

# Abstract and Local Rule Learning in Attributed Networks

Henry Soldano<sup>1,2</sup>, Guillaume Santini<sup>1</sup>, and Dominique Bouthinon<sup>1</sup>

<sup>1</sup> L.I.P.N UMR-CNRS 7030, Université Paris 13, Sorbonne Paris Cité,  
93430 Villetaneuse, France

henry.soldano@lipn.univ-paris13.fr

<sup>2</sup> Atelier de BioInformatique, ISYEB - UMR 7205 CNRS MNHN UPMC EPHE,  
Museum d'Histoire Naturelle, 75005 Paris, France

**Abstract.** We address the problem of finding local patterns and related local knowledge, represented as implication rules, in an attributed graph. Our approach consists in extending frequent closed pattern mining to the case in which the set of objects is the set of vertices of a graph, typically representing a social network. We recall the definition of *abstract closed patterns*, obtained by restricting the support set of an attribute pattern to vertices satisfying some connectivity constraint, and propose a *specificity measure* of abstract closed patterns together with an *informativity measure* of the associated *abstract implication rules*. We define in the same way *local closed patterns*, i.e. maximal attribute patterns each associated to a connected component of the subgraph induced by the support set of some pattern, and also define *specificity* of local closed patterns together with *informativity* of associated *local implication rules*. We also show how, by considering a derived graph, we may apply the same ideas to the discovery of local patterns and local implication rules in non disjoint parts of a subgraph as  $k$ -cliques communities.

## 1 Introduction

We address here the problem of discovering patterns and associated knowledge in an attributed graph. Previous work focuses on the topological structure of the patterns, thus ignoring the vertex properties, or consider only local or semi-local patterns [4]. In [1] patterns on co-variations between vertex attributes are investigated in which topological attributes are added to the original vertex attributes and in [7] the authors investigate the correlation between the support set of an itemset and the occurrence of dense subgraphs. What we propose in this article is to consider a graph  $G = (O, E)$  whose vertices are labelled by itemsets and to submit their occurrences in the vertex set  $O$ , i.e. their *support sets*, to connectivity constraints. We consider attribute patterns in the standard closed itemset mining approach developed in Formal concept Analysis (FCA)[3], Galois Analysis [2], and Data Mining (see for instance [6]).

In pattern mining, a *support-closed* pattern is a pattern which is maximal, in size, i.e. in terms of specificity, within the equivalence class of all patterns  $q$

sharing the same *support set*  $e = \text{ext}(q)$ . The corresponding equivalence relation is simply denoted  $\equiv$ . In standard itemset mining, there is a unique support-closed pattern, i.e. a maximum, in each equivalence class and this support-closed pattern is easily computed using a closure operator  $f$ . More precisely, when considering some pattern  $q$  its equivalence class is made of all patterns whose support set is  $\text{ext}(q)$  and the unique support-closed pattern is obtained as  $f(q) = \text{int} \circ \text{ext}(q)$  where  $\text{int}$  simply intersect the object descriptions of the support set. Support-closed patterns are then simply called *closed patterns*. The set of (support set, closed pattern) pairs is organized within a concept lattice and inclusion of support sets leads to implication rules that hold on the dataset under investigation. The set of frequent closed patterns, i.e. closed elements whose support is greater than or equal to some threshold  $\text{minsupp}$ , represents then all the equivalence classes corresponding to frequent supports. Such a class has also minimal elements, called *generators*. When the patterns belong to  $2^X$ , the min-max basis of implication rules [6] that represents all the implications  $t \rightarrow t'$  that hold on  $O$ , i.e. such that  $\text{ext}(t) \subseteq \text{ext}(t')$ , is defined as follows:

$$\{g \rightarrow f \mid f \text{ is a closed pattern, } g \text{ is a generator, } f \neq g, \text{ext}(g) = \text{ext}(f)\}$$

## 2 Abstract Knowledge

In a previous work [9] the attributed graph  $G = (O, E)$  was investigated in the following way: each pattern support set  $e \subseteq O$ , as a set of vertices, induces a subgraph  $G(e)$  of  $G$ , and this subgraph is then simplified by removing vertices in various ways. The vertices of such an *abstract subgraph* all satisfy some topological constraint, as for instance belonging to a  $k$ -clique, and form the *abstract support set* of the pattern. What happens here is that the extensional space is then reduced to a part  $A$  of  $2^O$ , called a *graph abstraction*, and that can be generated as the union closure of subsets of  $O$  we call *abstract groups*. For instance the  $k$ -clique abstraction is made of union of  $k$ -cliques and therefore the abstract support set of a pattern is the (maximum) subset of its support set made of  $k$ -cliques (Fig. 1).

