

RESEARCH

Open Access



Bi-pattern mining of attributed networks

Henry Soldano^{1,2*} , Guillaume Santini¹, Dominique Bouthinon¹, Sophie Bary² and Emmanuel Lazega³

*Correspondence:

Henry.Soldano@lipn.univ-paris13.fr

¹LIPN, Université Paris-Nord
UMR-CNRS 7030, Paris, France
²ISYEB, Institut de Systématique,
Evolution, Biodiversité, UMR 7205,
MNHN, Paris, France
Full list of author information is
available at the end of the article

Abstract

Applying closed pattern mining to attributed two-mode networks requires two conditions. First, as in two-mode networks there are two kinds of vertices, each described with a proper attribute set, we have to consider patterns made of two components that we call bi-patterns. The occurrences of a bi-pattern forms an extension made of a pair of vertex subsets. Second, Formal Concept Analysis and Closed Pattern Mining were recently applied to networks by reducing the extensions of pattern to their cores, according to some core definition. We need to consider appropriate core definitions for two-mode networks and define accordingly closed bi-patterns. We describe in this article a general framework to define closed bi-pattern mining. We also show that this methodology applies as well to cores of directed and undirected networks in which each vertex subset is associated with a specific role. We illustrate the methodology first on a two-mode network of epistemological data, then on a directed advice network of lawyers and finally on an undirected bibliographical network.

Keywords: Closed pattern mining, Core subgraph, Attributed network, Two-mode network, Directed network

Introduction

The first motivation 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. 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. The present work follows previous work in which CPM and FCA were applied to undirected and directed graphs. In what follows we recall the notions which CPM of attributed networks rely on. Then we also discuss the necessity of defining bi-patterns in order to mine two-mode networks.

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 interconnected (Borgatti and Everett 2000). 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 (Seidman 1983). By changing

the topological property we obtain various core definitions within the generalized cores framework proposed by V. Batagelj (Batagelj and Zaversnik 2011).

Various recent work 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. Recently an approach has been presented extending CPM and FCA to mine attributed graphs. For that purpose, the vertex subset in which an attribute pattern occurs is reduced to its core subset using some *interior operator* (Soldano and Santini 2014). Applying interior operators to compute closed patterns make them *abstract closed pattern* for which enumeration algorithms exists (Soldano and Ventos 2011). They are called *core closed pattern* when this methodology rely on core definitions (Soldano and Santini 2014; Soldano et al. 2017a).

Now, two-mode networks are 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 (Borgatti and Everett 1997) the authors advocated the direct investigation of two-mode networks, and a core definition for two-mode networks have been recently proposed by Cerinsek and Batagelj (Cerinsek and Batagelj 2015). However applying core closed pattern mining to such two-mode networks requires to extend the methodology. The difficulty is that when such a network is attributed each kind of vertex is described according to a proper attribute set. This means that we have to consider patterns made of two attribute subsets, we further call *bi-patterns*, that each selects two interconnected vertex subsets we call its *support set pair*. 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. Interestingly, such bi-patterns may also be defined in the directed case when considering subgraphs in which a single pattern is associated to each of the *in* or *out* vertex roles. Finally we will see that the methodology we propose may also apply to undirected networks as far as we may dynamically define two different roles in the network, namely here considering in one hand high degree nodes and in the other hand their neighbours.

Note that in order to properly define bi-pattern mining we also need to extract cores from subgraphs induced by vertex subset pairs. This also means defining cores made of two vertex subsets, which goes beyond generalized cores definition.

On the computational side, we adapt the general core extraction algorithm for our new core definitions and we propose a *closed bi-pattern enumeration* algorithm that we have implemented within the minerLC software¹. We have experimented the resulting program on three networks. The first network is an epistemological two-mode network relating deep sea exploration campaigns to their participants (Bary 2018). The second network is a lawyers network in which directed links represents lawyers asking for advice from other lawyers (Lazega 2001) that was previously used to illustrate closed pattern mining of attributed directed networks (Soldano et al. 2017b). The third one is an undirected co-authoring bibliographical network investigated in (Galbrun et al. 2014).

Finally, there may be a large number of bi-patterns to extract from directed and undirected networks, when compared to single patterns: any pair of core closed single patterns is a candidate to be a core closed bi-pattern. We will propose to focus on bi-patterns in

which the two components, which are expressed in the same pattern language, are different enough. For that purpose we define a *homogeneity* measure and select inhomogeneous bi-patterns.

This work was presented in a workshop article (Soldano et al. 2018) in which the bi-pattern methodology were first introduced for two-mode and directed networks. The present article also introduces the star-satellite core definition for undirected networks and discuss the bi-patterns extracted and selected from a bibliographical network, exhibiting in particular some cooperation and competition examples in the pattern mining research domain. Overall, the main contributions of this work may be summarized as follows:

- A general definition of closed bi-pattern mining.
- A general algorithm for closed bi-patterns enumeration
- A new definition of the core of a network as a pair of vertex subsets
- A general algorithm to extract such new cores
- A definition of homogeneity for bi-patterns.
- The methodology of core closed bi-pattern mining of attributed networks, including core definitions designed respectively for two-mode, directed and undirected networks.

“[Related work](#)” section discusses related work. “[Preliminaries](#)” section gives preliminary definitions and results on core Closed Pattern Mining. In “[Bi-concept lattices and abstract closed bi-patterns](#)” section we introduce abstract closed bi-pattern mining and abstract bi-concept lattices. In “[Cores as subset pairs and core closed bi-pattern mining](#)” section we extend the definition of cores in order to obtain two-component cores and consequently define core closed bi-pattern mining. In “[Core definitions: two-Mode, directed and undirected networks](#)” section we introduce such two components cores for two-mode, directed and undirected networks. In “[Computing the interior of \$\(X_1, X_2\)\$ and enumerating abstract closed bi-patterns](#)” section we provide algorithms to compute two component cores and to enumerate the associated closed bi-patterns. Finally, in “[Experiments](#)” section we present the results obtained on the three networks mentioned above and discuss the scalability of this pattern mining methodology.

Related work

Analyzing attributed graphs led to various ways of extracting cohesive subgraphs. First, various pattern mining work investigated mining patterns as pairs of constraints on topology and labels, and rank them according to interestingness measures (Mougel et al. 2012; Silva et al. 2012). This includes abstract closed pattern mining mentioned above as well as work coming from the subgroup discovery field in which selection and pruning of interesting patterns is performed during enumeration (Atzmueller et al. 2016). A second way consists in extending *community detection* algorithms by taking into account both topology and attribute information. Various definition of hybrid objective functions and efficient ways to find optimal solutions have been proposed. In most case the result is a set of non overlapping communities (Baroni et al. 2017; Sánchez et al. 2015; Combe et al. 2015). The overlapping case has been addressed by soft clustering schemes (Xu et al. 2012), by hard clustering of the edge set (Galbrun et al. 2014) or by building generative models in such a way that a node may freely belong to several communities (Yang et al.

2013). Finally, *network embedding* algorithms have been proposed to learn an appropriate representation of nodes as vectors, and then apply standard clustering methods (Gao and Huang 2018).

In all these approaches, when considering the relationship between attributes and nodes, the latter have a unique role. This is obviously not appropriate regarding two-mode networks, while in single mode network allowing nodes to have different roles within a group may lead to a more flexible way to define cohesive subnetworks. What we propose here, beyond the extension of the core closed pattern methodology to bi-patterns, is a first step in revisiting the various methodologies mentioned above.

Regarding core definitions, recent work have proposed definitions designed to investigate directed networks. In particular a core definition has been proposed in Giatsidis et al. (2013) to investigate collaboration within directed networks. The requirement is then that both *indegrees* and *outdegrees* of vertices have to be higher than thresholds, therefore all nodes in the core are required to have both the *out* and *in* role. 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 (Kleinberg 1999). The HA-core has been recently defined in order to express this idea (Soldano et al. 2017b).

Preliminaries

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 *support set* or 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 support sets.

Enumerating the definable subsets of V comes down to enumerate the equivalence classes of patterns when considering as equivalent two patterns with same support set. Whenever the pattern language is a finite lattice there is a unique most specific pattern in each class. Recall that a lattice is such that any pair a, b of elements have both a *join* $a \vee b$ and a *meet* $a \wedge b$. The meet $a \wedge b$ is the unique greatest lower bound of a and b , i.e. $a \wedge b \leq a, a \wedge b \leq b$ and there is no $c > a \wedge b$ which is a lower bound of both a and b . In a dual way the join $a \vee b$ is the least upper bound of a and b . When considering an equivalence class of patterns, the most specific element of the class is then the meet of all its elements. This most specific pattern represents then what is common to all the patterns that occur in exactly the same object subset.

