

---

# Motifs abstraits et sous-communautés dans les réseaux attribués

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

1. Université Paris 13, Sorbonne Paris Cité, L.I.P.N. UMR-CNRS 7030  
F-93430, Villetaneuse, France  
henry.soldano,guillaume.santini,db@lipn.univ-paris13.fr

2. Museum National d'Histoire Naturelle, ABI, ISYEB - UMR 7205 CNRS MNHN  
UPMC EPHE, F-75005, Paris, France  
henry.soldano@mnhn.fr

---

**RÉSUMÉ.** Nous nous intéressons à la découverte de motifs dans un graphe attribué ainsi que de connaissances, sous forme d'implications. L'idée centrale est de réduire l'espace des extensions possibles dans l'ensemble des sommets du graphe et de définir des motifs fermés relativement à ces extensions réduites, dites abstraites. Au niveau global, dans une extension abstraite les sommets doivent satisfaire une propriété topologique dans le sous-graphe qu'ils induisent. On en déduit un ensemble de motifs fermés abstraits et d'implications abstraites. On définit de même une notion de motif fermé local, associé ici aux sommets d'une composante connexe d'un sous-graphe abstrait. Enfin, en dérivant un nouveau graphe du graphe original, on peut appliquer les mêmes idées pour énumérer et ordonner les  $k$ -communautés des sous-graphes induits par les motifs. Dans toutes ces situations, les implications expriment des connaissances nouvelles associées à la structure du graphe.

**ABSTRACT.** We consider attribute pattern mining in an attributed graph through recent developments of Formal Concept Analysis. The core idea is to restrain the extensional space; i.e. the space of possible pattern extensions in the vertex set  $O$ , to vertex subsets satisfying structural properties. We consider two levels. At the global level, we reduce the extension of each pattern in such a way that the corresponding abstract extension induces a subgraph whose nodes satisfy some connectivity property. At the local level a pattern has various extensions each associated to a connected component of the abstract subgraph associated to the pattern. We obtain that way abstract closed patterns and local closed patterns, together with abstract and local implication rules. We consider in particular the detection and ordering of  $k$ -communities in subgraphs of an attributed network.

**MOTS-CLÉS :** graphes attribués, fouille de données, motifs fermés, analyse formelle de concepts

**KEYWORDS:** attributed networks, closed pattern mining, formal concept analysis

---

DOI:10.3166/RIA.30.441-468 © 2016 Lavoisier

## 1. Introduction

Nous nous intéressons ici à la découverte de motifs dans des graphes attribués, plus précisément des graphes représentant des réseaux complexes, au sens large. Les sommets sont alors des individus et les arêtes ou arcs représentent une certaine relation entre ces individus. Pour caractériser ce type de graphe on s'intéresse en particulier à la distribution des degrés de ses sommets et de la longueur d'une chaîne liant une paire d'individus. Ces distributions dépendent peu de la taille du réseau. De nombreuses études ont porté sur les caractéristiques structurelles de ces réseaux, en particulier la manière de caractériser des sous-réseaux, appelés « communautés », connexes et denses et ayant peu de liens entre eux (voir (Fortunato, 2010), pour une revue sur le sujet). Une autre voie d'étude est la réduction du graphe à un sous graphe, non nécessairement connexe, dont les sommets satisfont une certaine propriété topologique  $P$  dans le sous-graphe réduit. Cela a d'abord été proposé dans (Seidman, 1983) où  $P$  requiert que les sommets  $s$  du sous-graphe réduit aient un degré  $f(s)$  au moins égal à un certain  $k$ . Pour une valeur donnée de  $k$  on obtient alors le  $k$ -core (ou  $k$ -noyau) du graphe. Cette dernière idée a ensuite été étendue à une famille de fonctions  $f$ , ce qui conduit à des *noyaux généralisés* (Batagelj, Zaversnik, 2011). Ces réductions induisent des communautés « structurelles » qui sont les composantes connexes du graphe réduit selon la propriété  $P$ , ou même d'un graphe dérivé de celui-ci dans le cas des  $k$ -communautés définies dans (Palla *et al.*, 2005). Dans ce qui suit, nous parlerons d'approches « locales » lorsque nous nous intéresserons à la notion de communauté, et d'approches « globales » lorsque nous nous ramènerons à un graphe réduit, c'est-à-dire constitué des parties denses du graphe.