*Example 1.* Consider the graph  $G = (O, E)$  where  $O = \{1, 2, 3, 4, 5, 6, 7, 8\}$  and  $E = \{12, 13, 23, 34, 45, 56, 67, 57, 68, 78\}$ . Each vertex  $o$  is described by  $d(o) \in 2^{abc}$ , i.e.  $d(1) = d(2) = d(3) = ab, d(4) = d(5) = ac, d(6) = d(8) = bc, d(7) = abc$ . Consider then the 3-clique abstraction  $A$ . The support set of  $a$  is  $\text{ext}(a) = \{1, 2, 3, 4, 5, 7\}$  and induces the subgraph  $G(e)$  whose edges are

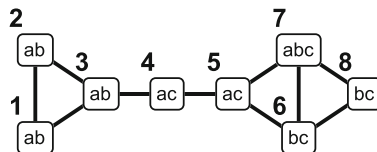


Fig. 1. An attributed graph

$\{12, 23, 13, 34, 45, 57\}$ . Its abstract support set is  $\text{ext}_A(a) = \{1, 2, 3\}$  as no vertex amongst 4, 5, 7 belongs to a triangle in  $G(e)$ .

Abstract support sets are obtained applying an *interior operator*  $p$  such that  $p[2^O] = A$ , i.e.  $\text{ext}_A = p \circ \text{ext}$ . As an interior operator on  $2^O$ ,  $p$  has the following properties: for any  $e, e' \in 2^O$ , *i*)  $p(e) \leq e$ , *ii*)  $p(p(e)) = p(e)$  and *iii*)  $e \leq e' \Rightarrow p(e) \leq p(e')$ . Abstract implications are then defined by considering inclusion of abstract support sets, i.e.  $\Box_A q \rightarrow \Box_A w$  is valid if and only if  $\text{ext}_A(q) \subseteq \text{ext}_A(w)$ . Such an abstract rule has the following meaning “whenever the members of some abstract group share pattern  $q$ , they also share pattern  $w$ ”. Because of the monotony (condition *iii*) of the interior operator  $p$ , abstraction preserves implication validity:

**Lemma 1.** *Let  $A$  be an abstraction,  $q$  and  $w$  two patterns, then  $q \rightarrow w \Rightarrow \Box_A q \rightarrow \Box_A w$*

In the case of the  $k$ -clique abstraction mentioned above, this means that by restricting the support sets of patterns to be made of  $k$ -cliques, we preserve previous valid implications and possibly obtain some new valid abstract implications representing abstract knowledge.

Consider then the equivalence relation  $\equiv_A$  defined by  $q \equiv_A w$  iff  $\text{ext}_A(q) = \text{ext}_A(w)$ . Equivalence classes of  $\equiv_A$  have a maximum obtained, by applying the closure operator  $\text{int} \circ p \circ \text{ext}$  and called an *abstract closed pattern*, while its minimal elements are called  $A$ -generators. We then obtain the *abstract min-max basis* of abstract implications rules where  $\text{ext}_A$  replaces  $\text{ext}$ . The abstract min-max basis is made of abstract implications relating  $A$ -generators of some equivalence class of  $\equiv_A$  to the abstract closed pattern of the same class:

$$\{\Box_A g \rightarrow \Box_A c \mid c \text{ is an } A\text{-closed pattern, } g \text{ is a } A\text{-generator, } c \neq g, \text{ext}_A(g) = \text{ext}_A(c)\}$$

*Example 2.* Consider the data and 3-clique abstraction of Example 1. Intersecting the vertex descriptions of  $\text{ext}_A(a) = \{1, 2, 3\}$  we obtain the abstract closed pattern  $ab$ . The equivalence class of patterns having abstract support set  $\{1, 2, 3\}$  is  $\{a, ab\}$  and  $a$  is therefore a  $A$ -generator. This means that  $\Box a \rightarrow \Box ab$  belongs to the *abstract min-max basis* extracted from  $G$  and means “whenever the vertices of a triangle in  $G$  share pattern  $a$ , they also share pattern  $ab$ ”. Note that  $a \rightarrow b$  was not a valide rule, i.e. when considering some vertex  $o$  to infer  $b$  from  $a$  we have to consider some triangle to which  $o$  belongs and whose two other vertices also have  $a$ .