In the case of powersets, the order is the inclusion order \subseteq and join and meet respectively are set theoretical union and intersection. In standard FCA the pattern language L is the powerset 2^I of a set I of binary attributes and the extensional space is the powerset 2^V of the object set V . However, for our purpose of defining and mining bi-patterns we need a more general presentation. First, we define below closure operators together with their dual *interior operators*.

Definition 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 connection² between the pattern language and the powerset of objects:

Proposition 1 *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 $\text{ext}(q) = \{x \in V \mid q \leq d(x)\}$ be 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 $(2^V, L)$

In what follows we will use interior operators to define the general framework of abstract concept lattices. First we recall a general result (Pernelle et al. 2002; Soldano and Ventos 2011) together with a corollary defining *abstract concept lattices*:

Proposition 2 *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 $A = p[X]$ be the image of X under p , then $(\text{int}, p \circ \text{ext})$ is a Galois connection on (A, L) .*

Corollary 1 *i) $f = \text{int} \circ p \circ \text{ext}$ is a closure operator on L ii) $h = p \circ \text{ext} \circ \text{int}$ is a closure operator on A iii) The set of the (e, c) pairs where $c = f(c) = \text{int}(e)$ and $e = h(e) = p \circ \text{ext}(c)$ form a lattice, ordered following A .*

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

Abstract closed pattern mining is illustrated in Example 5 of Appendix 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. A *core* property P is defined as a mapping $P : V \times 2^V \rightarrow \{\text{true}, \text{false}\}$ where $P(v, X)$ is true whenever vertex v 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 $P(x, X_2)$ (Batagelj and Zaversnik 2011; Soldano and Santini 2014). The following result allows then to apply abstract FCA to graphs:

Proposition 3 *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 (Soldano et al. 2017a) as well as directed networks (Soldano et al. 2017b).

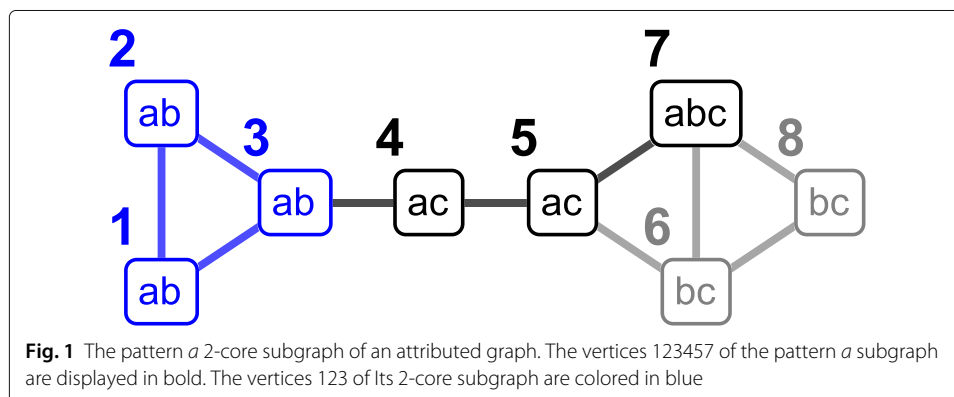
Example 1 We consider the small attributed graph displayed Fig. 1 and the 2-core property that states that in a core subgraph all vertices have degree at least 2. We have then that the support set $\text{ext}(a)$ of pattern a is 123457. The pattern a 2-core is then 123: when adding to 123 any vertex v among 457, the degree of v in G_{123v} is strictly less than 2. Therefore, $p \circ \text{ext}(a) = 123$ is the core support set of a . The corresponding core closed pattern is then $\text{int}(123) = ab \cap ab \cap ab$ i.e. the greatest pattern common to the vertices of this 2-core.

Summary

We have briefly presented standard closed pattern mining and FCA together with abstract closed pattern mining in which the support set $\text{ext}(q)$ of a pattern q is reduced to its abstract support set $p \circ \text{ext}(q)$ where p is an interior operator. The abstract closed pattern c associated to q is then the most specific pattern with the same abstract support set. We have then $c = \text{int} \circ p \circ \text{ext}(q)$ where the intersection operator int intersects the descriptions of the objects in $p \circ \text{ext}(q)$. Then we have introduced core closed pattern mining in which p reduces a vertex subset to the core of its induced subgraph. Any such core definition, including the well-known k -core, relies on a core property P such that $P(v, S)$ holds for all vertices v of the core S . In order to be a core property, P is required to satisfy a monotony condition. Core closed pattern mining consists then in enumerating the set of core closed patterns in an attributed graph.

Bi-concept lattices and abstract closed bi-patterns

This section is motivated by the extension of core closed pattern mining to two-mode networks, i.e. networks in which each 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 . This requires to extend the closed pattern mining and FCA methodology to patterns made of two components and that we call bi-patterns. A way to properly define such bi-patterns it to first extends the concept lattices of FCA. For that purpose, we need to consider lattice products and will obtain a new Galois connection.



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

Proposition 4 *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 1):

Proposition 5 *Let $X = X_1 \times X_2$ and $L = L_1 \times L_2$ be two lattices product, and let (int_1, ext_1) and (int_2, 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:*

- $int(x_1, x_2) = (int_1(x_1), int_2(x_2))$
- $ext(l_1, l_2) = (ext_1(l_1), ext_2(l_2))$

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

In what follows we consider two Galois connections as defined in Proposition 1 and use an interior operator to create the dependency between the two components of the extent which is necessary to represent cores of two-mode networks.

Proposition 2 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 2 *Let (int, ext) be the Galois connection on (X, L) as defined in Proposition 5 and let p be an interior operator on X . Then, the lattice of the Galois connection $(int, p \circ ext)$ on $(p[X], L)$ is called an abstract bi-concept lattice.*

The intents of the bi-concepts defined this way are what we call *abstract closed bi-patterns*. In a similar way as in abstract closed (single) pattern mining, each such abstract closed bi-pattern is the most specific bi-pattern c such that $p \circ ext(c)$ where p is an interior operator. However, bi-patterns occurrences are gathered in object subset pairs while closure and interior operator are self-map on lattice products. Abstract closed bi-pattern mining is illustrated in Example 6 of Appendix 2.

In what follows we apply this methodology to attributed graphs and for that purpose we define such interior operators with respect to pairs of logical properties and use them to give a new definition of cores as vertex subset pairs.

Cores as subset pairs and core closed bi-pattern mining

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 3 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 .

To obtain an interior operator, we need to define monotone properties in this context:

Definition 4 $P_1 : V_1 \times 2^{V_1} \times 2^{V_2} \rightarrow \{true, false\}$ 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 defined on $V_2 \times 2^{V_1} \times 2^{V_2}$ is monotone whenever for any $w \in V_2$, $P_2(w, W_1, W_2)$ implies $P_2(w, W'_1, W'_2)$

Cores will then be defined thanks to the following result (see proof in Appendix 1):

Proposition 6 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 5 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 3:

Proposition 7 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 $2^{V_1} \times 2^{V_2}$.

Summary

In the same way as in core closed pattern mining, given some bi-pattern $q = (q_1, q_2)$ we may compute its core support set pair $p \circ \text{ext}(q)$ where p is an interior operator. This interior operator relies on a pair of core properties that are each required to satisfy a monotony property. The associated core closed bi-pattern $c = (c_1, c_2)$ is obtained by intersecting componentwise, the vertex descriptions in $p \circ \text{ext}(q)$. Enumerating these core closed bi-patterns defines the bi-pattern mining task. In the next section we consider various core definitions to apply bi-pattern mining to two-mode, directed and undirected attributed graphs. Note that cores are here vertex subset pairs, which extends the previous (single) core notion referred to in “[Cores and closed pattern mining of attributed networks](#)” section.

Core definitions: two-Mode, directed and undirected networks

Two-mode network cores

According to this new definitions, we first define the *h-a* BHA-core of a two-mode network:

Definition 6 The *h-a* BHA-core of the two-mode 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 $v \in X_1$ 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 $v \in X_2$ in $G_{(X_1, X_2)}$ is at least a .

P_1 and P_2 are clearly monotone and therefore the *h-a* BHA-core is properly defined. This core definition is equivalent to the definition presented by Cerinsek and Batagelj (2015) in which the p - q BHA-core is called the (p, q) -core. We provide hereunder an example of an attributed two-mode network together with the set of closed bi-patterns associated to its *h-a* BHA cores.