Depuis quelques temps un intérêt marqué se porte sur les réseaux *attribués*, dans lesquels les sommets ou les arêtes sont étiquetés par une description, prise dans un langage de motifs contraignant les valeurs prises par un ensemble d'attributs. Dans le cadre de cette étude nous considérons que les étiquettes décrivent uniquement les sommets du graphe. On s'intéresse alors à des motifs dont on observe les occurrences dans l'ensemble des sommets du graphe. Dans (Bechara Prado *et al.*, 2013) on trouve un exemple d'approche indirecte dans laquelle les motifs représentent des co-variations portant sur les valeurs des attributs décorant les sommets du graphe. Des attributs topologiques étant ajoutés à la description des sommets, on peut utiliser des méthodes de fouilles de motifs traditionnelles ne faisant pas intervenir le graphe lui-même. Les approches directes considèrent des motifs constitués d'une partie extensionnelle, les sommets d'un sous-graphe dense, ainsi que d'une partie intensionnelle, une conjonction de contraintes portant sur les descriptions des sommets et qui doit être satisfaite dans le sous-graphe. Par la suite, nous employons le terme « motif » ou « motif d'attributs » pour cette partie intensionnelle. Par exemple (Mougel *et al.*, 2012) considère des motifs constitués d'attributs partagés dans des  $k$ -communautés de sous-graphes. Dans l'approche locale proposée dans (Silva *et al.*, 2012) les auteurs considèrent des motifs dont la partie extensionnelle est connexe et constituée de quasi-cliques. Dans (Galbrun *et al.*, 2014) un ensemble restreint de motifs est extrait d'un graphe attribué

en résolvant un problème d'optimisation. Dans ce dernier travail les motifs sont maximaux (ne peuvent être étendus) parmi ceux qui auraient la même partie extensionnelle.

Dans cet article nous commençons par une approche globale dans laquelle nous énumérons des motifs maximaux, au sens défini ci-dessus. Un motif a une *extension* dans le graphe, l'ensemble des sommets où il apparaît, qui induit un sous-graphe qui est ensuite réduit relativement à une propriété  $P$ . Les sommets de celui-ci constituent ce que nous appelons l'extension *abstraite* du motif. Dans l'approche locale, nous procéderons comme suggéré ci-dessus : on énumèrera par l'approche globale les motifs maximaux et leurs extensions abstraites, et pour chacun d'eux, à chacune des composantes connexes du sous-graphe réduit sera associé un motif, *local*, qui sera le motif le plus spécifique partagé par ses sommets. Nous présenterons ensuite une approche locale indirecte, dans laquelle à chaque motif global sera associé l'ensemble des  $k$ -communautés du sous-graphe associé à ce motif global, ainsi que les motifs locaux correspondants.