### 3 Measuring Abstract Knowledge

When considering frequent abstract closed patterns, we are interested in ordering or selecting them according to to what extent they are related to the graph structure. For that purpose we generalize hereunder the *structural correlation* measure introduced by A. Silva and co-authors [7], originally introduced to compute the ratio of vertices involved in quasi-cliques in the subgraph induced by a pattern, and rename it as *specificity*.

**Definition 1.** Let  $q$  be a pattern,  $A$  an abstraction of some powerset of objects  $O$ , the specificity of  $q$  with respect to  $A$  is defined as:

$$s_A(q) = \frac{|\text{ext}_A(q)|}{|\text{ext}(q)|}$$

Apart from measuring through specificity what is specific to the pattern in its abstract view, we are also interested when considering abstract rules in how informative they are. For that purpose we consider abstract rules whose left and right patterns are equivalent in the abstract space  $A$ , i.e. have same abstract support set, as in the min-max abstract rule basis defined above. Whenever these patterns are also equivalent in the original space  $2^O$  intuitively the rule is uninformative. Assume for instance that both  $a \rightarrow abc$  and  $\Box_A a \rightarrow \Box_A abc$  are valid, then the abstract rule did not bring any new information. On the contrary, assume that  $\Box_A a \rightarrow \Box_A abc$  is valid while  $a \rightarrow abc$  has only confidence 0.5, i.e.  $\text{ext}(abc) = 0.5 * \text{ext}(a)$ , then clearly the abstract rule brings some information. We simply measure here *informativity* as the inverse of confidence.

**Definition 2.** Let  $q$  be a pattern,  $A$  an abstraction of  $2^O$ , the informativity of the valid rule  $r : \Box_A q \rightarrow \Box_A w$  is defined as:

$$I_A(r) = \frac{|\text{ext}(q)|}{|\text{ext}(qw)|}$$

An alternative Informativity measure, ranging between 0 and 1, would be the (estimated) probability of not having  $w$  whenever we have  $q$  i.e.  $1 - \frac{|\text{ext}(qw)|}{|\text{ext}(q)|}$ . This quantity has value 0 whenever  $q \rightarrow w$  holds and has limit 1 whenever  $|\text{ext}(qw)|$  approaches 0, i.e. restricting the support set of patterns to elements of  $A$  concentrates the support set of  $q$  to the very few sharing also  $w$ . In the remaining of the article we keep Definition 2 to define informativity.

Considering an implication rule from the abstract min-max basis  $\Box_A g \rightarrow \Box_A c$ , we are then interested in the specificity  $s_A(c)$  of the abstract closed pattern and in the informativity  $I_A(r) = \frac{|\text{ext}(g)|}{|\text{ext}(c)|}$  of the rule.

*Example 3.* Considering the attributed graph and triangle abstraction of Examples 1 and 2,  $ab$  has specificity  $3 \div 6 = 0.5$  while  $\Box_A a \rightarrow \Box_A ab$  has informativity  $6 \div 4 = 1.5$ .

## 4 Local Knowledge

Given some attribute pattern, we are now interested in extracting *local support closed patterns*, i.e. maximal attribute patterns each associated to one dense subgraph, so allowing to extract *local implication rules* particular to specific dense groups of objects. Recently the closed pattern mining methodology has been extended to *local closed patterns*: they are obtained by applying a set of *local closure operators* [8]. In the graph case, this means that from the support set of some (closed) pattern  $c$ , various dense support sets, called *local support*

sets are extracted each associated to a *local closed pattern*, i.e. the most specific pattern  $l$  common to the elements of the local support set. Again we obtain a set of *local implication rules* corresponding to inclusion of local support sets, but now such an implication is only valid in the vicinity of some dense group of vertices.

#### 4.1 Direct Local Knowledge

The simplest case appears when the extensional space is reduced to the set  $F$  of connected subgraphs induced by vertex subsets belonging to some graph abstraction  $A$ . To a pattern  $q$  is associated one of its connected component  $e$  as a local support set, and  $\text{int}(e)$  as the corresponding local closed pattern. We may then consider, for instance, as  $A$  the 3-clique abstraction and obtain as local support sets connected subgraphs made of 3-cliques. In this simple case,  $F$  is a *confluence* of  $A$  [8], i.e. a partially ordered set made of several lattices, and that has in general a set  $\text{min}(F)$  of minimal elements. More precisely, in our connected 3-clique subgraphs case, these minimal elements are the 3-cliques of our graph  $G$ . We call such a confluence, whose elements are connected components, a *cc-confluence*. Let  $q$  be a pattern,  $m \in \text{min}(F)$ , and  $m \subseteq \text{ext}_A(q)$ , we obtain the connected component containing the 3-clique  $m$  as  $\text{ext}_m^A = p_m \circ \text{ext}_A(q)$  where  $p_m$  is again an interior operator, and therefore is monotonic. Note that in a cc-confluence, each vertex appears in only one such connected components and we may as well replace  $m$  by one of its vertex  $s$  in our definitions.