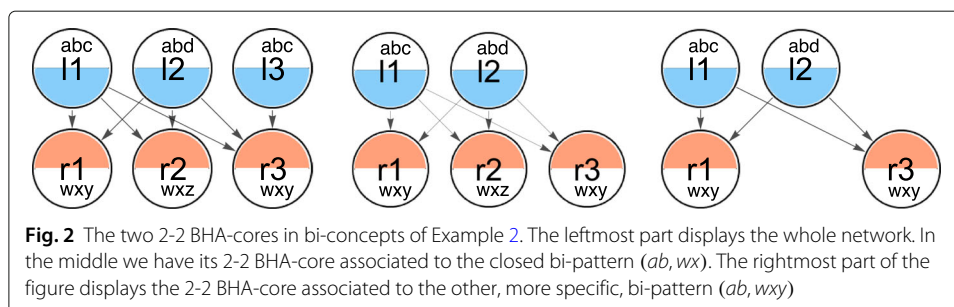
Example 2 We consider the two-mode network pictured on the leftmost part of Fig. 2. The two vertex sets are $V_1 = \{l_1, l_2, l_3\}$ and $V_2 = \{r_1, r_2, r_3\}$. Vertices of V_1 are labelled by subsets of $I_1 = \{a, b, c, d\}$ while vertices of V_2 are labelled by subsets of $I_2 = \{w, x, y, z\}$.

The most general bi-pattern (\emptyset, \emptyset) occurs in the whole network. Its 2-2 BHA-core is displayed in the middle of Fig. 2 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 Fig. 2 and has vertex sets pair $(l_1 l_2, r_1 r_3)$. This bi-pattern is closed as no item can be added without losing some vertex. Furthermore, adding any item to (ab, wxy) results in an empty 2-2 BHA-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 $((\emptyset, \emptyset), abcd, wxyz)$. Also see Fig. 4 the search tree developed for this example by the algorithm we propose in “Bi-pattern enumeration” section.

Now, let $G(V, E)$ be a single mode network, we may still consider the subgraph induced by a pair of vertex subsets according to Definition 3. This leads to core definitions for undirected and directed networks and in which vertices may have two roles.

Directed network cores : the hub and authority roles

In the directed case we reconsider the property pair of Definition 6 as a property on directed networks and obtain a BHA-core definition that extends the hub-authority core defined in Soldano et al. (2017b). We begin with a definition of the *h-a* BHA core for directed network:



Definition 7 The *h-a BHA-core* of the directed network G is defined through the following pair of core properties:

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

The BHA-core of directed networks extends the hub authority (HA) core definition:

Proposition 8 Let $G = (V, E)$ be a directed network, let (S_H, S_A) be its *h-a BHA-core*, and let $H \cup A$ be its *h-a HA core* where H and A are its hub and authority vertex subsets. Let then p_{BHA} and p_{HA} be the core operators respectively associated with the BHA core and the HA core, we have then:

$$(S_H, S_A) = (H, A) \text{ and} \quad (1)$$

$$\cup p_{BHA}(X, X) = p_{HA}(X) \quad (2)$$

for any vertex subset X .

Undirected network cores : the star and satellite roles

The previous section showed that bi-pattern mining could be applied to directed networks as far as each bi-pattern component were associated to one of the *in* and *out* roles of the vertices. In what follows we extend the *k-near-star* core which was defined on undirected networks, and exploits the two roles it relies on (Soldano and Santini 2014). This new core is called the *k-StSa* core referring to the "Star" and "Satellite" roles: a *star* vertex is required to have degree at least k while its neighbours have the *satellite* role. The *k-StSa* core subgraph is then the subgraph induced by its Star and Satellite vertex subsets as defined below:

Definition 8 The *k StSa-core* of the undirected 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 $v \in X_1$ in $G_{(X_1, X_2)}$ is at least k .
- $P_2(v, X_1, X_2)$ holds if and only if there exists some edge xv such that $P_1(x, X_1, X_2)$ holds

In the corresponding core subset pair (St, Sa) , St is called the star vertex subset and Sa the satellite vertex subset.

Star-satellite bi-pattern mining will be exemplified on an undirected bibliographical network in "[Star-Satellite bi-patterns in a bibliographical network](#)" section.

Computing the interior of (X_1, X_2) and enumerating abstract closed bi-patterns

Computing interiors

We present now the generic algorithm **Interior** that computes the interior $p(X_1, X_2) = (S_1, S_2)$ associated to the pair of monotone properties (P_1, P_2) . In the bipartite case, i.e. when $V_1 \cap V_2 = \emptyset$, the algorithm is basically a rewriting of the algorithm proposed in Cerinsek and Batagelj (2015). When considering $X_1 = X_2$, **Interior** is similar to the algorithm proposed in Soldano et al. (2017b) to compute the directed 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 needs $\mathcal{O}(m)$ operations as far as p needs only to access the neighbourhood of each vertex. The overall complexity is then $\mathcal{O}(m * n)$. A more efficient algorithm in $\mathcal{O}(m * \max(\Delta, \log n))$, where Δ is the highest degree within the graph, is obtained by adapting the variant cited in Batagelj and Zaversnik (2011) which uses two heaps as data structures for the vertex subset associated to each mode.