Techniquement, la première idée des travaux présentés dans cet article est de considérer les motifs *support-fermés*, au sens d'un motif maximal dans le sous-ensemble des motifs ayant même extension, et de l'étendre au cas des extensions abstraites puis locales, nous permettant ainsi de considérer des motifs abstraits et locaux support-fermés dans les graphes attribués. La seconde, plus structurelle, est de se situer dans le champ de la fouille de motifs fermés et de l'analyse formelle de concept. Dans ce domaine, les motifs support-fermés sont obtenus, lorsque le langage de description est un treillis, en utilisant un opérateur de fermeture qui, étant donné un motif, renvoie l'unique motif support-fermé de même extension. Cette seconde idée bénéficie d'avancées récentes en analyse formelle de concept, domaine au sein duquel a été proposée la notion d'abstraction comme un moyen de réduire l'espace extensionnel, c'est-à-dire l'ensemble des parties  $2^O$  de l'ensemble des objets, tout en conservant l'unicité des motifs-support fermés pour une extension (abstraite). On garantit ainsi l'existence d'un opérateur de fermeture pour les obtenir (Pernelle *et al.*, 2002 ; Soldano, Ventos, 2011). En conséquence i) les algorithmes classiques d'énumération des motifs fermés dans l'ordre général vers spécifique ne nécessitent qu'une adaptation mineure pour énumérer les *motifs fermés abstraits*, ii) les motifs fermés abstraits s'organisent en un treillis inclus dans le treillis des fermés définis sur  $2^O$ , iii) ce treillis est associé à un ensemble d'*implications abstraites* que l'on peut résumer en une base d'implications. Des travaux plus récents, motivés par la possibilité d'introduire une contrainte de connexité dans un langage de motifs tout en gardant la notion de motif fermé (Boley *et al.*, 2010), ont abouti à une généralisation de la structure de treillis, qui capture en particulier l'ordre partiel des sous-graphes connexes d'un graphe, et à une nouvelle extension de l'analyse formelle de concepts (Soldano, 2015b) dont nous tirons parti dans cet article pour formaliser la notion de motif local.

Nous considérons dans ce qui suit un graphe attribué  $G(O, E)$  d'ensembles de sommets  $O$  et d'arêtes  $E$  et nous considérons que les sommets sont étiquetés par une description dans un langage de motifs  $L$  ayant une structure de treillis, typiquement  $L = 2^I$  où  $I$  est un ensemble d'attributs binaires ou *items*. On s'intéresse alors

aux extensions des motifs de  $L$  dans l'ensemble des sommets  $O$ . Pour restreindre l'espace extensionnel  $2^O$  tout en gardant la notion de motif fermé, on le restreint à une abstraction, c'est à dire à l'image d'un *opérateur intérieur* sur  $2^O$ . Dans le cas d'un graphe attribué, cet opérateur réduit un sous-ensemble de sommets  $e$  à un sous-ensemble maximal plus petit  $p(e)$  dont les sommets satisfont une propriété  $P$  dans le sous-graphe qu'ils induisent. Nous appelons *abstraction de graphe* l'espace extensionnel obtenu comme image  $p[2^O]$  de cet opérateur. Un exemple simple de propriété que doit respecter un sommet est que son degré doit être au moins  $k$ , le sous-ensemble  $p(e)$  est alors l'union des  $k$ -core du sous-graphe  $G_e$  induit par  $e$ . Un autre exemple simple est la propriété  $k$ -clique qui est satisfaite par un sommet  $x$  relativement à un sous-ensemble de sommets  $e'$  si  $x$  appartient à une  $k$ -clique du sous-graphe induit  $G_{e'}$ . A un motif donné  $q$  on associe alors son extension abstraite  $p \circ \text{ext}(q)$  ainsi qu'un motif fermé abstrait : le plus grand motif parmi ceux de même extension abstraite. On note alors  $\Box q \rightarrow \Box w$  l'implication abstraite qui est valide si l'extension abstraite de  $q$  est incluse dans l'extension abstraite de  $w$ . Les abstractions de graphes et les motifs abstraits associés ont été présentées dans (Soldano, Santini, 2014). Ils font l'objet de la première partie du présent article.