Whenever we have  $p_m \circ \text{ext}_A(q) \subseteq p_m \circ \text{ext}_A(w)$  we rewrite this as the local implication  $\square_m^A q \rightarrow \square_m^A w$  stating that if  $q$  has a local support set containing  $m$ , then  $w$  has a larger than or equal to local support set. Because of monotonicity of  $p_m$ , again validity of implications is preserved:

**Lemma 2.** *Let  $F$  be a confluence of an abstraction  $A$ ,  $q$  and  $w$  two patterns, then  $\square_A q \rightarrow \square_A w \Rightarrow \square_m^A q \rightarrow \square_m^A w$*

When considering a given abstract closed pattern  $c$  which has a local support set  $e$  in  $F$  that contains  $m$ , and whose corresponding local closed pattern is  $l$ , we have then that the implication rule  $\square_m^A c \rightarrow \square_m^A l$  holds. The set  $\{\square_A c \rightarrow \square_A l \mid l \text{ a local closed pattern, } c \text{ an abstract closed pattern, } c \neq l, \text{ext}_m^A(c) = \text{ext}_m^A(l)\}$  represents (a basis for) the local knowledge deriving from the reduction of the extensional space from  $A$  to the confluence  $F$ .

*Example 4.* Still considering the attributed graph  $G$  and triangle abstraction of Examples 1 and 2, we consider the cc-confluence  $F$  of vertex subsets inducing connected subgraphs of  $G$  made of triangles. We have  $\text{ext}_A(b) = \{1, 2, 3, 6, 7, 8\}$  that induces a subgraph made of two connected components  $\{1, 2, 3\}$  and  $\{6, 7, 8\}$ . The corresponding local closed patterns are  $\text{int}(\{1, 2, 3\}) = ab$  and  $\text{int}(\{6, 7, 8\}) = bc$ . As  $b = \text{int}(\{1, 2, 3, 6, 7, 8\})$ ,  $b$  is an abstract closed pattern and we have the following local implications:  $\square_A^{\{1,2,3\}} b \rightarrow \square_A^{\{1,2,3\}} ab$  and  $\square_A^{\{6,7,8\}} b \rightarrow \square_A^{\{6,7,8\}} bc$  we may rewrite, since  $A$  is a cc-confluence, as, for instance:  $\square_A^1 b \rightarrow \square_A^1 ab$  and  $\square_A^6 b \rightarrow \square_A^6 bc$ .

### 4.2 Measuring Direct Local Knowledge

To measure how much a local closed pattern is specific to the associated connected component, and in the same way as in the abstract case where we considered the ratio between the abstract and standard support sets, we are here interested in the ratio between the local and the global (standard or abstract) support set:

**Definition 3.** *Let  $q$  be a pattern,  $F$  an extensional confluence of some abstraction  $A$  of  $2^O$ , and  $m \in F$  such that  $m \subseteq \text{ext}_A(q)$ , the specificity of  $q$  in the vicinity of  $m$  is defined as:*

$$s_F(q, m) = \frac{|\text{ext}_m^A(q)|}{|\text{ext}_A(q)|}$$

In the same way as in the abstract implication case, we measure informativity of a local rule with respect to the corresponding global rule. The idea here is that in a valid local implication the patterns left and (left+)right have same local support set while their global support sets are different. Again *informativity* is defined as the inverse of the (abstract) confidence.

**Definition 4.** *Let  $q$  be a pattern,  $F$  an extensional confluence of some abstraction  $A$  of  $2^O$ , and  $m \in F$  such that  $m \subseteq \text{ext}_A(q)$ , the informativity of the valid local rule  $r : \square_m^A q \rightarrow \square_m^A w$  is defined as:*

$$I_F(r) = \frac{|\text{ext}_A(q)|}{|\text{ext}_A(qw)|}$$

Intuitively, informativity measures what we have learned when discovering that  $q$  and  $qw$  had same local support sets with respect to  $m$  while they had different abstract support set. Considering a local implication rule  $r : \square_m^A c \rightarrow \square_m^A l$  we are interested in the specificity  $s_F(l, m)$  of the local closed pattern  $l$  and in the informativity  $I_F(r) = \frac{|\text{ext}_A(c)|}{|\text{ext}_A(l)|}$  of the rule.

