

Bi-Pattern Mining of Two Mode and Directed Networks

Henry Soldano*
LIPN UMR-CNRS 7030, Université
Paris-Nord
Villetaneuse, France
ISYEB UMR 7205, Museum National
d'Histoire Naturelle
Paris, France
henry.soldano@mnhn.fr

Guillaume Santini
LIPN, Université Paris-Nord
Villetaneuse, France
guillaume.santini@lipn.univ-paris13.
fr

Dominique Bouthinon
LIPN, Université Paris-Nord
Villetaneuse, France
dominique.bouthinon@lipn.
univ-paris13.fr

Sophie Bary
ISYEB, Museum National d'Histoire
Naturelle
Paris, France
sophie.bary@mnhn.fr

Emmanuel Lazega
Institut d'Etudes Politiques de Paris,
CSO-CNRS
Paris, France
emmanuel.lazega@sciencespo.fr

ABSTRACT

In two-mode networks there are two kinds of vertices, i.e objects, each being possibly described with a proper attribute set. This means that to select a subnetwork according to vertex descriptions we have to consider a pair of vertex subsets. A common technique is to extract from a network an essential subnetwork, the core subgraph of the network. Formal Concept Analysis and closed pattern mining were previously applied to networks with the purpose of reducing extensions of patterns to be core subgraphs. To apply this methodology to two-mode networks, we need to consider the two vertex subsets of two-mode cores and define accordingly abstract closed bi-patterns. Each component of a bi-pattern is then associated to one mode. We also show that the same methodology applies to hub-authority cores of directed networks in which each vertex subset is associated to a role (in or out). We illustrate the methodology both on a two-mode network of epistemological data and on a directed advice network of lawyers.

CCS CONCEPTS

• Information systems → Data mining; • Human-centered computing → Social networks;

KEYWORDS

Closed Pattern Mining; Core Subnetwork; Attributed Network; Two-Mode Network; Directed Network

ACM Reference Format:

Henry Soldano, Guillaume Santini, Dominique Bouthinon, Sophie Bary, and Emmanuel Lazega. 2018. Bi-Pattern Mining of Two Mode and Directed Networks. In *WWW '18 Companion: The 2018 Web Conference Companion*,

*Contact Author

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23-27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3191568>

April 23-27, 2018, Lyon, France. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3184558.3191568>

1 INTRODUCTION

The aim of this article is to extend the Closed Pattern mining (CPM) and Formal Concept Analysis (FCA) methodologies in order to investigate attributed two-mode networks as well as directed networks according to the *in* and *out* roles of their vertices. This follows previous work in which these methodologies were applied to undirected and directed graphs. Note that there is no difference between the two methodologies in that they enumerate the same closed patterns, however FCA is also interested in the structure of this result as a conceptual structure.

Most of the work in social and complex networks analysis consider unlabelled and undirected networks and is concerned by what may be said about the topological structure of the network. Various ways have been proposed to extract interesting subgraphs. In particular in the core-periphery model the network is made of a *core subgraph*, i.e. a dense subgraph whose vertices are highly connected, together with its *periphery*, made of vertices highly connected to the core, but poorly connected between them [6]. The first formal core definition was the *k*-core subgraph which is the greatest subnetwork whose vertices all have degree at least *k* in the subnetwork [16]. By changing the topological property we obtain various core definitions [2]. Core definitions have also been considered in directed networks. In particular a core definition has been proposed in [10] to investigate collaboration within directed networks. The requirement is then that both *indegrees* and *outdegrees* of vertices have to be higher than thresholds. A different kind of core is related to the *Hub-Authority* idea which considers that a vertex may be prominent in a network according to only one or both of its *out* or *in* roles [11]. The HA-core has been recently defined in order to express this idea [20].

Various recent works on complex networks analysis take into account information provided as labels about vertices or edges. The network is then called a *labelled* or *attributed* network. This leads to mine and rank patterns as pairs of constraints on topology and labels [8, 14, 17]. Recently a new approach has been presented

extending CPM and FCA to attributed graphs by reducing the vertex subset in which an attribute pattern occur, using *interior operators* [18]. Such an operator reduces then a vertex subset to its core subset. In particular the HA-core mentioned above was defined to investigate this way an advice network in a lawyer firm [20].

Two-mode networks are bipartite graphs made of two vertex sets representing in general two kind of entities, for instance actors and movies, together with edge relating entities of each kind, as for instance "G. Clooney acted in Ocean's Eleven". Until recently they were mostly investigated by extracting single mode networks, relating for instance actors to actors who participated to the same movies. However in [5] the authors advocated the direct investigation of two-mode networks, and core definitions for two-mode networks have been recently proposed by Cerinsek and Batagelj [7]. When such a network is attributed each kind of vertex is described according to a proper attribute set. This means that we need patterns made of two attribute subsets, we call *bi-patterns*, that select two vertex subsets. In this article we investigate the abstract *bi-concept* lattices resulting from the adaptation of the abstract FCA methodology to attributed two-mode networks. This allows for instance to require actors to be American and movies to be recent, but only consider vertices of a subnetwork in which each actor played in at least 2 movies and each movie is linked to at least 3 actors. Such bi-patterns may also be defined in the directed case when considering subgraphs in which a pattern is associated to each of the in or out vertex roles.

2 PRELIMINARIES

2.1 Abstract Closed Pattern Mining and Concept lattices

The closed pattern mining and Formal Concept Analysis (FCA) frameworks consider the occurrences of patterns in a set of objects V . The pattern language L is partially ordered in such a way that if $q' \geq q$, i.e. q' is *more specific* than q , then whenever q' occurs in object v , q also occurs in v . The set of occurrences $\text{ext}(q)$ of a pattern q , i.e. the object subset in which q occurs, is called its *extension* in V . The purpose common to closed pattern mining and FCA is then to represent, in a condensed way, the set of *definable* subsets of V , i.e. subsets which are pattern extensions.