Dans des travaux récents (Soldano *et al.*, 2015b), nous avons étendu l'analyse formelle de concept de manière analogue à ce qui avait été fait dans (Soldano, 2015b) pour le langage de motifs mais cette fois dans l'espace extensionnel, et introduit ainsi la notion de *motif fermé local* : dans le cas des graphes attribués, celui-ci est le motif le plus grand associé à un sous-graphe dense et connexe et contenant un certain sommet (ou groupe de sommets)  $i$ . Ce motif fermé local est obtenu en appliquant un opérateur de fermeture local  $f_i$ . On obtient alors des *implications locales*, vraies près de  $i$ . Dans ce cas plusieurs opérateurs de fermeture  $f_i$  peuvent être appliqués au même motif : un motif fermé sera alors local car associé au sommet ou groupe de sommets  $i$ . L'exemple le plus simple d'un tel espace extensionnel est l'ensemble des sous-ensembles de sommets du graphe  $G(O, E)$  induisant des sous-graphes connexes : l'extension  $\text{ext}(q)$  d'un motif  $q$  induit alors un sous-graphe constitué de plusieurs composantes connexes, chacune associée à un (quelconque) de ses sommets  $i$ . Chacune de ces composantes forme une extension locale  $p_i \circ \text{ext}(q)$  et le plus grand motif commun à ses sommets est le motif fermé local  $f_i(q)$ . Dans ce qui suit le sous-graphe local (ou abstrait) induit par un motif  $q$  est simplement appelé le sous-graphe local (ou abstrait) du motif  $q$ . Formellement, les éléments de cet espace extensionnel inclus dans  $2^O$  forment une *confluence* de  $2^O$  (Soldano, 2015b), une structure plus faible que celle d'abstraction, pouvant avoir plusieurs éléments minimaux (voir Section 4). Le cas mentionné ci-dessus correspond à la plus simple *confluence de graphe*, ses éléments minimaux sont les singletons  $\{v\}$ , où  $v$  est un sommet du graphe, et les éléments de cette confluence sont appelés des *sous-ensemble de sommets connexes*. On obtient alors un ensemble de concepts locaux, c'est-à-dire de paires (support local, motif fermé local). Les implications locales associées sont notées  $\Box_m q \rightarrow \Box_m w$  où  $m$  est un élément minimal de la confluence et ont la sémantique suivante: l'extension locale contenant  $m$  du motif  $q$  est incluse dans celle du motif  $w$ . Dans le cas évoqué ci-dessus,  $\Box_{\{v\}} q \rightarrow \Box_{\{v\}} w$  est valide lorsque la composante connexe contenant  $v$  du sous-graphe induit par  $q$  est

incluse dans celle du sous-graphe induit par  $w$ . On obtiendra tout un ensemble de confluences de graphes, que nous nommerons des *cc-confluences* en remarquant que l'espace extensionnel  $F_A = F \cap A$ , où  $A$  est une abstraction et  $F$  la confluence de graphe mentionnée ci-dessus, est lui-même une confluence. En pratique cela signifie que l'on peut choisir une abstraction de graphe et ne considérer dans  $F_A$ , que les sous-ensembles de sommets appartenant à  $A$  et formant un sous-graphe connexe. La figure 1 représente le sous-graphe induit par un motif d'un graphe de co-auteurs tiré de DBLP (voir Section 1). Les sommets et arêtes en rouge représentent le sous-graphe abstrait associé à la propriété *degré*  $\geq 4$  et constitué de deux composantes connexes. Un motif fermé local est associé à chacune de ces composantes. L'ensemble de ces travaux est présenté dans la Section 4.

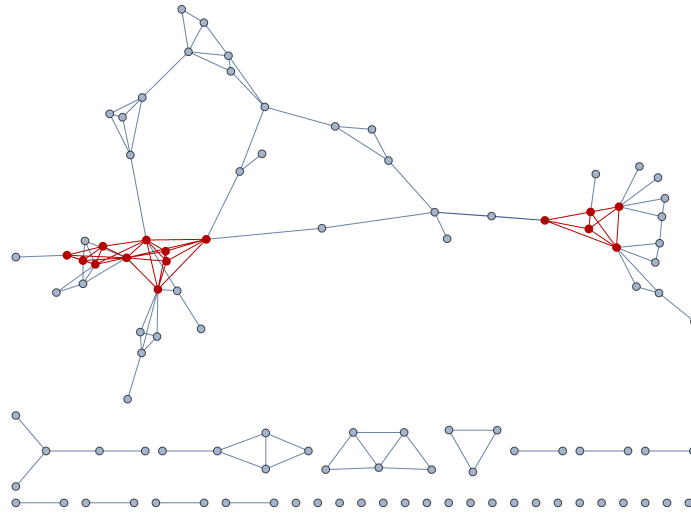


Figure 1. Le sous-graphe induit par le motif DMKD, IDArev dans le graphe de co-auteurs DBLP. Les sommets rouges (en gras) représentent l'extension abstraite *degré*  $\geq 4$