Interior (X_1, X_2)

```

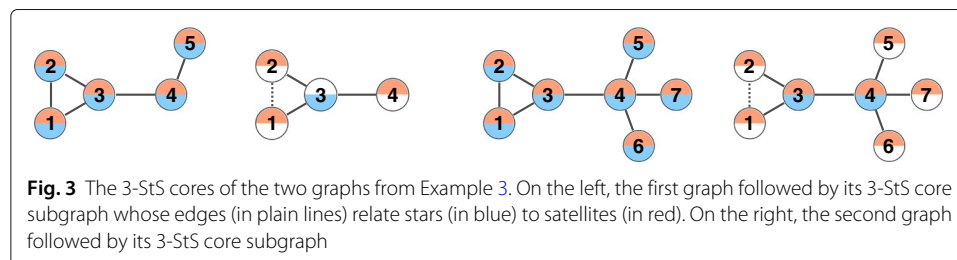
2   $S_1 \leftarrow X_1; S_2 \leftarrow X_2$ 
3  repeat
4     $Z_1 \leftarrow S_1; Z_2 \leftarrow S_2$ 
6    for  $x \in Z_1$  do
9      if  $\neg P_1(x, Z_1, Z_2)$  then withdraw  $x$  from  $S_1$  end
14   done
6    for  $x \in Z_2$  do
9      if  $\neg P_2(x, Z_1, Z_2)$  then withdraw  $x$  from 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

The following example illustrates how **Interior** computes the St-Sa core of an undirected network:

Example 3 Let $G = (V, E)$ be an undirected graph with $V = 12345$ and $E = \{12, 13, 23, 34, 45\}$. We consider its 3 StSa core. Execution of **Interior**(V, V) starts with $S_1 = S_2 = 12345$ and results in the following iterations:

1. $Z_1 = 12345$ and $Z_2 = 12345$ and then vertices 1245 are removed from S_1 as their degree in G is less than 3 while 3 and 5 are removed from S_2 as in G there is neither an edge $x3$ nor an edge $x5$ such that the degree of x is at least 3.
2. $Z_1 = 3$ and $Z_2 = 124$ and no vertex is removed from $S_1 = Z_1$ as degree of vertex 3 in $G_{(3,124)}$ still is 3. In the same way no vertex is removed from $S_2 = Z_2$ as undirected edges 31, 32, 34 are in $G_{(3,124)}$. As a result $Z_1 = S_1$ and $Z_2 = S_2$ and the iterations stop.

We note in this example that i) only one iteration is necessary to converge, which is always the case when computing k St-Sa cores and ii) $St = 3$ and $Sa = 124$ are disjoint, but this is not necessarily the case as, for instance, when adding edges 46 and 47 to G . In the new graph we obtain $St = 34$ and $Sa = 1234567$ as 3 and 4 are both stars and neighbours of each other (Fig. 3).



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 Boley et al. (2010) as implemented in the MinerLC software. 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 below with the following notations:

Let $q = (q_1, q_2)$ be a bi-pattern, i add(i, q) returns either $(q_1 \cup i, q_2)$ when $i \in I_1$ or $(q_1, q_2 \cup i)$ when $i \in I_2$, $\text{minus}(I, q)$ 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)

$S \leftarrow p(V)$
enum(int(S))

Function 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

$S_x \leftarrow p(S \cap \text{ext}(\text{add}(x, q)))$ // compute the core of its support set in S

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

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

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

enum(q_x, S_x, EL)

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

end if

end if

end for

Example 4 We follow on from Example 2 and consider $s = 1$ as the minimum 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 Fig. 2. Function **enum** is then called with the core closed pattern $q = \text{int}(\text{vs}(G_c)) = \text{int}(l_1 l_2, r_1 r_2 r_3) = (ab, wx)$ and first outputs the pair $((ab, wx), (l_1 l_2, r_1 r_2 r_3))$, and then adds to q in turn each item in $\text{minus}(I, q) = (cd, yz)$:

- $\text{add}(c, q) = (abc, wx)$ selects a subgraph whose core is empty. As a result the branch is pruned as smaller subgraphs would also result in an empty core.
- $\text{add}(d, q) = (abd, wx)$ selects also a subgraph whose core is empty.
- $\text{add}(y, q) = (ab, wxy)$ selects $(l_1 l_2 l_3, r_1 r_3)$ whose core displayed on the right of Fig. 2 has vertex set $(l_1 l_2, r_1 r_3)$. The core closed bi-pattern $q_x = (ab, wxy)$ is computed and

having null intersection with the empty list EL leads to another recursive call of **enum**. This call will output the pair $(q_x, (l_1l_2, r_1r_3))$ 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 $=\{y\}$ prior to the next iteration.

- $add(z, q) = (ab, wxz)$ selects a subgraph whose core is empty.

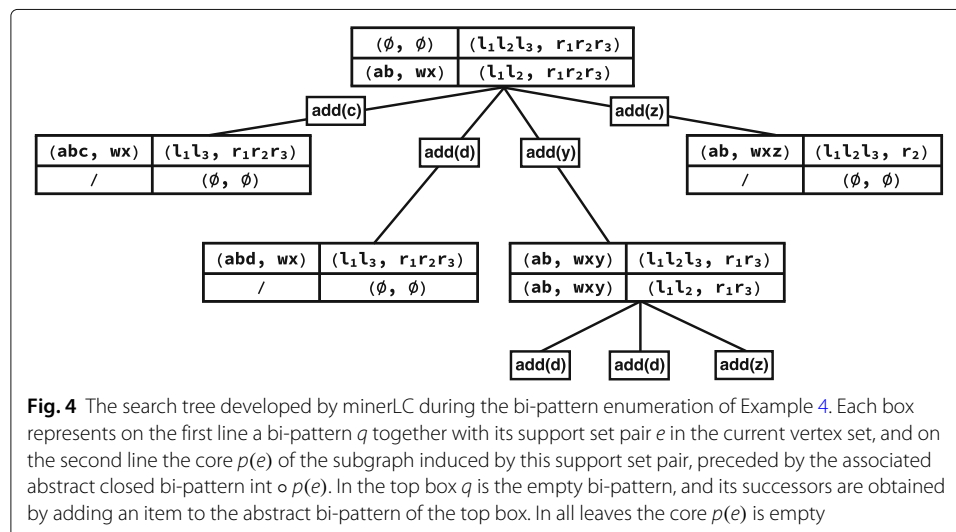
As **enum** ends **bi-patterns** also ends. The two closed bi-patterns that have been output are the most specific bi-patterns that occur respectively in the 2-2 BHA-cores, displayed on the middle and the right of Fig. 2. The search tree is represented Fig. 4.

Experiments

The first experiment concerns an original two-mode network, the second concerns a well-known directed social network available on the minerLC web page while the third one is an attributed undirected bibliographical network. The actual implementation, as part of the minerLC suite, relies on a pre-processing of the dataset that transforms the original network into a new network. Closed bi-patterns are then represented as single patterns whose items are prefixed by a role. Note that in this section there is no comparison with other programs or methods, as the task of bi-pattern mining is new as far as we know. However regarding the second dataset, we display a single pattern core subgraph, obtained in a previous work, together with a bi-pattern core subgraph sharing some nodes with the former.

h-a BHA bi-patterns in 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 exploring the deep-sea in the Indo-West Pacific region, since 1976 (Bary 2018). In this network 596 edges relate 74 campaigns (V_1) to 268 participants (V_2). 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 bi-concepts associated to 3-4 HA cores (subnetworks with participants to at least 3 campaigns with at least 4 participants to these campaigns). As an illustration Fig. 5 displays the respective 3-4 HBA-cores $S = (S_1, S_2)$ and $S' = (S'_1, S'_2)$ of two bi-patterns q and q' . The corresponding core subgraphs contains respectively $S_1 + S_2 = 80$ vertices and $S'_1 + S'_2 = 76$ vertices. Vertices are displayed at their original position in the whole network according to a standard force directed drawing (Kobourov 2013). The difference between the extents are mainly in the left part of the network, i.e. the part that corresponds to campaigns before 2000 which means that differences concern campaigns and participants which are strongly related within the original network.

h-a BHA bi-patterns in a Lawyer Advice directed network

This dataset concerns a network study of corporate law partnership that was carried out from 1988 to 1991 in New England (Lazega 2001). It concerns 71 attorneys (partners and associates). The vertices 1 to 36 represent partners while vertices 37 to 71 represents associates, i.e. attorneys with a lower position in the firm. 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 Soldano et al. (2017b) applying the abstract closed pattern methodology using the HA-core definition. We use here the same attributed network as found in the minerLC web page (see above).