Enumerating the definable subsets of V comes down to enumerate the classes of the relation stating as equivalent two patterns with same extension. The more favorable situation is the one in which there is a unique most specific pattern in each class. This happens whenever the pattern language is a complete lattice. In original FCA the pattern language is the powerset of a set of binary attributes, however FCA has more recently been extended to more general languages called pattern structures [9]. The most specific pattern of a class may then be accessed from any pattern q of the same class through a closure operator. We define hereunder closure operators together with their dual *interior operators*.

Definition 2.1. Let S be an ordered set and $f : S \rightarrow S$ a self map such that for any $x, y \in S$, 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 $f(x) \geq x$, f is called a *closure operator* while if $f(x) \leq x$, f is called an *interior operator*.

Formal Concept Analysis goes beyond enumeration of closed patterns: FCA considers knowledge discovery as the process of discovering the ordering structure of the data to analyse. It relies primarily on the Galois connection¹ between the pattern language and the extensional space, i.e. the powerset of objects:

PROPOSITION 2.2. Let (L, \leq) be a lattice called the *pattern language*, V be a set of objects and $d : V \rightarrow L$ be an operator that describes the object x as an element $d(x)$ of L . Let $\text{ext}(q) = \{x \in V \mid q \leq d(x)\}$ define the subset of objects in which pattern q occurs. Then

- $\text{int}(V') = \bigwedge_{x \in V'} d(x)$ is the greatest element of L which occurs in V'
- (int, ext) define a Galois connection on $(2^V, L)$

Concept lattices are consequently defined as the ordering of concepts, i.e. maximal pairs of elements from 2^V and L . In what follows we will use interior operators to define the more general framework of abstract concept lattices. Interior operators were introduced in FCA to reduce the concept lattice size, either by reducing the pattern language [9] or the extensional space [15]. First we recall a general result [15, 21] together with a corollary defining *abstract concept lattices*:

PROPOSITION 2.3. Let X and L be two lattices, (int, ext) be a Galois connection on (X, L) and p be an interior operator on X . Let $A = p[X]$, then $(\text{int}, p \circ \text{ext})$ is a Galois connection on (A, L) .

COROLLARY 2.4.

- $f = \text{int} \circ p \circ \text{ext}$ is a closure operator on L
- $h = p \circ \text{ext} \circ \text{int}$ is a closure operator on A
- $h[A]$ and $f[L]$ are two anti-isomorphic lattices and the lattice of the (e, c) pairs where $c = \text{int}(e)$ and $e = p \circ \text{ext}(c)$ form a lattice, ordered following A , isomorphic to $h[A]$.

In such a concept (e, c) , e is called its *(abstract) extent* while c is its *intent* i.e. the abstract closed pattern whose abstract extension $p \circ \text{ext}(c)$ is e .

As the new equivalence relation is coarser, i.e. $\text{ext}(q) = \text{ext}(q')$ implies $p \circ \text{ext}(q) = p \circ \text{ext}(q')$, there is less abstract closed patterns than closed patterns, thus leading to a more condensed representation of patterns occurring on V . Note the useful one to one correspondence between interiors operators on a lattice and their range (see [3] for the dual result on closure operators):

PROPOSITION 2.5. Let X be a complete lattice. A subset A of X is the range of an interior operator on X if and only if A is closed under join. The interior operator $f : X \rightarrow X$ is then unique and defined as $f(x) = \bigvee_{\{a \in A \mid a \leq x\}} a$.

A is closed under join means here that the join of any subset, including the empty subset \emptyset , belongs to A . We further call A an *abstraction* of X , hence we may define abstract concept lattices through interior operators as well as abstractions.

Example 2.6. We exemplify hereunder the closure operator $f = \text{int} \circ \text{ext}$ returning closed patterns in the standard closed itemset mining case. We further write subsets as strings, i.e. 12 stands for $\{1, 2\}$. Patterns are subsets of $I = \{a, b, c, d\}$, objects in $V = \{1, 2, 3\}$ are described as $d[V] = \{a, ab, abc\}$. We have then for instance

¹Galois connections are defined in Appendix A

$\text{ext}(b) = 23$ and as a consequence, $f(b) = d(2) \cap d(3) = ab \cap abc = ab$, $f(abc) = d(3) = abc$ and $f(d) = abcd$. The latter closure means that d is in the set of patterns with empty support set whose greatest element is $abcd$.

We consider the operator p on 2^V such that $p(e) = e$ except for singletons that are projected on the empty set: $p(\{1\}) = p(\{2\}) = p(\{3\}) = \emptyset$. We have then $p \circ \text{ext}(ab) = p(\{2, 3\}) = \{2, 3\}$ and as a consequence $f(ab) = abc$ as in the non-abstract case. However $p \circ \text{ext}(abc) = p(\{3\}) = \emptyset$ and now $f(abc) = abcd$ as $abcd$ is the greatest element with empty abstract support set.

The corresponding abstraction $A = p[2^{123}]$ is generated by union closure of size 2 subsets $\{12, 23, 13\}$ and it is then straightforward that for any e , $p[e]$ is the greatest subset of A smaller than or equal to e . For instance, $p[12] = 12$ as 12 belongs to A while $p[1] = \emptyset$ as no element of A except \emptyset is included in subset 1.

2.2 Cores and Closed Pattern Mining of Attributed Networks

Now, consider the object set as the vertex set V of some graph whose vertices are each labelled by a description in a pattern language. Defining the essential part of a graph, i.e. its *core subgraph*, relies on all vertices satisfying some boolean property. Let $G = (V, E)$ be a graph. Vertex properties are defined as $P : V \times 2^V \rightarrow \{\text{true}, \text{false}\}$ mappings where $P(v, X)$ is true whenever vertex x satisfies some condition within the subgraph G_X induced by the vertex subset X . The *core subgraph* of a graph (V, E) is then defined as the subgraph $G_{V'}$ induced by the largest vertex subset V' , also called its *core*, whose vertices v all have property $P(v, V')$.