*Example 5.* Always following Examples 1,2,3 and 4, we obtain  $bc$  local specificity w.r.t. triangle  $\{6, 7, 8\}$ ,  $s_F(bc, \{6, 7, 8\}) = 3 \div 3 = 1$ , i.e. pattern  $bc$  is specific of the local support set  $\{1, 2, 3\}$ . Furthermore, implication  $\square_A^{\{6,7,8\}} b \rightarrow \square_A^{\{6,7,8\}} bc$  has informativity  $6 \div 3 = 2$ , i.e. in the abstract extensional space  $A \square_A b \rightarrow \square_A bc$  has confidence 0.5 while the implication holds at the local level.

### 4.3 Indirect Local Knowledge and Associated Measures

Local knowledge is related above to a notion of locality in a graph expressed through a confluence structure of the vertex space. This is mainly illustrated on the idea that the subgraph induced by the (abstract) support set of some pattern is made of several connected components, and that there may be specific patterns associated to each connected component. However, we are also interested in locality notions closer to the notion of community in Social Network Analysis.

A well known example of community definition is the  $k$ -clique community [5] which is defined as a maximal vertex subset made of adjacent (i.e. sharing  $k - 1$  vertices)  $k$ -cliques. Such a  $k$ -clique community may alternatively be defined as a connected component of a graph whose vertices are  $k$ -cliques and edges relate two adjacent  $k$ -cliques. What we discuss, more generally, in this section is a way to define local knowledge associated to subgraphs which are connected components of a *derived graph* made of particular vertex subsets, as  $k$ -cliques in the  $k$ -clique community case. This local knowledge, stated as *indirect*, is obtained by using the methodology described in Sect. 4 on the derived graph.

We start from a family  $T$  of elements of  $2^O$ , and consider  $T$  as the vertex set of a new graph  $G_T = (T, E_T)$ . We consider then a confluence  $F$  of  $2^T$  as the extensional space and search for the corresponding local closed patterns. The corresponding local support sets are afterwards transformed into support sets in  $2^O$ : when considering a (local support set, local closed pattern) pair  $(e_T, l)$  we may transform it into the pair  $(e, l)$  where  $e$  is the union of the elements of  $e_T$ . Let  $T \subseteq 2^O$ , and  $u : 2^T \rightarrow 2^O$  be such that  $u(e_T) = \cup_{t \in e_T} t$ .  $u(e_T)$  is called the *flattening* of  $e_T$ . We consider then two maps  $\text{ext}_T$  and  $\text{int}_T$  relating  $L$  to  $2^T$ :

- $\text{ext}_T : L \rightarrow 2^T$  with  $\text{ext}_T(q) = \{t | t \subseteq \text{ext}(q)\}$
- $\text{int}_T : 2^T \rightarrow L$  with  $\text{int}_T(e_T) = \text{int} \circ u(e_T)$

$\text{ext}_T(q)$  represents the support set of  $q$  in  $2^T$  when considering that  $q$  occurs in  $t$  whenever  $q$  occurs in all elements of  $t$ . Conversely  $\text{int}_T(e_T)$  represents the greatest pattern in  $L$  whose support set in  $T$  includes  $e_T$ , i.e. whose support set in  $O$  contains, as subsets, the elements of  $e_T$ . We have then the following result when flattening the (local ) support sets so found in  $F$ :

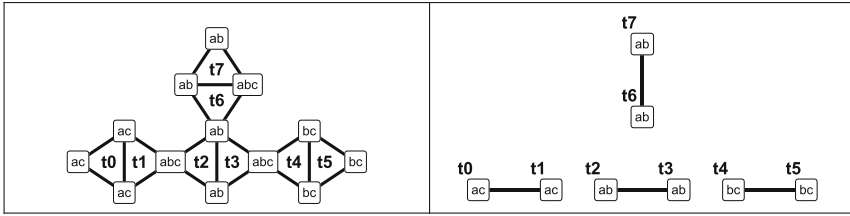
**Proposition 1.** *Let  $F$  be a confluence of  $2^T$ ,  $u$  be the flattening operator on  $O$  and  $(e_T, l)$  be a (local support set, local closed pattern) pair with  $e_T \geq m \in \min[F]$ , then  $u(e_T)$  is the greatest element of  $u[F^m]$  among elements  $e$  such that  $\text{int}(e) = l$ .*