Les *cc-confluences* de graphe ne capturent pas toujours la notion de communauté telle qu'on l'entend en analyse des réseaux sociaux. La figure 2 montre un graphe dont les sommets représentent des élèves d'un collège dans l'ouest de l'Écosse étiquetés par des attributs concernant leurs habitudes de vie, et les arêtes leurs relations d'amitié<sup>1</sup>. Lorsqu'on applique une abstraction 3-clique au graphe entier on obtient le sous-graphe constitué d'arêtes et sommets en gras et colorés. On obtient 4 motifs fermés locaux  $l_i$  chacun associé à une des composantes connexes de ce sous-graphe abstrait. On constate alors que la plus grande composante connexe est constituée de plusieurs *communautés* que nous voudrions considérer lors de la recherche de motifs locaux. Ici ajouter une abstraction, par exemple de degré, ne permet pas d'isoler ces

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

composantes. Cependant, les  $k$ -communautés d'un graphe  $G$  (Palla *et al.*, 2005) sont définies comme les composantes connexes du graphe dérivé  $G_T$  dont les sommets sont les  $k$ -cliques de  $G$  et les arêtes joignent deux  $k$ -cliques partageant  $k - 1$  sommets. On peut donc considérer une cc-confluence de ce nouveau graphe, associer à chaque sous-ensemble de sommets connexes dans  $G_T$  un motif fermé local, puis revenir dans le graphe d'origine. Chaque sous-graphe coloré et entouré d'une ellipse de la figure 2 représente une telle 3-communauté de taille au moins 4 dans  $G$ . Chaque motif fermé local est le plus grand motif commun à la communauté correspondante.

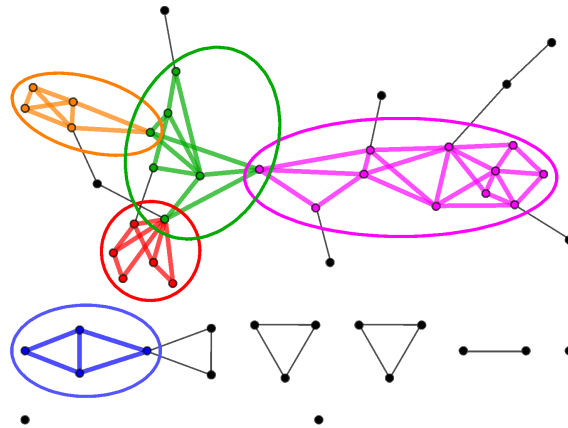


Figure 2. Un graphe d'amitié d'élèves de collège dans l'ouest de l'Écosse. Chaque ellipse représente une 3-communauté de taille au moins 4

Enfin, lorsqu'on énumère exhaustivement un ensemble de motifs ou que l'on extrait un ensemble de règles d'un jeu de données, on ne peut se contenter de sélectionner des motifs fréquents, ce qui permet un élagage efficace de l'énumération. La question de la sélection d'un sous-ensemble intéressant des motifs produits est cruciale. Nous proposons, dans l'approche globale comme dans l'approche locale, des mesures d'intérêt des motifs et règles obtenues. En ce qui concerne les motifs, nous nous inspirons de la mesure d'intérêt proposée dans (Silva *et al.*, 2012) dans le cas des quasi-cliques, pour définir des mesures de *spécificité* des motifs abstraits et locaux produits. Dans le cas des règles abstraites et locales, nous proposons des mesures d'*informativité*. Pour introduire celles-ci, rappelons que les règles abstraites et locales définies ci-dessus doivent être valides lorsqu'on réduit les extensions en requérant qu'une propriété topologique soit satisfaite ou en se focalisant dans une région du graphe. Les règles que nous considérerons les plus informatives seront alors celles qui étaient le plus loin d'être valides, avant cette réduction. Nous utiliserons pour cela une estimation de la probabilité conditionnelle de la conclusion connaissant la prémisse.