To define a core, we need P to be such that there does exist such a largest vertex subset with property P . This is true whenever P is *monotone* i.e. for any $x \in X_1 \subseteq X$ we have that $P(x, X_1)$ and $X_2 \supseteq X_1$ implies $P(x, X_2)$ [2, 18]. The following result allows then to apply abstract FCA to graphs:

PROPOSITION 2.7. *The operator that reduces a vertex subset V' of a graph G to the core of the subgraph $G_{V'}$ is an interior operator on 2^V .*

As a result, abstract concept lattices together with closure operators are defined in such a way that each *extent* $p \circ \text{ext}(c)$ is a core while the associated *intent* c is the most specific pattern that occurs in this core. Abstract closed pattern mining has been applied to undirected networks [19] as well as directed networks [20]. The next section is first motivated by applying FCA to two-mode networks, i.e. networks in which an edge relates a vertex from a vertex set V_1 to a vertex from a vertex set V_2 . The vertices may then be described in two different pattern languages L_1 and L_2 .

3 ABSTRACT FORMAL BI-CONCEPT ANALYSIS

We consider lattice products both regarding extensions and pattern languages and will obtain a new Galois connection whose associated lattice we call a bi-concept lattice.

3.1 Bi-concept lattices

Lattice products are also lattices according to the so-called cartesian ordering:

PROPOSITION 3.1. *Let $(X_1, \leq_1, \vee_1, \wedge_1)$ and $(X_2, \leq_2, \vee_2, \wedge_2)$ be two lattices, and consider the cartesian product $X = X_1 \times X_2$ together with the binary relation \leq defined as $(x_1, x_2) \leq (y_1, y_2)$ iff $x_1 \leq_1 y_1$ and $x_2 \leq_2 y_2$. Then (X, \leq, \vee, \wedge) is a lattice with join and meet defined as:*

- $(x_1, x_2) \vee (y_1, y_2) = (x_1 \vee_1 y_1, x_2 \vee_2 y_2)$
- $(x_1, x_2) \wedge (y_1, y_2) = (x_1 \wedge_1 y_1, x_2 \wedge_2 y_2)$

We may then build a Galois connection on lattices products (see proof in Appendix A):

PROPOSITION 3.2. *Let $X = X_1 \times X_2$ and $L = L_1 \times L_2$ be two lattices product, and let $(\text{int}_1, \text{ext}_1)$ and $(\text{int}_2, \text{ext}_2)$ be Galois connections on respective lattices pairs (X_1, L_1) and (X_2, L_2) . Consider the mappings int and ext on X and L such that:*

- $\text{int}(x_1, x_2) = (\text{int}_1(x_1), \text{int}_2(x_2))$
- $\text{ext}(l_1, l_2) = (\text{ext}_1(l_1), \text{ext}_2(l_2))$

then (int, ext) define a Galois connection on (X, L)

In what follows we consider two Galois connections as defined in Proposition 2.2. We have then $X = (2^{V_1}, 2^{V_2})$:

Definition 3.3. The lattice of the Galois connection on lattice products (X, L) is called a bi-concept lattice, and each (x, l) pair such that $l = \text{int}(x)$ and $x = \text{ext}(l)$ is called a bi-concept with intent l and extent x .

Abstract FCA uses an interior operator to obtain smaller concept lattices, and we will allow the interior operator to create the dependency between the two components of the extent which is necessary to represent cores of two-mode networks.

3.2 Abstract bi-concept lattices

Proposition 2.3 states that applying an interior operator to a lattice involved in a Galois connection preserves the connection. The interior operator in the bi-concept case applies to a pair of object subsets, i.e. has domain $X = 2^{V_1} \times 2^{V_2}$:

Definition 3.4. Let p be an interior operator on X and (int, ext) be the Galois connection on (X, L) associated to a bi-concept lattice. Then the lattice of the Galois connection $(\text{int}, p \circ \text{ext})$ on $(p[X], L)$ is called the abstract bi-concept lattice associated to p .

We exemplify hereunder these definitions. According to Proposition 2.5 we may derive an interior operator p from a subset of X closed under join. We will hereunder make use of this property to represent the interior operator.

Example 3.5. Let $V_1 = \{1, 2\}$ and $V_2 = \{3, 4\}$ be two object sets and $X_1 = 2^{V_1} = \{\emptyset, 1, 2, 12\}$ while $X_2 = 2^{V_2} = \{\emptyset, 3, 4, 34\}$. Objects of V_1 are labelled by subsets of $I_1 = \{a, b, c\}$ while objects of V_2 are labelled by subsets of $I_2 = \{w, x\}$. The descriptions of the objects from V_1 and V_2 respectively as subsets of I_1 and I_2 are as follows:

- $d_1(1) = ab, d_1(2) = b, d_2(3) = wx, d_2(4) = x$

Consider the abstraction $\{(\emptyset, \emptyset), (1, 4), (2, 3), (12, 34)\}$ and the associated interior operator p . Now, we have that

- $p(12, 34) = (12, 34)$,
- $p(1, 34) = (1, 4)$ and $\text{int}(1, 4) = (ab, x)$
- $p(12, 3) = (2, 3)$ and $\text{int}(2, 3) = (b, wx)$
- $p(1, 3) = p(\emptyset, 3) = (\emptyset, \emptyset)$ and $\text{int}(\emptyset, \emptyset) = (abc, wx)$

We obtain then the following abstract bi-concept lattice:

$$\begin{aligned}
& ((12,34) (b,x)) \\
((1,4) (ab,x)) & \quad ((2,3),(b,wx)) \\
& ((\emptyset, \emptyset), (abc,wx))
\end{aligned}$$

The set of abstract closed-bi-patterns with extent different from (\emptyset, \emptyset) is then $\{(b, x), (ab, x), (b, wx)\}$.

4 CORES ASSOCIATED TO VERTEX SUBSET PAIRS

In what follows we consider the subnetwork induced by a pair of vertex subsets (W_1, W_2) . When considering $W_1 \subseteq V_1$ and $W_2 \subseteq V_2$ we simply write $(W_1, W_2) \leq (V_1, V_2)$ or call (W_1, W_2) a subset pair of (V_1, V_2) . The following definition may be applied to a two-mode network (V_1, V_2, E) as well as to a single mode network by considering $V = V_1 = V_2$.