This means that the support closed patterns with respect to the confluence  $F$  are the same as the support closed patterns with respect to the extensional space  $U = u[F]$ . Note that as flattened support sets are obtained by joining elements of  $T$ , they belong to the abstraction  $A = \text{UnionClosure}(T)$ .<sup>1</sup>

This will be illustrated by considering  $T$  as the set of 3-cliques of  $G$  (further called *triangles*) and stating that  $(t_1, t_2)$  belongs to  $G_T$  whenever  $t_1$  and  $t_2$  share an edge in  $G$ . In this case, a flattened local support set of pattern  $q$  represents a triangle community in the pattern  $q$  subgraph  $G(\text{ext}(q))$ . An example of both graphs  $G$  and  $G_T$  is displayed Fig. 2.

It is then natural to extend the definition of specificity to make it relative to the flattened support sets:

<sup>1</sup> defined by (i)  $T \subseteq \text{UnionClosure}(T)$  and (ii) if  $q$  and  $w$  belong to  $\text{UnionClosure}(T)$  then  $q \cup w$  belongs to  $\text{UnionClosure}(T)$ . By considering any subset  $T \subseteq 2^O$  and closing it under union we obtain an abstraction of  $2^O$  [10].



**Fig. 2.** On the left we have a graph of objects each described as an itemset included in  $\{a, b, c\}$ . This graph represents the triangle abstraction of some input graph. On the right, the graph  $G_T$  whose vertices are the triangles of  $G$ . The itemset describing a vertex in  $G_T$  is the intersection of the itemsets describing the elements of the corresponding triangle in  $G$ .

**Definition 5.** Let  $q$  be a pattern,  $F$  an extensional confluence of  $2^T$  where  $T \subseteq 2^O$ ,  $A$  is the abstraction generated from  $T$  and  $m \in F$  such that  $m \subseteq \text{ext}_T(q)$ , the flattened specificity of  $q$  in the vicinity of  $m$  is defined as:

$$s_F^f(q, m) = \frac{|u \circ p_m \circ \text{ext}_T(q)|}{|u \circ \text{ext}_T(q)|} = \frac{|u \circ p_m \circ \text{ext}_T(q)|}{|\text{ext}_A(q)|}$$

Coming back to the example of triangles communities,  $s_F(q, m)$  states to what extent a pattern  $q$  is specific to the community containing a particular triangle  $m$  with respect to its abstract support set in  $O$  when considering only triangles.

From Sect. 4.1 we know that we may rewrite  $p_m \circ \text{ext}_T(q) \subseteq p_m \circ \text{ext}_T(w)$  as a local implication  $\Box_m q \rightarrow \Box_m w$ . As the flattening operator is monotonic when the rule  $\Box_m q \rightarrow \Box_m w$  is valid on the set  $T$ , we also have  $u \circ p_m \circ \text{ext}_T(q) \subseteq u \circ p_m \circ \text{ext}_T(w)$ . We may then define the flattened informativity of  $r = \Box_m q \rightarrow \Box_m w$  as

$$I_F^f(r) = \frac{|u \circ \text{ext}_T(q)|}{|u \circ \text{ext}_T(qw)|} = \frac{|\text{ext}_A(q)|}{|\text{ext}_A(qw)|}$$

Let us consider a (flattened local support set, local closed pattern) pair  $(e, l)$ , where  $e$  is a community containing a given triangle  $m$ ,  $l$  the corresponding local closed pattern, and  $c$  an abstract closed pattern whose support set in  $G$  induces a subgraph in which  $e$  forms a triangle community. This means that  $\Box_m c \rightarrow \Box_m l$  is a valid local implication rule stating that when we consider the subgraph induced by the support set of  $c$ , all the members of the community containing the triangle  $m$  also has pattern  $l$  (see Fig. 3). The set of such  $\Box_m c \rightarrow \Box_m l$  local implications, with  $c \neq l$ , represents (a basis for) the local knowledge deriving from the reduction of the extensional space to triangle communities.