There may be many bi-patterns when considering a single mode network as their number is quadratic in the number of single patterns in the same network. We will focus on bi-patterns associated to cores which are unlikely to appear as cores of single patterns. In this way, bi-pattern analysis is complementary to single pattern analysis. For that purpose we define the *homogeneity* of a bi-pattern as the Jaccard similarity of its components support sets. Homogeneity is then 1 when $q_1 = q_2$ and 0 when q_1 and q_2 never both occur in the same vertex. We will then select bi-patterns with low homogeneity.

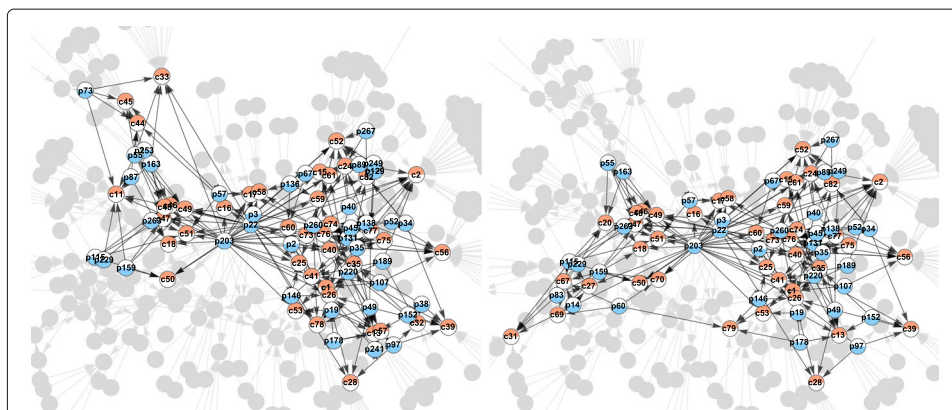


Fig. 5 Two 3-4 HA bi-concept extents from the experiments on the Participant-Campaign two-mode network. On the left the first bi-pattern select campaigns (prefixed with 'c' and red-colored) whose main objective is the faunistic inventory while on the right the second bi-pattern select campaigns that satisfy various constraints in particular about the species described during the campaign

Definition 9 (Homogeneity of a bi-pattern $q = (q_1, q_2)$)

$$h(q) = \frac{|ext_1(q_1) \cap ext_2(q_2)|}{|ext_1(q_1) \cup ext_2(q_2)|}$$

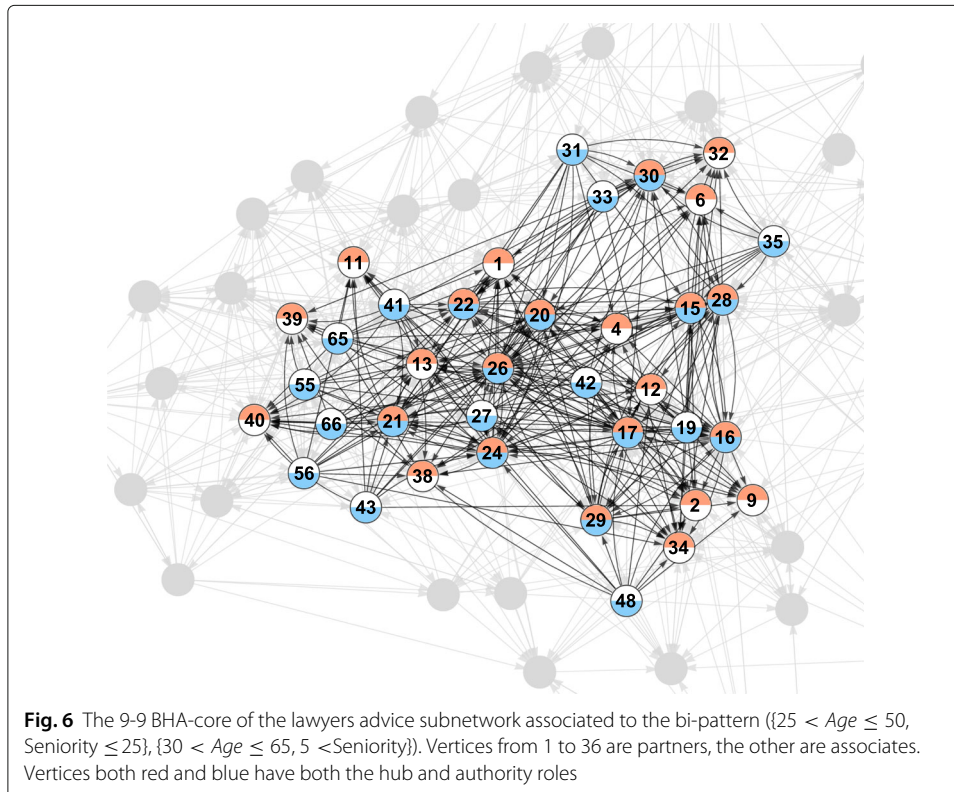
We apply our bi-pattern methodology using the 9-9 BHA-core which corresponds to a 9-9 HA-core as far as we have equal input vertex subsets $W_1 = W_2 = W$ (see Proposition 8). As an example, we consider the following closed bi-pattern $q = (q_1, q_2)$ where

$$q_1 = \{25 < Age \leq 50, Seniority \leq 25\} \text{ and}$$

$$q_2 = \{30 < Age \leq 65, 5 < Seniority\}.$$

This bi-pattern is the abstract closed bi-pattern with least homogeneity among the 82 abstract closed bi-patterns. It represents a group of young lawyers seeking advices from older lawyers who are in the firm for more than five years. We observe that 68 vertices over the 71 vertices of the whole advice network satisfy what is common to q_1 and q_2 i.e. satisfy $q_1 \cap q_2 = \{25 < Age \leq 65\}$. Only 24 vertices among these 68 satisfy both patterns q_1 and q_2 resulting in homogeneity $h(q) = 0.368$. The 9-9 BHA-core subgraph of q is displayed Fig. 6. It is made of 33 vertices 13 of which are both in H and A vertex subsets. Note that the 9-9 HA core associated to the single abstract closed pattern $\{25 < Age \leq 65\}$ is much larger: it contains 50 vertices with $|H \cap A| = 23$ and also is the 9-9 HA-core of the whole graph.

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

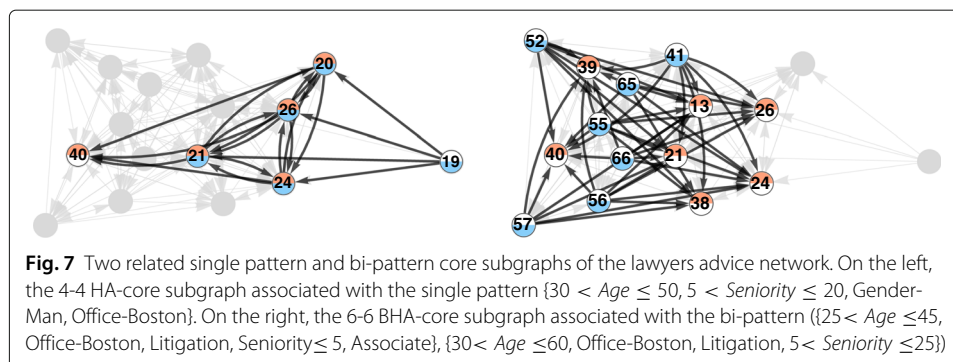


in particular 7 bi-patterns with null homogeneity, one among which represents lawyers from Boston whose law domain is litigation. In this bi-pattern 7 associate lawyers with age between 26 and 45 and seniority no more than 5 years go for advice to 7 older lawyers (both partners and associates) with age between 31 and 60 and seniority more than 6. The associated core subgraph is displayed on the right part of Fig. 7. This bi-pattern reflects the composition and cohesion of one of the relatively stable teams of lawyers on the litigation side in this Boston office. It shows the very special proximity in this team between, on the one hand, Partners 13, 21, 24 and 26 as well as senior Associates 38, 39 and 40 (in red) and on the other hand the more junior Associates (in blue) who seek advice from the former. A single Pattern 4-4 BHA-core, previously discussed in Soldano et al. (2017b), is displayed on the left of Fig. 7 and identifies an even stronger tie between these Partners and senior Associate 40 who, in 1991, was sought out for advice by the Partners themselves in breach of the unspoken status rule related to advice seeking ('You do not seek advice from others lower in the social pecking order'). In [13, page 107], blockmodelling clustered Associates 38 and 40 in these Partners' position (Position One) as structurally equivalent to them, an exceptional status heterogeneity. A year later, still as exceptionally, Associate 40 (male) was made partner. More senior Associates 38 and 39 (both female) had to wait for longer (Associate 38 made it to partnership two years later). Based on the up or out rule, Associate 39 (who was not part of Position One to begin with) had to leave the firm. Inspection of these pattern and bi-pattern thus captures a very real process.

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 Soldano et al. (2017b).

Star-Satellite bi-patterns in a bibliographical network

We also investigated the co-authoring network DBLP.E extracted from the DBLP database. DBLP.E is part of a family of networks used in various experiments on graph mining (Galbrun et al. 2014). To build the vertices description, first the terms in the titles of the author's articles were gathered and stemmed. Stop-words as well as terms that occur with more than 60% of the authors were then removed. Finally, each researcher is labelled by the terms whose occurrence count is higher than one percent of the total volume of terms for that researcher. The network is the ego-network of radius 2 of co-authors of George Karypis and has 721 authors connected by 1427 undirected co-authoring links. The maximum vertex degree is 68 and the average vertex degree is 3.95. Each vertex is



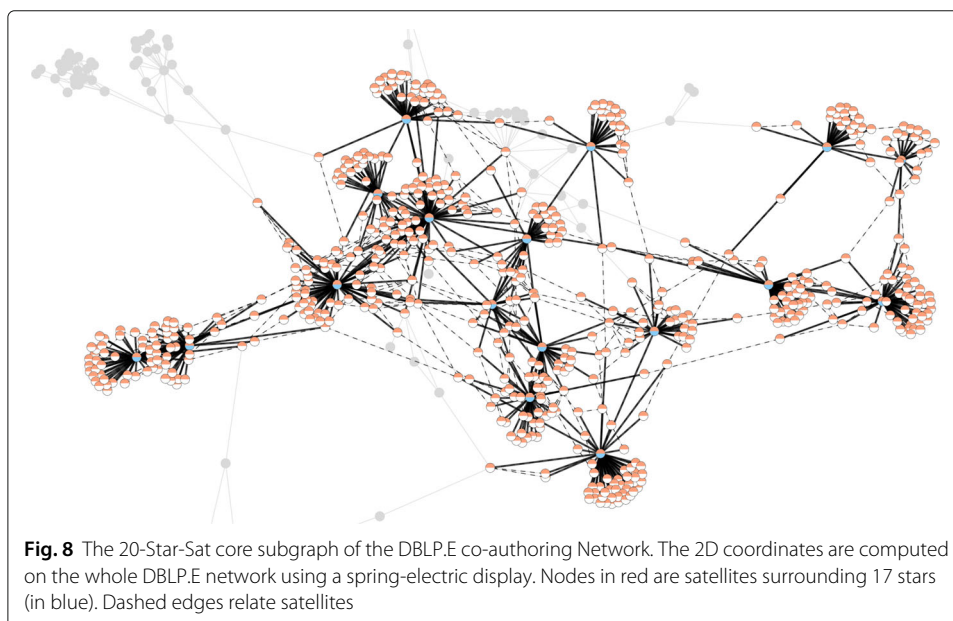
described by a subset of labels among a set of 2782 labels and the average vertex description size is 23.9. We experimented bi-pattern mining with 20-Star-Satellites cores. The core of the whole network is made of 17 stars among a total of 589 nodes in the core. We display Fig. 8 this core subgraph in which blue nodes represent stars and red nodes represent satellites. Note that all blue nodes are also red nodes. This means that any star, i.e. an author with at least 20 co-authors, is also a satellite of, i.e. is connected to, another star.

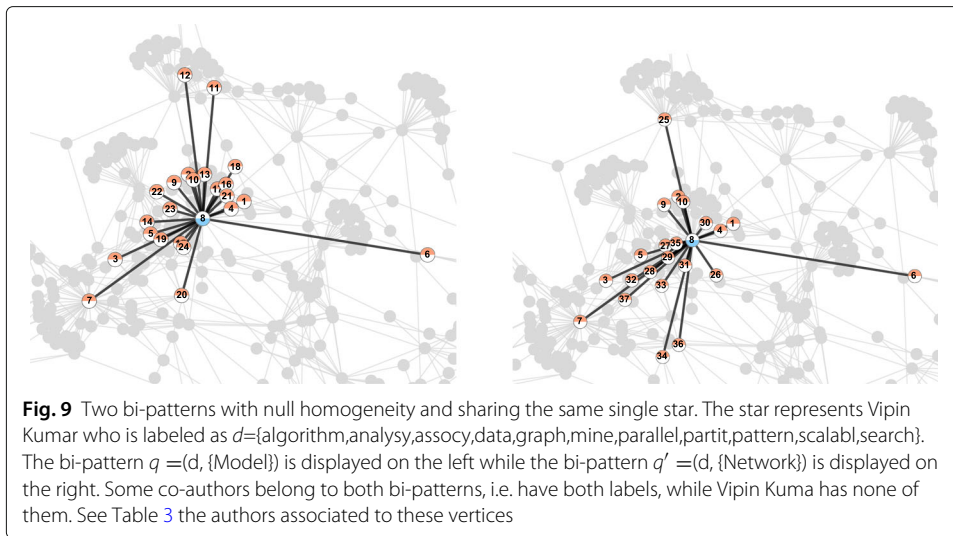
We obtained 214 bi-patterns among which we found in particular bi-patterns representing single stars with all their satellites. Most of such bi-patterns have the form $(d(s), \emptyset)$ where $d(s)$ is the description of the star s and in which the satellites have no common label. When considering homogeneity as defined above, these single star bi-patterns have low homogeneity. We also found bi-patterns made of a single star with null homogeneity, meaning the co-authors of this single star in the core subgraph have at least one common label they do not share with the star. We display Fig. 9 two such bi-patterns sharing the same single star.

With low homogeneity we also have a bi-pattern representing a pair of co-authors, namely Jianyong Wang and Lizhu Zhou, who are both stars and satellites (since an edge relate them). Such a bi-pattern represents a close cooperation between two senior researchers. Conversely, we have a bi-pattern with two unconnected stars, namely Mohammed J. Zaki and Jianyong Wang, who share labels {cluster,data,databas,efficy,frequ,graph,mine,pattern} but no satellites, thus suggesting some competition on close subjects. The corresponding core subgraphs are displayed Fig. 10.

Scalability