Definition 4.1. Let $G = (V_1, V_2, E)$ be a network, the subnetwork induced by the subset pair (W_1, W_2) is the network $G_{(W_1, W_2)} = (W_1, W_2, E')$ where E' is the edge subset relating vertices from W_1 to vertices from W_2 .

We need to define monotone properties in this context:

Definition 4.2. $P_1 : V_1 \times 2^{V_1} \times 2^{V_2} \rightarrow \{true, false\}$ is said monotone if and only if for any $w \in V_1$ and any subset pairs (W_1, W_2) and $(W'_1, W'_2) \geq (W_1, W_2)$,

$$P_1(w, W_1, W_2) \text{ implies } P_1(w, W'_1, W'_2)$$

In the same way P_2 on $V_2 \times 2^{V_1} \times 2^{V_2}$ is monotone whenever for any $w \in W_2$, $P_2(w, W_1, W_2)$ implies $P_2(w, W'_1, W'_2)$

Cores will then be defined thanks to the following result (see proof in Appendix A):

PROPOSITION 4.3. Let (P_1, P_2) be a pair of monotone properties, and (W_1, W_2) be a subset pair of (V_1, V_2) . Then there exists a greatest subset pair $(S_1, S_2) \leq (W_1, W_2)$ such that $P_1(v_1, S_1, S_2)$ holds for all elements v_1 of S_1 and $P_2(v_2, S_1, S_2)$ holds for all elements v_2 of S_2 .

We will further call this subset pair (S_1, S_2) the *core subset pair* of (W_1, W_2) and define core subgraphs accordingly:

Definition 4.4. Let $G = (V_1, V_2, E)$ be a network, and (P_1, P_2) be a pair of monotone properties. The subnetwork $G_{(S_1, S_2)}$ induced by the core subset pair (S_1, S_2) is called the *core subnetwork* of G .

We benefit then from a result similar to Proposition 2.7:

PROPOSITION 4.5. The operator that reduces a subset pair $(W_1, W_2) \leq (V_1, V_2)$ to its core subset pair (S_1, S_2) is an interior operator on $2^{V_1} \times 2^{V_2}$.

We first define the *h-a* BHA-core of a two-mode network:

Definition 4.6. The *h-a* BHA-core of the network G is defined through the following pair of core properties:

- $P_1(v, X_1, X_2)$ holds if and only if the degree of $v_1 \in X_1$ in $G_{(X_1, X_2)}$ is at least h .
- $P_2(v, X_1, X_2)$ holds if and only if the degree of $v_2 \in X_2$ in $G_{(X_1, X_2)}$ is at least a .

P_1 and P_2 are clearly monotone and therefore the *h-a* HA-core is properly defined. This core definition is equivalent when considering two-mode networks to the definition presented by Cerinsek and Batagelj [7] in which the *p-q* BHA-core is called the (p, q) -core, and more generally any such generalized two-mode core may be rewritten according to a pair of core properties. We may also consider a single mode network $G(V, E)$ as a two-mode network $G(V, V, E)$ and consider subgraph induced by a pair of vertex subsets according to Definition 4.1. We may then consider core definitions of undirected and directed networks in which vertices may have two roles. In the directed case we relate hereunder this HBA-core to the hub-authority core defined in [20]:

PROPOSITION 4.7. Let $G = (V, E)$ be a directed network rewritten as the two-mode network $G_T = (V, V, E)$. Let then p_{BHA} be the interior operator associated to the *h-a* BHA-cores of G_T and consider $p_{HBA}(W, W) = (S_1, S_2)$, then $S_1 \cup S_2$ is the *h-a* HA-core of the subgraph G_W induced by W , S_1 is its hub vertex subset while S_2 is its authority vertex subset.

This means that we have generalized hub-authority cores by extracting them from subgraphs induced by pairs of vertex subsets W_1, W_2 in such a way that hubs belong to W_1 while authorities belong to W_2 . This will be illustrated in Section 6.2. We give hereunder an example with a two-mode network.

Example 4.8. We consider the two-mode network pictured on the leftmost part of Figure 1. The two vertex sets are $V_1 = \{l_1, l_2, l_3\}$ and $V_2 = \{r_1, r_2, r_3\}$. Vertices of V_1 are labelled by subsets of $I_1 = \{a, b, c, d\}$ while vertices of V_2 are labelled by subsets of $I_2 = \{w, x, y, z\}$.

The most general bi-pattern (\emptyset, \emptyset) occurs in the whole network. Its 2-2 BHA-core is displayed in the middle of Figure 1 and is induced by $(l_1 l_2, r_1 r_2 r_3)$. We have then as the corresponding closed bi-pattern $\text{int}(l_1 l_2, r_1 r_2 r_3) = (ab, wx)$. When adding attributes to this bi-pattern we obtain subnetworks whose 2-2 HA-core is empty, except when adding y to wx . The corresponding bi-pattern (ab, wxy) occurs in $(l_1 l_2 l_3, r_1 r_3)$ whose corresponding 2-2 BHA-core is displayed in the rightmost part of Figure 1 and has vertex sets pair $(l_1 l_2, r_1 r_3)$. This bi-pattern is closed as nothing can be added without losing some vertex. Furthermore, adding anything to (ab, wxy) result in an empty 2-2 HA-core. The corresponding bi-concept lattice is therefore the total ordering of the 3 bi-concepts $((l_1 l_2, r_1 r_2 r_3), (ab, wx))$, $((l_1 l_2, r_1 r_3), (ab, wxy))$ and $((\emptyset, \emptyset), abcd, wxyz)$.

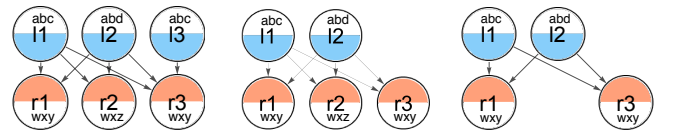


Figure 1: The two 2-2 BHA-cores in bi-concepts of Example 4.8. The leftmost part displays the whole network. In the middle we have its 2-2 BHA-core associated to the closed bi-pattern (ab, wx) . The rightmost part of the figure displays the 2-2 BHA-core associated to the other, more specific, bi-pattern (ab, wxy) .