La section 2 décrit les graphes attribués utilisés dans nos expériences. Nous présentons les motifs fermés abstraits et les abstractions de graphes dans la section 3. Les motifs fermés locaux et les cc-confluences sont présentés dans la section 4. Dans ces deux

sections nous proposons également les mesures d'intérêt de motifs et d'implications, propres aux niveaux abstraits et locaux décrites dans (Soldano *et al.*, 2015a). Nous décrivons dans la section 5 les cc-confluences d'un graphe dérivé et les  $k$ -communautés associées aux sous-graphes induits par un motif. La section 6 discute les aspects algorithmiques de ce travail.

## 2. Jeux de données

Les expériences des sections suivantes sont effectuées sur deux jeux de données. Dans les deux cas nous avons un graphe  $G = (O, E)$  dont les sommets ont des descriptions dans un vecteur attribut-valeur. Les algorithmes utilisés traitant des représentations par liste d'items, nous décrivons ci-dessous, s'il y a lieu, le processus de binarisation.

### 2.1. Étudiantes écossaises et style de vie

Le jeu de données *s50-1* est un jeu de données classique pour les expériences sur les graphes attribués<sup>2</sup>. Le graphe représente 148 relations d'amitié entre élèves d'un collège dans l'ouest de l'Écosse (des jeunes filles d'environ 13 ans au début de l'étude), les étiquettes concernant l'usage de différentes substances récréatives (tabac, cannabis, alcool) ainsi que la pratique sportive. Les attributs sont ordinaux. L'attribut T concerne la consommation de tabac et prend les valeurs 1 (non), 2 (occasionnelle) et 3 (régulière). L'attribut C concerne la consommation de cannabis et prend des valeurs entre 1 (non) et 4 (régulière). L'attribut D indique la consommation d'alcool et prend des valeurs entre 1 (non) et 5 (élevée). S concerne l'activité sportive et ne prend que les valeurs 1 (occasionnelle) et 2 (régulière). Un item représente un intervalle de valeurs pour une variable. Par exemple l'item C23 est présent pour un sommet lorsque C a une valeur dans  $[2, 3]$ . Dans nos expériences nous avons fusionné les deux valeurs les plus élevées des variables T, C et D. Par exemple, les valeurs 4 et 5 de D sont fusionnées dans la valeur 45. Dans la table ci-dessous nous donnons la liste des items utilisés. Tout intervalle de valeur d'une variable peut être représenté par un sous-ensemble d'items. Par exemple,  $D=2$  est représenté par  $\{D12, D234\}$ :

Tobacco	Cannabis	Alcohol
T1, T23	C1, C12, C234, C34	D1, D12, D123, D2345, D345, D45

### 2.2. Un graphe de co-auteurs de DBLP

Le jeu de données de co-auteurs tiré de DBLP est décrit dans (Bechara Prado *et al.*, 2013). Il comprend 45 131 auteurs (sommets) et 228 188 liens de co-publication

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

(arêtes), le graphe résultant est composé de 555 composantes connexes. Les auteurs présents dans le graphe ont publié au moins un article dans un journal ou une conférence dans le domaine des bases de données ou de la fouille de données<sup>3</sup> entre janvier 1990 et février 2011. Une arête lie alors deux auteurs lorsqu'ils ont participé ensemble à au moins un article durant la même période. Pour enrichir le langage de motifs, les conférences sont regroupées en trois groupes : DB (bases de données), DM (fouille de données) et AI (intelligence artificielle) selon la nomenclature d'un site de classement de conférences<sup>4</sup>. Les items sont les journaux et les conférences ainsi que les groupes de conférences.