First note that we did not use any constraint on the cores size, i.e. we considered $s = 1$ as a minimum size threshold. This is a rather general situation: the topological constraint associated with the core definition allows a better exploration of patterns occurrences since strengthening the constraint, i.e. increasing h or a , decreases the number of closed





patterns, therefore allowing to find unfrequent patterns. Now, the first 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 as far as the logical property P only depends on neighbours of the considered vertex (Batagelj and

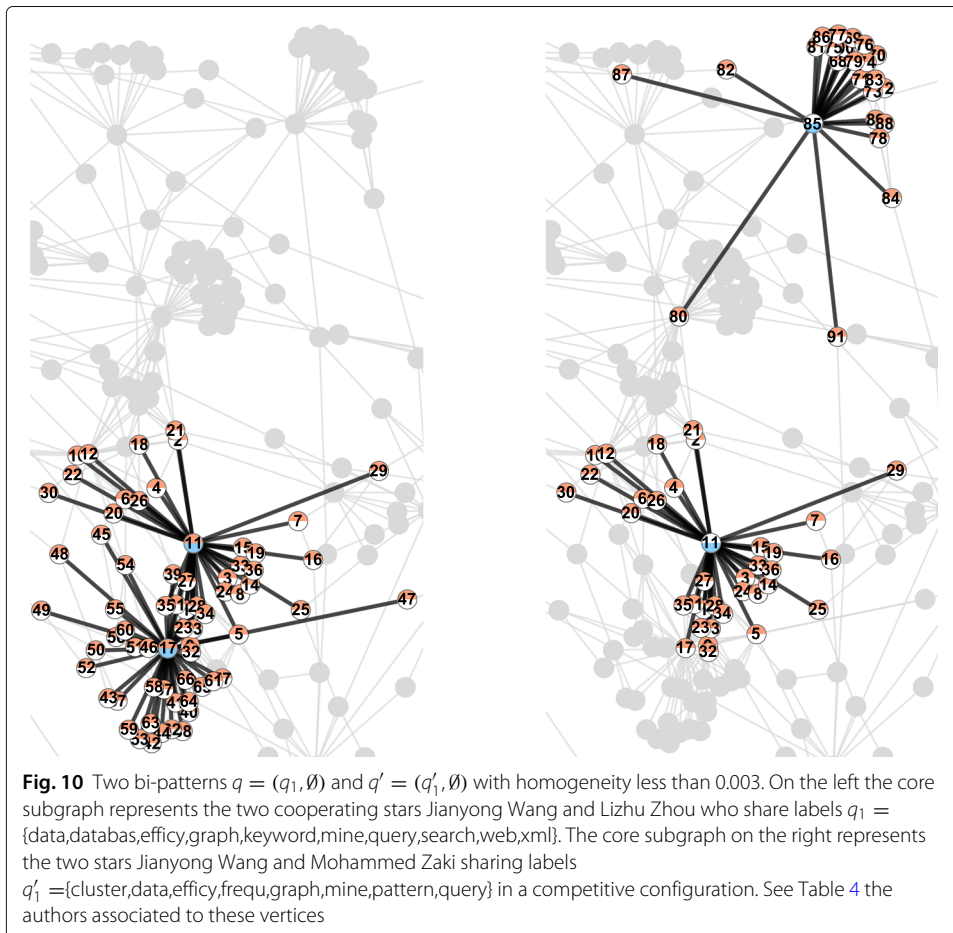


Table 1 Two components core definitions for various kind of networks

Two-mode	$X_1 \subseteq V_1, X_2 \subseteq V_2$	BHA-core subgraph $G' = G_{W_1, W_2}$ of G_{X_1, X_2} s.t. in G' degrees in W_1 at least h and degrees in W_2 at least a
Directed	$X_1 \subseteq V, X_2 \subseteq V$	BHA-core subgraph $G' = G_{W_1, W_2}$ of G_{X_1, X_2} s.t. in G' outdegrees in W_1 at least h and indegrees in W_2 at least a
Undirected	$X_1 \subseteq V, X_2 \subseteq V$	StSa-core subgraph $G' = G_{St, Sa}$ of G_{X_1, X_2} s.t. in G' degrees in St at least k and degrees in Sa at least 1

Zaversnik 2011), and has been performed on very large networks. Regarding the closed pattern enumeration, our algorithm is based on an efficient top-down general algorithm (Boley et al. 2010) and the implementation uses data reduction techniques borrowed from (Negrevergne et al. 2013). However the scalability, as mentioned above, depends on the number of bi-patterns to generate. This number 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. Note that though the vertices of the undirected network ICDM_E are described in a large language, each vertex is described with a small number of terms. As a consequence the number of bi-patterns with different cores is limited and the enumeration stops after few minutes (namely 470 s). We still have to experiment bi-pattern mining on large attributed networks of hundred thousands of nodes and edges. The ICDM_E case shows that, as far as we consider strong enough core definitions, we may investigate a large network in a reasonable time. In the general case there may be a large number of bi-patterns to investigate (see, for instance, the 4-4 BHA experiments at the end of “[h-a BHA bi-patterns in a Lawyer Advice directed network](#)” section). Only considering, as a post-processing, bi-patterns with low homogeneity allows to reduce the number of patterns to examine, while selecting unexpected patterns, adapting the method from (Soldano et al. 2017b), should also be efficient. Finally, in order to present to domain experts a limited number of interesting patterns, we still need some way, as the Minimum Description Length pattern selection scheme (see for instance Spyropoulou et al. (2014)), to sample among bi-patterns associated with similar cores.

Summary

Table 1 summarizes the various two component cores used in the bi-pattern mining problems we have investigated. The definitions are very close but concern different types of networks. More core definitions are obviously possible as far as monotony of properties pair, as defined in Definition 4, is satisfied. For sake of simplicity the BHA and StSa cores have been defined using subgraphs induced by vertex subset pairs, according to Definition 3. However, this is not mandatory and could preclude some interesting core definitions. For instance, core definitions designed to constrain some core-periphery structure should take also into account edges relating nodes within one of the vertex subset pair.

Conclusion

In this article we have extended the core closed pattern methodology in order to address two-mode attributed networks. For such networks there were no methodology, to the best of our knowledge, to extract subnetworks according to constraints on both topology and attributes. For that purpose, we have first extended the core notion: a core subgraph is now induced from a pair of vertex subsets. In each vertex subset the nodes have to satisfy an associated topological property. We may then start from any vertex subset pair and reduce this subset pair to its core, according to this new definition. We have then defined a bi-pattern as a pattern pair each component of which selects a vertex subset. This leads to define core closed bi-pattern mining which is a new and natural way to investigate attributed two-mode networks: each component of a bi-pattern select the nodes associated to a mode. We have also provided efficient algorithms to extract cores and enumerate core closed bi-patterns.

Closed bi-pattern mining as defined here may be applied to single mode networks when considering nodes separately according to two different roles. In directed networks we may then straightforwardly consider the in and out roles of nodes. In undirected networks we may still apply bi-pattern mining as far as the core definition relies on two different roles, as exemplified when introducing the star-satellite core. In these single mode networks bi-pattern mining allows to extract information which is not accessible using standard pattern mining: we may rank or select bi-patterns with low homogeneity i.e. whose components select vertex subsets with a limited or null overlap. This allows for instance to extract bi-patterns representing young lawyers asking for advice to older lawyers or representing a group of coauthors made of senior researchers sharing a large list of keywords together with a set of junior co-authors who share few or no keywords.

It should be emphasized i) that the results and definitions presented in this article may be extended to multiple patterns i.e. tuples rather than pairs, and therefore to the analysis of multi mode or multi role networks, and ii) that by using appropriate core and multi-pattern definitions, the methodology may also be extended to multiplex networks i.e. basically to address general linked data. For instance, the core of a multiplex network may be obtained in the same way as the BHA core of a directed network: as edges have a type we may associate a node degree with each edge type, associate a role to each edge type and require nodes to have a sufficient degree to belong to the corresponding role component in the core. We could then investigate, for instance, gene regulation networks by considering two different types of regulation: a regulator may either increase or decrease the gene expression. Note that in this case edges have both a direction and a type. There is no technical difficulty in defining appropriate cores in such situations, but of course core definitions as well as multi-pattern definitions, should be accurately designed according to the questions we intend to investigate: we may or not be interested in the direction according to the specific biological question we consider.