5 COMPUTING THE INTERIOR OF (X_1, X_2) AND ENUMERATING ABSTRACT CLOSED BI-PATTERNS

5.1 Computing interiors

We present now a generic algorithm to compute the interior $p(X_1, X_2) = (S_1, S_2)$ associated to the pair of monotone properties (P_1, P_2) . It is basically a rewriting of the algorithm proposed in [7] when $V_1 \cap V_2 = \emptyset$ and also corresponds in the directed case, when considering $X_1 = X_2$, to the algorithm proposed in [20] to compute the HA-core. Let n be the number of vertices and m be the number of edges, the algorithm performs at most n iterations while the inner loop need $O(m)$ operations as far as p needs only to access the neighbourhood of each vertex. The overall complexity is then $O(m * n)$. A more efficient algorithm in $O(m * \max(\Delta, \log n))$, where Δ is the highest degree within the graph, is obtained by adapting the variant cited in [2] which uses two heaps as data structures for nodes of each mode.

Interior (X_1, X_2)

```

2   $S_1 \leftarrow X_1; S_2 \leftarrow X_2$ 
3  repeat
4     $Z_1 \leftarrow S_1; Z_2 \leftarrow S_2$ 
6    for  $x \in Z_1$  do
9      if  $\neg P_1(x, Z_1, Z_2)$  then withdraw  $x$  from  $S_1$  end
14   done
6    for  $x \in Z_2$  do
9      if  $\neg P_2(x, Z_1, Z_2)$  then withdraw  $x$  from  $S_2$  end
14   done
15 until  $S_1 = Z_1$  and  $S_2 = Z_2$ 
16 return  $S_1$  and  $S_2$ 
```

5.2 Bi-pattern enumeration

We focus now on abstract closed bi-pattern enumeration. Building the bi-concept lattice has therefore to be a post-processing step. The enumeration follows the same process as abstract closed pattern enumeration, i.e. the efficient divide and conquer scheme described in [4] as implemented in the program MinerLC². The adaptation is straightforward: the closure operator is now $f_A = \text{int} \circ p \circ \text{ext}$ where p is the interior operator as defined above. To perform enumeration of abstract closed bi-patterns we specialize each abstract closed bi-pattern (q_1, q_2) by adding either an element of I_1 to q_1 or an element of I_2 to q_2 .

The algorithm **bi-patterns** is described hereunder with the following notations: Let $q = (q_1, q_2)$ be a bi-pattern, i) $\text{add}(i, q)$ returns either $(q_1 \cup i, q_2)$ when $i \in I_1$ or $(q_1, q_2 \cup i)$ when $i \in I_2$, ii) $\text{minus}(I, q)$ returns the set of items which belong neither to the left part nor to the right part of the bi-pattern $q = (q_1, q_2)$, i.e. $\text{minus}(I, q) = I_1 \setminus q_1 \cup I_2 \setminus q_2$. iii) The exclusion pair list EL is a subset pair of (I_1, I_2) .

Algorithm bi-patterns (V)

```

 $S \leftarrow p(V)$ 
 $\text{enum}(\text{int}(S))$ 
```

Function $\text{enum}(q, S, EL)$

With: q an abstract closed bi-pattern, EL an exclusion list

Ensures: outputs the frequent abstract closed bi-patterns q'

where $q' \geq q$ and q' contains no items of EL

Output (q, S)

for all $x \in \text{minus}(I, q)$ **do** // specialize q

$S_x \leftarrow p(\text{add}(x, q))$

if $|S_x| \geq s$ **then**

$q_x \leftarrow \text{int}(S_x)$

if $q_x \cap EL = (\emptyset, \emptyset)$ **then** // q not yet enumerated

$\text{enum}(q_x, S_x, EL)$

$EL \leftarrow \text{add}(EL, x)$

end if

end if

end for

Example 5.1. We follow on from Example 4.8 and consider $s = 1$ as a minimal support. The algorithm starts by computing the 2-2 HA-core G_c of the whole graph G . G and G_c are displayed respectively on the left and on the middle of Figure 1. Function **enum** is then called with the core closed pattern $q = \text{int}(\text{vs}(G_c)) = \text{int}(\{l_1, l_2, r_1, r_2, r_3\}) = (l_a l_b, r_a r_b)$ and first outputs the pair $((l_a l_b, r_a r_b), \{l_1, l_2, r_1, r_2, r_3\})$, and then adds to q in turn each item in $\text{minus}(I, q) = \{l_c, l_d, r_c, r_d\}$:

- $\text{add}(l_c, q) = (l_a l_b l_c, r_a r_b)$ selects a subgraph whose core is empty. As a result the branch is pruned as smaller subgraphs would also result in an empty core.
- $\text{add}(l_d, q) = (l_a l_b l_d, r_a r_b)$ selects also a subgraph whose core is empty.
- $\text{add}(r_c, q) = (l_a l_b, r_a r_b r_c)$ selects $\{l_1, l_2, l_3, r_1, r_3\}$ whose core displayed on the right of Figure 1 has vertex set $\{l_1, l_2, r_1, r_3\}$. The core closed bi-pattern $q_x = (l_a l_b, r_a r_b r_c)$ is computed and having null intersection with the empty list EL leads to another recursive call of **enum**. This call will output the pair $(q_x, \{l_1, l_2, r_1, r_3\})$ but there will be no deeper recursive calls as 2-2 HA structure with strictly less than four nodes are excluded. We have then EL set to $\{r_c\}$ prior to the next iteration.
- $\text{add}(r_d, q) = (l_a l_b, r_a r_b r_d)$ selects a subgraph whose core is empty.

As **enum** ends **bi-patterns** also ends. The two closed bi-patterns that have been output are the most specific bi-patterns that occur respectively in the 2-2 B-cores, displayed on the middle and the right of Figure 1.

6 EXPERIMENTS

The implementation is an adaptation of the minerLC software. The first experiment concerns a new two-mode network while the second concerns a well-known directed social network available on the minerLC web page (see above).

