting pattern

2. In fact we just need the  $\cup$ -irreducible elements of  $A$  as objects  
3. For  $(F, X)$  to be a strongly accessible set system, it is required that between any pair of elements  $t_1, t_2$  with  $t_1 \leq t_2$  in  $F$  there is a path  $t_1, t_1 \cup \{x_1\}, \dots, t_1 \cup \{x_1, \dots, x_k\} = t_2$  all elements of which belong to  $F$ .

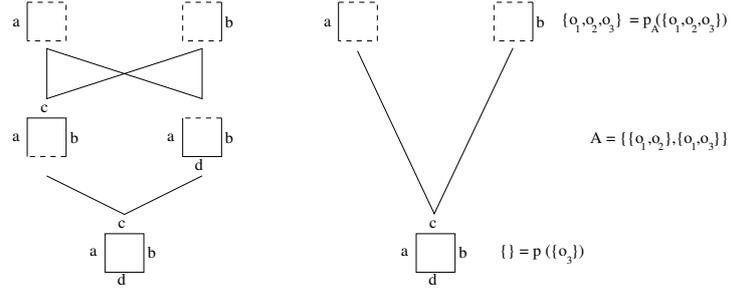


FIGURE 2 – Diagram of the abstract support closed connected subgraphs pre-confluence  $f_A[F]$  (on the left part of the figure) with respect to the abstraction  $A = \{\{o_1, o_2\}, \{o_1, o_3\}\}$  of  $O$ . The support closed element  $abc$  of  $f[F]$  as been projected to the maximal element of  $F$ ,  $abcd$ , because its extension  $\{abc, abcd\}$  is projected on  $\emptyset$  as no element of  $A$  is included in  $\{abc, abcd\}$ .

$t \cup \{x\}$  is in  $F$ , and closing the pattern. A SELECT function states whether a pattern belongs to  $F$  and closure is only computed if it returns TRUE. The function has an ad hoc implementation according to the problem in hand. In terms of interior operators, SELECT implicitly tests whether  $p_t(t \cup \{x\}) = t \cup \{x\}$  is true. A CLOSURE function computes the closure of any  $t \in F$  by implicitly applying  $p_t$  to  $\text{int}(\text{ext}(t))$ . Again the implementation is ad hoc, depending of the problem at hand. An open question is the construction and visualisation of the diagram of the pre-confluence of support closed elements and of the corresponding min-max implication basis.

## 6 Conclusion

Motivated by the problem of finding closed patterns in languages as the set of connected subgraphs of a graph, we have investigated an extension of FCA where the pattern language is a pre-confluence, i.e. a partial order defined through the existence of a local meet operator, and that can be expressed as a constrained union of a set of lattices. We have first extended the standard property that relates closure subsets and subsets closed under the meet operator to the case of pre-confluences. Then we have discussed the existence of support-closure operators in pre-confluences, extending a result of [3]. We have also shown that applying interior operators to the powerset of objects we obtain, as in the lattice case, abstract support closures. The connection to FCA we have attempted to rises some technical questions, as the construction of diagrams of closure subsets, as well as more fundamental questions. For instance, when considering a support closed element as the intensional part of some concept, i.e. an *intent*, we may have two different concepts with the same *extent* which is somewhat disturbing. On the other hand, we could consider that the extension defines the concept, i.e. is an *extent* and in this case, a concept may have several *intents*. Finally, regarding appli-

cations, it seems worthwhile to consider such structures, as they are frequent when modeling data using graphs.

## Références

- [1] Hiroki Arimura and Takeaki Uno. Polynomial-delay and polynomial-space algorithms for mining closed sequences, graphs, and pictures in accessible set systems. In *SDM*, pages 1087–1098. SIAM, 2009.
- [2] Thomas S. Blyth. *Lattices and Ordered Algebraic Structures*. Universitext, Springer, 2005.
- [3] 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.
- [4] Nathalie Caspard and Bernard Monjardet. The lattices of closure systems, closure operators, and implicational systems on a finite set : a survey. *Discrete Appl. Math.*, 127(2) :241–269, 2003.
- [5] B. Ganter and R. Wille. *Formal Concept Analysis : Mathematical Foundations*. Springer Verlag, 1999.
- [6] J.L. Guigues and V. Duquenne. Famille non redondante d’implications informatives résultant d’un tableau de données binaires. *Mathématiques et Sciences humaines*, 95 :5–18, 1986.
- [7] Benjamin Negrevergne, Alexandre Termier, Marie-Christine Rousset, and Jean-François Méhaut. Paraminer : a generic pattern mining algorithm for multi-core architectures. *Data Mining and Knowledge Discovery*, pages 1–41, 2013.
- [8] Nicolas Pasquier, Rafik Taouil, Yves Bastide, Gerd Stumme, and Lotfi Lakhal. Generating a condensed representation for association rules. *Journal Intelligent Information Systems (JIIS)*, 24(1) :29–60, 2005.
- [9] Henry Soldano and Véronique Ventos. Abstract Concept Lattices. In P. Valtchev and R. Jäschke, editors, *International Conference on Formal Concept Analysis (ICFCA)*, volume 6628 of *LNAI*, pages 235–250. Springer, Heidelberg, 2011.
- [10] Véronique Ventos and Henry Soldano. Alpha Galois lattices : An overview. In B. Ganter and R. Godin (Eds), editors, *International Conference on Formal Concept Analysis (ICFCA)*, volume 3403 of *Lecture Notes on Computer Science*, pages 298–313. Springer Verlag, 2005.