Endnotes

¹ <https://lipn.univ-paris13.fr/MinerLC/>

² Galois connections are defined in Appendix 1

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

⁴ 673 s on a 4-core 2,2 GHz Intel Core i7 computer

Appendix 1: Notations, Definitions and Proofs

Table 2 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. We denote both order relations by the set theory inclusion symbols.

Definition 10 (Galois connection) *Let (L, \subseteq) and (X, \subseteq) be two lattices. Let int and ext be two maps defined on X and L by*

$$int: X \rightarrow L$$

$$ext: L \rightarrow X$$

and such that:

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

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

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

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

Proposition 5 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 5 and 2 to obtain abstract closed bi-patterns as well as their partial ordering.

The proof of Proposition 6 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 $P_2(v, S_2 \cup S'_2, S_2 \cup S'_2)$ 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 are maximal subset pairs satisfying C , this means that $S_1 = S'_1$ and $S_2 = S'_2$. \square

Table 2 Notations

$G_{(W_1, W_2)}$	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

Appendix 2: Examples of abstract closed pattern and bi-pattern mining

In this section, we exemplify abstract closed pattern mining discussed in “Abstract closed pattern mining and concept lattices” section and abstract closed bi-pattern mining presented in “Bi-concept lattices and abstract closed bi-patterns” section. We first note an useful one to one correspondance between interiors operators on a lattice and their range (see (Blyth 2005) for the dual result on closure operators):

Proposition 9 *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 $f(x) = \bigvee_{\{a \in A \mid a \leq x\}} a$.*

We further call A an *abstraction* of X , hence we may define abstract concept lattices through interior operators as well as abstractions. By A is closed under join means we intend that the join of any subset $\{W_1, \dots, W_n\}$ of A , including the empty subset \emptyset , belongs to A . In the bi-pattern case, X is a pair $(2^{V_1}, 2^{V_2})$ of powersets and an element W of A is a pair of object subsets.

We give now a simple example of abstract closed pattern mining.

Example 5 *We exemplify 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 $\text{ext}(b) = 23$ 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 $abcd$.*

Now, to exemplify abstract closed patterns, we consider the operator p on 2^V such that $p(e) = e$ except for singletons whose images are the empty set: $p(1) = p(2) = p(3) = \emptyset$. It is straightforward following Definition 1 that p is an interior operator and as a consequence of Proposition 2, $f = p \circ \text{int} \circ \text{ext}$ is a closure operator. As we have $p \circ \text{ext}(ab) = p(23) = 23$, we obtain that $f(ab) = abc$ as in the non-abstract case. However $p \circ \text{ext}(abc) = p(3) = \emptyset$ and now $f(abc) = abcd$ is the greatest element with empty abstract support set.

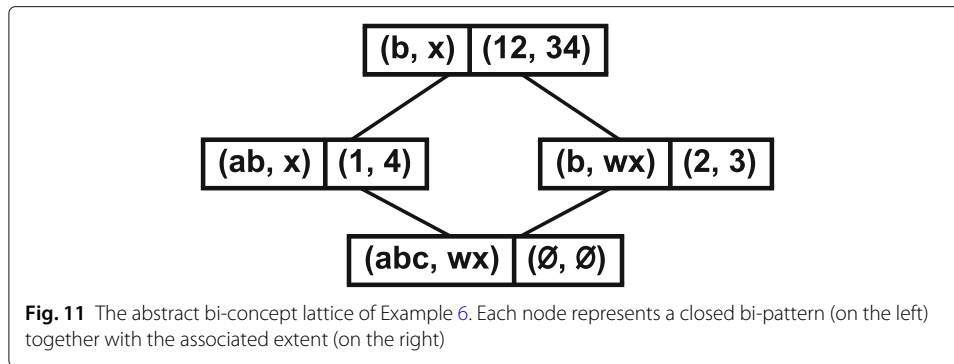
The corresponding abstraction $A = p[2^{123}]$ is generated by union closure of size 2 subsets $\{12, 23, 13\}$ and it is straightforward that for any e , $p[e]$ is the greatest subset of A smaller than or equal to e . For instance, $p[12] = 12$ as 12 belongs to A while $p[1] = \emptyset$ as no element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

We provide hereunder an example of closed bi-pattern mining that makes use of Proposition 9 to represent the interior operator.

Example 6 *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 $int(1, 4) = (ab, x)$
- $p(12, 3) = (2, 3)$ and $int(2, 3) = (b, wx)$
- $p(1, 3) = p(\emptyset, 3) = (\emptyset, \emptyset)$ and $int(\emptyset, \emptyset) = (abc, wx)$

We obtain then the abstract bi-concept lattice displayed Fig. 11. The set of abstract closed-bi-patterns with extent different from (\emptyset, \emptyset) is then $\{(b, x), (ab, x), (b, wx)\}$.

Appendix 3: Supplementary details on experimental results

Table 3 Authors from the DBLPE dataset and their index as it appears on the core subgraphs of the bi-patterns q and q' depicted en left and right parts of Fig. 9

$core(q) \setminus core(q')$	$core(q) \cap core(q')$	$core(q') \setminus core(q)$
11- Ahmed_H_Sameh	1- Anurag_Srivastava	25- Ananth_Grama
12- Aleksandar_Lazarevic	2- Clement_H_C_Leung	26- Aysel_Ozgur
13- Aparajita_Ojha	3- Jaideep_Srivastava	27- Chad_L_Myers
14- Arindam_Banerjee	4- K_Ramesh	28- Gaurav_Pandey
15- Christopher_Potter	5- Nishith_Pathak	29- Gowtham_Atluri
16- Dana_S_Nau	6- Sanjay_Ranka	30- Karsten_Steinhaeuser
17- Eui-Hong_Han	7- Shashi_Shekhara	31- Levent_Ertöz
18- Gul_Agha	8- Vipin_Kumar	32- Minesh_B_Amin
19- György_J_Simon	9- Yongdae_Kim	33- Pang-Ning_Tan
20- Hui_Xiong	10- Yow-Jian_Lin	34- Rajat_Aggarwal
21- Kirk_Schloegel		35- Rohit_Gupta
22- Maria_L_Gini		36- Yan_Huang
23- Peter_W_Li		37- Zhi-Li_Zhang

The two associated cores share the same unique star Vipin Kumar (in bold) but have different satellite subsets. By $core(q)$ we intend the set of vertices of the q core subgraph, i.e. the union of its stars and its satellites. The left column displays authors that belong only to the q core subgraph, the centre column displays authors that belong to both cores, and the right column displays authors that belong only to the q' core subgraph

Table 4 Authors from the DBLPE dataset and their index as it appears on the core subgraphs of the bi-patterns q and q' depicted on the left and right part of Fig. 10

$\text{ext}(q) \setminus \text{ext}(q')$	$\text{ext}(q) \cap \text{ext}(q')$	$\text{ext}(q') \setminus \text{ext}(q)$
37- Ali_Daud	1- Bing_Lv	68- Adriano_Veloso
38- Chao_Li	2- Charu_C_Aggarwal	69- Arlei_Silva
39- Chen_Li	3- Chuancong_Gao	70- Benjarath_Phoothakdee
40- Chun-Xiao_Xing	4- Chun_Li	71- Boleslaw_K_Szymanski
41- Chunxiao_Xing	5- Guoliang_Li	72- Chris_Bystroff
42- Fangzhen_Lin	6- Guozhu_Dong	73- Christopher_Bystroff
43- Faqir_Muhammad	7- Hongfei_Yan	74- Christopher_D_Carothers
44- Fengrong_Gao	8- Huafeng_Chen	75- Dorgival_Olavo_Guedes_Neto
45- Gang_Li	9- Jianhua_Feng	76- Geng_Li
46- Hang_Guo	10- Jian_Pei	77- Hélio_Almeida
47- Hao_Wu	11- Jianyong_Wang	78- Hilmi_Yildirim
48- Jie_Tang	12- Jiawei_Han	79- Jierui_Xie
49- Juan_Liu	13- Jun_Zhang	80- Jiong_Yang
50- Juanzi_Li	14- Krishna_Gade	81- Karam_Gouda
51- Ju_Fan	15- Lili_Jiang	82- Lizhuang_Zhao
52- Jun_Han	16- Lin_Guo	83- Medha_Atre
53- Kewen_Wang	17- Lizhu_Zhou	84- Mohammad_AI_Hasan
54- Ling_Feng	18- Min_Wang	85- Mohammed_J_Zaki
55- Ling_Lin	19- Ning_An	86- Mosab_Hassaan
56- Lin_Qiao	20- Petre_Tzvetkov	87- Naren_Ramakrishnan