6.1 A two-mode network of epistemological data

We are currently investigating a two-mode network concerning data related to a MNHN-IRD program (called MUSORSTOM then Tropical Deep-Sea Benthos) of expeditions to explore the deep-sea in the Indo-West Pacific region, since 1976 [1]. In this network 596 edges relate 74 campaigns to 268 participants to these campaigns.

²<https://lipn.univ-paris13.fr/MinerLC/>

Campaigns are described following their date and location, the type of fishing gear (dredge, trawl), the objectives of the campaign as well as species described during the campaign. Regarding participants the attributes concern the location of the institution they belong to, their scientific domain as well as bibliometrics. We have in particular searched abstract bi-concepts associated to a 3-4 HA cores (subnetworks with participants to at least 3 campaigns, and campaigns with at least 4 participants). As an illustration Figure 2 displays two bi-patterns whose intents mainly differ on their campaign component. The subnetworks, before core reduction, are made of respectively 132 and 135 vertices while core subnetworks represent extents of respectively 80 and 76 vertices. Vertices are displayed at their original position in the whole network according to a standard force directed drawing [12]. The difference between the extents are mainly in the left part of the network, which means that differences concern campaigns and participants which are strongly related within the original network.

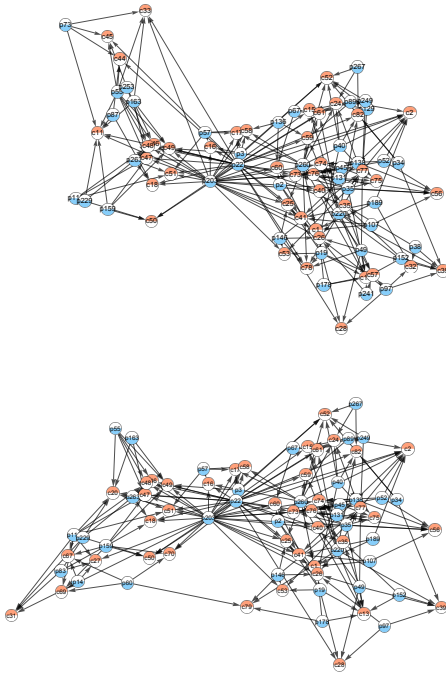


Figure 2: Two 3-4 HA bi-concept extents from the experiments on the Participant-Campaign two-mode network. On top the intent requires campaigns (prefixed with 'c' and red-colored) whose main objective is the faunistic inventory while on the bottom campaigns have to satisfy various constraints in particular about the species described during the campaign.

6.2 A lawyers advice network

This dataset concerns a network study of corporate law partnership that was carried out from 1988 to 1991 in New England [13]. It concerns 71 attorneys (partners and associates). In the Advice

network³, each attorney is described using various attributes. and 892 directed edges xy relate attorney x who goes to attorney y for basic professional advice. This network was investigated in [20] applying the abstract closed pattern methodology using the HA-core definition. We use here the attributed network as found in the minerLC web page (see above).

We apply our bi-pattern methodology using the 9-9 BHA-core which corresponds to a 9-9 HA-core as far as we have equal input vertex subsets $W_1 = W_2 = W$ (see Prop. 4.7). As an example, we consider the following closed bi-pattern $q = (q_1, q_2)$ where $q_1 = "25 < \text{Age} \leq 50, \text{Seniority} \leq 25"$ and $q_2 = "30 < \text{Age} \leq 65, 5 < \text{Seniority}"$. This pattern corresponds to the abstract closed bi-pattern with least homogeneity among the 82 abstract closed bi-patterns. Homogeneity is defined as the Jaccard similarity between the two components of its (non abstract) extension: $h(q) = |\text{ext}_1(q_1) \cap \text{ext}_2(q_2)| / |\text{ext}_1(q_1) \cup \text{ext}_2(q_2)|$. Our motivation is to consider bi-patterns which are unlikely to appear as cores of single patterns. This way bi-pattern analysis is complementary to single pattern analysis. We observe that 68 vertices over the 71 vertices of the whole advice network satisfy at least one of q_1 and q_2 i.e. satisfy $q_1 \cap q_2 = "25 < \text{Age} \leq 65"$. Only 24 vertices among these 68 satisfy both single patterns resulting in homogeneity $h(q) = 0.368$. The 9-9 BHA-core subgraph of q is much smaller, with 33 vertices 13 of which are both in H and A vertex subsets. Note that the 9-9 HA core associated to $"25 < \text{Age} \leq 65"$ is much larger: it contains 50 vertices with $|H \cap A| = 23$ and also is the 9-9 HA-core of the whole subgraph. This BHA-core subgraph displayed Figure 3 represents a group of young lawyers seeking advices from elder lawyers who are in the firm for at least five years.

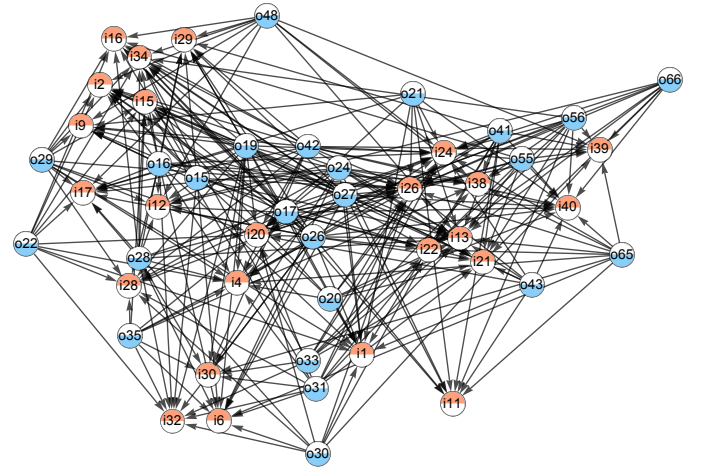


Figure 3: The 9-9 BHA-core of the lawyers advice subnetwork associated to the bi-pattern ($"25 < \text{Age} \leq 50, \text{Seniority} \leq 25"$, $"30 < \text{Age} \leq 65, 5 < \text{Seniority}"$) represented as a two-modes network in which vertices in H (resp. A) are prefixed with 'o' (resp. with 'i'). Vertices from 1 to 36 are partners, the other are associates.