*Example 6.* Let  $G = (O, E)$  be the graph displayed on the left part of Fig. 2. Each vertex of  $G$  belongs to some triangle in  $G$ , therefore  $G$  is the same as its triangle abstraction. Each vertex has an itemset included in  $\{a, b, c\}$  as a label. The set of triangles is  $T = \{t_0, t_1, t_2, t_3, t_4, t_5, t_6, t_7\}$  and forms a triangle graph  $G_T$  displayed on the right part of Fig. 2. An edge relates any pair



of triangles sharing two vertices in  $G$ , as for instance  $(t_0, t_1)$ . Each triangle in  $G_T$  has as its itemset the intersection of the itemsets of its three vertices in  $G$ . For instance, the description of  $t_1$  in  $G_T$  is  $ac = abc \cap ac \cap ac$ . The vertex subsets inducing connected subgraphs of  $G^T$  form the *cc-confluence*  $F^T = \{\{t_0\}, \{t_1\}, \{t_0, t_1\}, \{t_2\}, \{t_3\}, \{t_2, t_3\}, \{t_4\}, \{t_5\}, \{t_4, t_5\}, \{t_6\}, \{t_7\}, \{t_6, t_7\}\}$ .

The support set of the pattern  $a$  is  $\text{ext}(a) = \{t_0, t_1, t_2, t_3, t_6, t_7\}$ . The local support with respect to  $t_0$  is  $p_{t_0}(\{t_0, t_1, t_2, t_3, t_6, t_7\}) = \{t_0, t_1\}$ , i.e. the connected component containing  $\{t_0\}$  of the subgraph induced by  $\text{ext}(a)$ . The local closed patterns, where  $f_i(q)$  denotes a closed pattern which is local w.r.t. triangle  $t_i$ , are as follows:

$$- f_0(a) = f_1(a) = ac, f_2(a) = f_3(a) = ab, f_6(a) = f_7(a) = ab$$

In the same way, the pattern  $b$  whose support set is  $\text{ext}(b) = \{t_2, t_3, t_4, t_5, t_6, t_7\}$ . leads to the following local closed patterns:

$$- f_2(b) = f_3(b) = ab, f_4(b) = f_5(b) = bc, f_6(b) = f_7(b) = ab$$

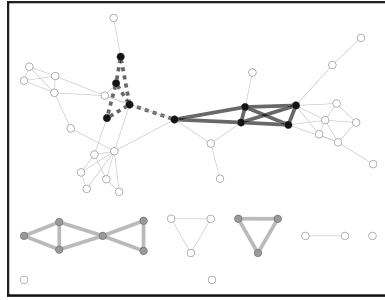
Note that  $ab$  appears both as a local closed pattern resulting from  $a$  with respect to  $f_0, f_1$  and to  $f_6, f_7$  and as a local closed pattern resulting from  $b$  with respect to  $f_2, f_3$  and again to  $f_6, f_7$ . This leads to three different sets of local implications:

$$\begin{aligned} & - \square_{t_2} a \rightarrow \square_{t_2} ab, \square_{t_3} a \rightarrow \square_{t_3} ab, \square_{t_6} a \rightarrow \square_{t_6} ab, \square_{t_7} a \rightarrow \square_{t_7} ab, \\ & - \square_{t_2} b \rightarrow \square_{t_2} ab, \square_{t_3} b \rightarrow \square_{t_3} ab, \square_{t_6} b \rightarrow \square_{t_6} ab, \square_{t_7} b \rightarrow \square_{t_7} ab, \end{aligned}$$

As a whole, a local closed pattern is part of a pair  $(e_T, l)$  where  $l$  is the local closed pattern and  $e_T$  is a local support set corresponding to one of the connected components induced by the support set. Two examples of such pairs are  $(\{t_2, t_3\}, ab)$  and  $(\{t_6, t_7\}, ab)$ . When interested in implication rules, we have to consider triples  $(c, t_i, l)$  where  $c$  is a pattern whose support set is split in different local support sets one of which, namely  $e$ , contains  $t_i$ .  $\square$

*Example 7.* The dataset is denoted as *s50-1* and is a standard attributed graph dataset.<sup>2</sup> It represents 148 friendship relations between 50 pupils of a school in the West of Scotland, and labels concern the substance use (tobacco, cannabis and alcohol) and sporting activity (see [9]). We want to answer to the question: "what knowledge can be extracted when considering groups of pupils connected by friendship relationships?". For that purpose, we computed the local abstract closures associated to the *cc-confluence* representing 3-clique communities in subgraphs of the triangle graph  $G_T$  derived from the original graph and the "support $\geq 4$ " constraint on  $O$ . In Fig. 3 we represent the flattened local support set  $e$  of the local closed pattern  $l$  shared in a community (in black lines and dots) of the subgraph induced by the abstract support set of  $l$  (in black+grey lines and dots) (w.r.t. the 3-clique abstraction). We also represent (in dashed+ black + grey lines and dots) the abstract support set of the abstract

<sup>2</sup> [http://www.stats.ox.ac.uk/~snijders/siena/s50\\_data.htm](http://www.stats.ox.ac.uk/~snijders/siena/s50_data.htm).