57- Na-Ta	21- Phillip_S_Yu	88- Saeed_Salem
58- Qi_Guo	22- Ping_Luo	89- Vineet_Chaoji
59- Qinke_Wang	23- Qian_Qian	90- Wagner_Meira_Jr.
60- Xiang_Li	24- Qingyan_Yang	91- Yongqiang_Zhang
61- Xuhui_Liu	25- Wei_Feng	
62- Xutao_Du	26- Wei_Shen	
63- Yaoqiang_Xu	27- Xiaofang_Zhou	
64- Yi_Wang	28- Xiaoming_Fan	
65- Yong_Zhang	29- Xiaoming_Li	
66- Yuguo_Liao	30- Ying_Lu	
67- Zhiqiang_Zhang	31- Yukai_He	
	32- Yu_Wang_0011	
	33- Yuzhong_Sun	
	34- Yuzhou_Zhang	
	35- Zhiping_Zeng	
	36- Zhiwei_Xu	

The left column displays authors that belong only to the q core subgraph, the centre column displays authors that belong to both cores, and the right column displays authors that belong only to the q' core subgraph
 Authors in bold are stars in the core subgraphs depicted in Fig. 10

Acknowledgments

Not Applicable.

Funding

This research has received funding from the Project Chistera Adalab (ANR-14-CHR2-0001-04).

Availability of data and materials

The datasets and program sources are available at <https://lipn.univ-paris13.fr/MinerLC/> under a particular "Submission to Applied NetWork Science" section. Datasets are given in the format required by MinerLC.

Authors' contributions

All authors equally contributed to the whole manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹LIPN, Université Paris-Nord UMR-CNRS 7030, Paris, France. ²ISYEB, Institut de Systématique, Evolution, Biodiversité, UMR 7205, MNHN, Paris, France. ³Sciences Po, IUF, CSO-CNRS, Paris, France.

Received: 15 November 2018 Accepted: 12 May 2019

Published online: 14 June 2019

References

- Atzmueller M, Doerfel S, Mitzlaff F (2016) Description-Oriented Community Detection using Exhaustive Subgroup Discovery. *Inf Sci* 329:965–984
- Baroni A, Conte A, Patrignani M, Ruggieri S (2017) Efficiently clustering very large attributed graphs. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining ASONAM '17. ACM, New York. pp 369–376
- Bary S (2018) Scientific representations of biodiversity in the deep-sea: an epistemologic and scientific approach. PhD thesis, Ecole Doctorale numéro 474, Sorbonne Paris Cité. Defended October 10th 2018
- Batagelj V, Zaversnik M (2011) Fast algorithms for determining (generalized) core groups in social networks. *Adv Data Anal Classif* 5(2):129–145
- Blyth TS (2005) Lattices and Ordered Algebraic Structures. Universitext. Springer
- Boley M, Horváth T, Poigné A, Wrobel S (2010) Listing closed sets of strongly accessible set systems with applications to data mining. *Theor Comput Sci* 411(3):691–700
- Borgatti SP, Everett MG (1997) Network analysis of 2-mode data. *Soc Netw* 19(3):243–269
- Borgatti SP, Everett MG (2000) Models of core/periphery structures. *Soc Netw* 21(4):375–395
- Cerinek M, Batagelj V (2015) Generalized two-mode cores. *Soc Netw* 42:80–87
- Combe D, LARGERON C, Géry M, Egyed-Zsigmond E (2015) I-Louvain: An attributed graph clustering method. In: Fromont E, De Bie T, van Leeuwen M (eds). *Advances in Intelligent Data Analysis XIV. IDA 2015. Lecture Notes in Computer Science*, vol. 9385. Springer, Cham. pp 181–192
- Galbrun E, Gionis A, Tatti N (2014) Overlapping community detection in labeled graphs. *Data Min Knowl Discov* 28(5-6):1586–1610
- Gao H, Huang H (2018) Deep attributed network embedding. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18. International Joint Conferences on Artificial Intelligence Organization. pp 3364–3370
- Giatsidis C, Thilikos DM, Vazirgiannis M (2013) D-cores: measuring collaboration of directed graphs based on degeneracy. *Knowl Inf Syst* 35(2):311–343
- Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. *J ACM (JACM)* 46(5):604–632
- Kobourov SG (2013) Force-directed drawing algorithms. In: Tamassia R (ed). *Handbook on Graph Drawing and Visualization*. CRC Press. pp 383–408
- Lazega E (2001) *The Collegial Phenomenon: The Social Mechanisms of Cooperation Among Peers in a Corporate Law Partnership*. Oxford University Press
- Moguel PN, Rigotti C, Gandrillon O (2012) Finding collections of k-clique percolated components in attributed graphs. In: Tan PN, Chawla S, Ho CK, Bailey J (eds). *Advances in Knowledge Discovery and Data Mining. PAKDD 2012. Lecture Notes in Computer Science*, vol. 7302. Springer, Berlin. pp 181–192
- Negrevergne B, Termier A, Rousset M-C, Méhaut J-F (2013) Paraminer: a generic pattern mining algorithm for multi-core architectures. *Data Min Knowl Discov* 28:593–633
- Pernelle N, Rousset M-C, Soldano H, Ventos V (2002) Zoom: a nested Galois lattices-based system for conceptual clustering. *J Exp Theor Artif Intell* 2/3(14):157–187

- Sánchez PI, Müller E, Korn UL, Böhm K, Kappes A, Hartmann T, Wagner D (2015) Efficient algorithms for a robust modularity-driven clustering of attributed graphs. In: Venkatasubramanian S, Ye J (eds). Proceedings of the 2015 SIAM International Conference on Data Mining, Vancouver, BC, Canada, April 30 - May 2, 2015. SIAM. pp 100–108
- Seidman SB (1983) Network structure and minimum degree. *Soc Netw* 5:269–287
- Silva A, Meira Jr W, Zaki MJ (2012) Mining attribute-structure correlated patterns in large attributed graphs. *Proc VLDB Endow* 5(5):466–477
- Soldano H, Santini G (2014) Graph abstraction for closed pattern mining in attributed networks. In: Schaub T, Friedrich G, O'Sullivan B (eds). European Conference in Artificial Intelligence (ECAI). *Frontiers in Artificial Intelligence and Applications*, vol. 263. IOS Press. pp 849–854
- Soldano H, Ventos V (2011) Abstract Concept Lattices. In: Valtchev P, Jäschke R (eds). *Formal Concept Analysis. ICFA 2011. Lecture Notes in Computer Science*, vol. 6628. Springer, Heidelberg. pp 235–250
- Soldano H, Santini G, Bouthinon D (2017a) Formal concept analysis of attributed networks. In: Missaoui R, Obiedkov S, Kuznetsov S (eds). *Formal Concept Analysis of Social Networks. Lecture Notes in Social Networks*. Springer, Cham. pp 143–170
- Soldano H, Santini G, Bouthinon D, Lazega E (2017b) Hub-authority cores and attributed directed network mining. In: International Conference on Tools with Artificial Intelligence (ICTAI). IEEE Computer Society, Boston
- Soldano H, Santini G, Bouthinon D, Bary S, Lazega E (2018) Bi-pattern mining of two mode and directed networks. In: *WWW (Companion Volume)*. ACM. pp 1287–1294
- Spyropoulou E, Bie TD, Boley M (2014) Interesting pattern mining in multi-relational data. *Data Min Knowl Discov* 28(3):808–849
- Xu Z, Ke Y, Wang Y, Cheng H, Cheng J (2012) A model-based approach to attributed graph clustering. In: Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, SIGMOD '12. ACM, New York. pp 505–516
- Yang J, McAuley J, Leskovec J (2013) Community detection in networks with node attributes. In: 2013 IEEE 13th International Conference on Data Mining. IEEE Computer Society. pp 1151–1156

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