³ Available at:
https://www.stats.ox.ac.uk/~snijders/siena/Lazega_lawyers_data.htm

We also experimented with a weaker 6-6 BHA-core abstraction, then resulting in 32010 abstract closed bi-patterns among which 262 have homogeneity less than 0.1. There were in particular 7 bi-patterns with null homogeneity, one among which represents lawyers from Boston whose law domain is litigation. In this bi-pattern 7 associate lawyers with age between 26 and 45 and seniority no more than 5 years go for advice to 7 older lawyers (both partners and associates) with age between 31 and 60 and seniority more than 6. Finally we conduct experiments involving 4-4 BHA-cores resulting in 293 490 bi-patterns, found in few minutes⁴, to be compared to the 930 single patterns observed in [20]. Indeed bi-pattern mining on directed networks may generate much more results than single pattern mining though focussing on low homogeneities lightens this burden. Selecting unexpected pattern, adapting the method from [20], should also be efficient though we still need some way, as the MDL pattern selection scheme (see for instance [22]), to sample among bi-patterns associated to similar cores.

6.3 Scalability

First note that we did not use any constraint on the cores sizes, i.e. we considered $s = 1$ as a minimum size threshold. This is a rather general situation: the topological constraint associated to the core definition allows a better exploration of patterns occurrences as strengthening the constraint, i.e. increasing h or a , decreases the number of closed pattern, therefore reducing the need of pruning using a minimum size threshold and allowing to find unfrequent patterns. Now, the two networks in our experiments are rather small and dense networks whose vertices have a detailed description. Scalability of the enumeration depends on the cost of core computation as well as the number of bi-patterns to output. Core computation is efficient [2], and have been performed on very large networks and the closed pattern enumeration uses an efficient top-down general algorithm [4]. However the scalability, as mentioned above, also depends on the size of the pattern language and bi-pattern mining means a pattern space which size is the product of the single pattern spaces. We still have to build large two-mode and directed attributed networks to experiment practical scalability.

7 CONCLUSION

Previously, it was shown that reducing a network to its core was the result of applying an interior operator to its vertex set, and that consequently abstract closed patterns could be searched for to mine attributed networks. In the present article we show that extending the core idea to subnetworks induced from vertex subset pairs we may define closure operators on bi-patterns resulting in bi-concept lattices. Closed bi-pattern mining as defined here appears then as a natural way to investigate attributed two-mode networks and may be applied to directed networks when considering nodes separately according to their in or out role in the core subnetwork. The methodology is illustrated on epistemological data relating campaigns to participant to these campaigns and a Lawyers advice network. The idea may be extended to multiple patterns i.e. tuples

$G_{(W_1, W_2)}$	Subgraph induced by vertex subsets W_1, W_2
(P_1, P_2)	Monotone core property pair
$(S_1, S_2) = p(W_1, W_2)$	Core subset pair of (W_1, W_2)
$G_{(S_1, S_2)}$	Core subnetwork of $G_{(W_1, W_2)}$
$q = (q_1, q_2)$	Closed bi-pattern shared by $G_{(S_1, S_2)}$ vertices

Table 1: Notations

rather than pairs, and therefore to analysis of multi mode or multi role networks, i.e. basically to address general linked data.

A NOTATIONS, DEFINITIONS AND PROOFS

Table 1 summarizes the main notations regarding bi-pattern mining on attributed graphs. Closed bi-patterns are ordered in a bi-concept lattice whose definition relies, as the concept lattice definition, on the Galois connection between an extensional and an intensional space. Hereunder we denote both order relations by the set theory inclusion symbols.

Definition A.1 (Galois connection). Let (L, \subseteq) and (X, \subseteq) be two lattices. Let int and ext be two maps defined on X and L by

$\text{int}: X \rightarrow L$

$\text{ext}: L \rightarrow X$

and such that:

C1- $\forall e, e' \in X, e \subseteq e' \text{ implies } \text{int}(e) \supseteq \text{int}(e')$

C2- $\forall c, c' \in L, c \subseteq c' \text{ implies } \text{ext}(c) \supseteq \text{ext}(c')$

C3- $\forall c \in L, c \subseteq \text{int}(\text{ext}(c))$, and $\forall e \in X, e \subseteq \text{ext}(\text{int}(e))$

Then (int, ext) define a Galois connection on (X, L)

Proposition 3.2 is then straightforward according to the componentwise definition of the orders on pairs $X = (X_1, X_2)$ and $L = (L_1, L_2)$.

Note that in closed pattern mining the Galois Connection definition is not always mentioned as such since results focus on the closure operator on the pattern language. Still, it is a simple way using Propositions 3.2 and 2.3 to obtain abstract closed bi-patterns as well as their partial ordering.

The proof of Proposition 4.3 is also straightforward:

PROOF. Let (P_1, P_2) be a pair of monotone properties, and (W_1, W_2) be a subset pair of (V_1, V_2) . Then there exists a greatest subset pair $(S_1, S_2) \leq (W_1, W_2)$ such that $P_1(v_1, S_1, S_2)$ holds for all elements v_1 of S_1 and $P_2(v_2, S_1, S_2)$ holds for all elements v_2 of S_2 .