**Fig. 3.** Representation of a local rule  $\square_m c \rightarrow \square_m l$  extracted from a West of Scotland school friendship network.  $c$  is the abstract closed pattern shared by the friendship triangles displayed in black+ grey+ dashed lines and dots.  $l$  is a local closed pattern specific to the 3-community, to which the triangle  $m$  belongs, represented in black lines and dots. This local closed pattern is shared by all the friendship triangles displayed in black+grey lines and dots.

closed pattern  $c$  that also induces a subgraph in which  $e$  is a connected component. Overall  $\square_m c \rightarrow \square_m l$  is a valid local implication rule whose informativity is  $I_F^f(r) = \frac{|\text{ext}_A(c)|}{|\text{ext}_A(l)|} = (5 + 9 + 4) \div (5 + 9) = 1.286$ . The specificity  $s_F^f(r)$  of the local closed pattern  $l$  is  $5 \div (5 + 9) = 0.357$ . Here  $l$  means “Never has tried Cannabis, drinks moderately, does not smoke” while  $c$  means “Has tried Cannabis at most once, drinks moderately, does not smoke”. The specificity of the 3-community, with respect to the whole set of pupils sharing “Have tried Cannabis at most once, drink moderately, does not smoke” is to be composed only of pupils who have never tried Cannabis.

## 5 Conclusion

We have discussed here a framework extending the closed itemset mining framework to abstract and local information in an attributed network. Our focus in this article was on the abstract and local knowledge to be extracted as abstract and local rules, together with measures about how specific and informative is abstraction or locality.

## References

1. Prado, A.B., Plantevit, M., Robardet, C., Boulicaut, J.F.: Mining graph topological patterns: finding co-variations among vertex descriptors. *IEEE Trans. Knowl. Data Eng.* **25**(9), 2090–2104 (2013)
2. Caspard, N., Monjardet, B.: The lattices of closure systems, closure operators, and implicational systems on a finite set: a survey. *Discrete Appl. Math.* **127**(2), 241–269 (2003)
3. Ganter, B., Wille, R.: *Formal Concept Analysis: Mathematical Foundations*. Springer, Heidelberg (1999)

4. Mougel, P.-N., Rigotti, C., Gandrillon, O.: Finding collections of  $k$ -clique percolated components in attributed graphs. In: Tan, P.-N., Chawla, S., Ho, C.K., Bailey, J. (eds.) PAKDD 2012, Part II. LNCS, vol. 7302, pp. 181–192. Springer, Heidelberg (2012)
5. Palla, G., Derenyi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**(7043), 814–818 (2005)
6. Pasquier, N., Taouil, R., Bastide, Y., Stumme, G., Lakhal, L.: Generating a condensed representation for association rules. *J. Intell. Inf. Syst. (JIIS)* **24**(1), 29–60 (2005)
7. Silva, A., Meira Jr., W., Zaki, M.J.: Mining attribute-structure correlated patterns in large attributed graphs. *Proc. VLDB Endow.* **5**(5), 466–477 (2012)
8. Soldano, H.: Extensional confluences and local closure operators. In: Baixeries, J., Sacarea, C., Ojeda-Aciego, M. (eds.) ICFCA 2015. LNCS, vol. 9113, pp. 128–144. Springer, Heidelberg (2015)
9. Soldano, H., Santini, G.: Graph abstraction for closed pattern mining in attributed network. In: Schaub, T., Friedrich, G., O’Sullivan, B. (eds.) European Conference in Artificial Intelligence (ECAI). *Frontiers in Artificial Intelligence and Applications*, vol. 263, pp. 849–854. IOS Press (2014)
10. Soldano, H., Ventos, V.: Abstract Concept Lattices. In: Jäschke, R. (ed.) ICFCA 2011. LNCS, vol. 6628, pp. 235–250. Springer, Heidelberg (2011)