As we consider the finite case, there are maximal subset pairs such that the required condition (referred to as C) is satisfied. We will assume that there are two maximal pairs (S_1, S_2) and (S'_1, S'_2) that satisfy C . i) This means that for any element v of S_1 we have that $P_1(v, S_1, S_2)$ holds, and as P_1 is monotone we also have that $P_1(v, S_1 \cup S'_1, S_2 \cup S'_2)$ holds. In the same way, for any element v of S'_1 we have that $P_1(v, S_1 \cup S'_1, S_2 \cup S'_2)$ also holds. This means for any element v of $S_1 \cup S'_1$ we have that $P_1(v, S_1 \cup S'_1, S_2 \cup S'_2)$ holds. ii) The same reasoning regarding S_2, S'_2 and P_2 shows that for any element v of $S_2 \cup S'_2$ we have that $P_2(v, S_2 \cup S'_2, S_1 \cup S'_1)$ holds. From i) and ii) we conclude that $(S_1 \cup S'_1, S_2 \cup S'_2)$ satisfy condition C , and is therefore greatest than both (S_1, S_2) and (S'_1, S'_2) . As both pairs

⁴673 s on a 4-core 2.2 GHz Intel Core i7 computer

⁴<http://www-lipn.univ-paris13.fr/~santini/MinerLC/>

are maximal subset pairs satisfying C , this means that $S_1 = S'_1$ and $S_2 = S'_2$. \square

REFERENCES

- [1] Sophie Bary. 2018. *Scientific representations of Biodiversity in the deep-sea : an epistemologic and scientific approach*. Ph.D. Dissertation. Ecole Doctorale numéro 474, Sorbonne Paris Cité, To appear.
- [2] Vladimir Batagelj and Matjaz Zaversnik. 2011. Fast algorithms for determining (generalized) core groups in social networks. *Adv. Data Analysis and Classification* 5, 2 (2011), 129–145. <https://doi.org/10.1007/s11634-010-0079-y>
- [3] Thomas S. Blyth. 2005. *Lattices and Ordered Algebraic Structures*. Universitext, Springer.
- [4] Mario Boley, Tamás Horváth, Axel Poigné, and Stefan Wrobel. 2010. Listing closed sets of strongly accessible set systems with applications to data mining. *Theor. Comput. Sci.* 411, 3 (2010), 691–700.
- [5] Stephen P Borgatti and Martin G Everett. 1997. Network analysis of 2-mode data. *Social networks* 19, 3 (1997), 243–269.
- [6] Stephen P. Borgatti and Martin G. Everett. 2000. Models of core/periphery structures. *Social Networks* 21, 4 (2000), 375–395. [https://doi.org/10.1016/S0378-8733\(99\)00019-2](https://doi.org/10.1016/S0378-8733(99)00019-2)
- [7] Monika Cerinsek and Vladimir Batagelj. 2015. Generalized two-mode cores. *Social Networks* 42 (2015), 80 – 87. <https://doi.org/10.1016/j.socnet.2015.04.001>
- [8] Esther Galbrun, Aristides Gionis, and Nikolaj Tatti. 2014. Overlapping community detection in labeled graphs. *Data Min. Knowl. Discov.* 28, 5-6 (2014), 1586–1610. <https://doi.org/10.1007/s10618-014-0373-y>
- [9] Bernhard Ganter and Sergei O. Kuznetsov. 2001. Pattern Structures and Their Projections. *ICCS-01, LNCS* 2120 (2001), 129–142.
- [10] Christos Giatsidis, Dimitrios M. Thilikos, and Michalis Vazirgiannis. 2013. D-cores: measuring collaboration of directed graphs based on degeneracy. *Knowl. Inf. Syst.* 35, 2 (2013), 311–343. <https://doi.org/10.1007/s10115-012-0539-0>
- [11] Jon M Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)* 46, 5 (1999), 604–632.
- [12] Stephen G. Kobourov. 2013. Force-Directed Drawing Algorithms. In *Handbook on Graph Drawing and Visualization.*, Roberto Tamassia (Ed.). Chapman and Hall/CRC, 383–408. <https://www.crcpress.com/Handbook-of-Graph-Drawing-and-Visualization/Tamassia/9781584884125>
- [13] Emmanuel Lazega. 2001. *The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership*. Oxford University Press.
- [14] Pierre-Nicolas Mougél, Christophe Rigotti, and Olivier Gandrillon. 2012. Finding Collections of k-Clique Percolated Components in Attributed Graphs. In *PAKDD, Advances in Knowledge Discovery and Data Mining, (Lecture Notes in Computer Science)*, Vol. 7302. Springer, 181–192.
- [15] Nathalie Pernelle, Marie-Christine Rousset, Henry Soldano, and Véronique Ventos. 2002. ZooM: a nested Galois lattices-based system for conceptual clustering. *J. of Experimental and Theoretical Artificial Intelligence* 2/3, 14 (2002), 157–187.
- [16] Stephen B. Seidman. 1983. Network structure and minimum degree. *Social Networks* 5 (1983), 269–287.
- [17] Arlei Silva, Wagner Meira, Jr., and Mohammed J. Zaki. 2012. Mining Attribute-structure Correlated Patterns in Large Attributed Graphs. *Proc. VLDB Endow.* 5, 5 (Jan. 2012), 466–477. <http://dl.acm.org/citation.cfm?id=2140436.2140443>
- [18] Henry Soldano and Guillaume Santini. 2014. Graph abstraction for closed pattern mining in attributed networks. In *European Conference in Artificial Intelligence (ECAI) (Frontiers in Artificial Intelligence and Applications)*, Torsten Schaub, Gerhard Friedrich, and Barry O’Sullivan (Eds.), Vol. 263. IOS Press, 849–854.
- [19] Henry Soldano, Guillaume Santini, and Dominique Bouthinon. 2017. Formal Concept Analysis of Attributed Networks. In *Formal Concept Analysis in Social Network Analysis*, Rokia Missaoui, Sergei Obiedkov, and Sergei Kuznetsov (Eds.). Springer, 143–170.
- [20] Henry Soldano, Guillaume Santini, Dominique Bouthinon, and Emmanuel Lazega. 2017. Hub-Authority Cores and Attributed Directed Network Mining. In *International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE Computer Society, Boston, MA, USA, To appear.
- [21] Henry Soldano and Véronique Ventos. 2011. Abstract Concept Lattices. In *International Conference on Formal Concept Analysis (ICFCA) (LNAI)*, P. Valtchev and R. Jäschke (Eds.), Vol. 6628. Springer, Heidelberg, 235–250.
- [22] Eirini Spyropoulou, Tijl De Bie, and Mario Boley. 2014. Interesting pattern mining in multi-relational data. *Data Min. Knowl. Discov.* 28, 3 (2014), 808–849. <https://doi.org/10.1007/s10618-013-0319-9>