### 3. Motifs fermés abstraits dans les graphes attribués

#### 3.1. Motifs fermés abstraits

Un problème de fouille de données revient souvent à s'intéresser aux occurrences de motifs d'un langage  $L$  dans un ensemble d'objets  $O$ . Un motif  $q$  est alors associé à  $\text{ext}(q)$ , son ensemble-support, ou extension, dans  $O$ , qui est constitué des objets en lesquels le motif a une occurrence. Le langage est partiellement ordonné selon un ordre de spécificité et forme un treillis<sup>5</sup>. De plus chaque objet  $o$  est décrit comme un motif particulier  $d(o)$ . On dit alors que  $q$  a une occurrence dans  $o$  lorsque  $d(o)$  est plus spécifique que  $q$ . Un motif  $q$  est dit *support-fermé* lorsqu'il n'y a pas de motif strictement plus spécifique de même extension. Deux motifs ayant même extension sont considérés comme équivalents et les motifs support-fermés sont les éléments maximaux de ces classes. Dans le cadre évoqué ci-dessus il y a un unique motif support-fermé pour une extension donnée  $e$ . Dans ce cas une fonction d'intension  $\text{int}(e)$  renvoie le motif support-fermé associé à  $e$ . On obtient alors, pour tout motif  $q$  le motif support-fermé de même extension en appliquant l'opérateur de fermeture  $\text{int} \circ \text{ext}$ . Nous dirons qu'un motif est *fermé* lorsqu'il est égal à sa fermeture, les motifs support-fermés sont donc les motifs fermés selon l'opérateur de fermeture décrit ci-dessus. Remarquons que hors du cadre défini ci-dessus les motifs support-fermés ne sont pas forcément uniques dans chaque classe d'équivalence, et ne sont pas obtenus comme des motifs fermés relativement à un opérateur de fermeture<sup>6</sup>.

L'ensemble des motifs peut alors être résumé en ne s'intéressant qu'aux classes fréquentes (les extensions de taille plus grande qu'un seuil) et en prenant comme re-

3. Conférences : KDD, ICDM, ECML/PKDD, PAKDD, SIAM DM, AAAI, ICML, IJCAI, IDA, DASFAA, VLDB, CIKM, SIGMOD, PODS, ICDE, EDBT, ICDT, SAC, Journaux : IEEE TKDE, DAMI, IEEE Int. Sys., SIGKDD Exp., Comm. ACM, IDA J., KAIS, SADM, PVLDB, VLDB J., ACM TKDD.

4. <http://webdocs.cs.ualberta.ca/~zaiane/htmldocs/ConfRanking.html>. DB = {VLDB, SIGMOD, PODS, ICDE, ICDT, EDBT, DASFAA, CIKM}; DM = {SIGKDD Explorations, ICDM, PAKDD, ECML/PKDD, SDM}; AI = {IJCAI, AAAI, ICML, ECML/PKDD}.

5. Rappelons que dans un treillis toute paire d'éléments  $(x, y)$  a une borne inférieure  $x \wedge y$  (ou *inf*) et une borne supérieure (ou *sup*)  $x \vee y$ .

6. En fouille de données, on parle souvent de motifs fermés pour désigner des motifs support-fermés, ce qui peut entretenir une certaine confusion.



présentant les maxima de ces classes, les motifs fermés. Le langage de motif  $L$  est typiquement l'ensemble des parties  $2^I$  d'un ensemble  $I$  d'items et le motif fermé est obtenu en faisant l'intersection des descriptions d'objets de l'extension correspondante. Par ailleurs, une implication  $q \rightarrow w$  est dite valide sur  $O$  lorsque  $\text{ext}(q)$  est incluse dans  $\text{ext}(w)$  : lorsque le motif  $q$  a une occurrence dans  $o$ , le motif  $w$  a aussi une occurrence en  $o$ . L'ensemble des implications valides peut-être engendré par un sous-ensemble de celles-ci, qu'on appelle une base d'implications. Une classe d'équivalence a aussi des éléments minimaux, les *générateurs* et une base particulière leur est associée, la base min-max (Pasquier *et al.*, 2005) définie comme suit :

DÉFINITION 1. — *La base min-max  $m$  des implications valides sur  $2^O$  est définie comme suit :  $m = \{g \rightarrow f \mid f \text{ est un motif fermé, } g \text{ est un générateur, } f \neq g, \text{ext}(g) = \text{ext}(f)\}$*

Nous définissons maintenant formellement les opérateurs de fermeture ainsi que les *opérateurs intérieurs* qui seront utilisés pour définir les *extensions abstraites*. Dans ce qui suit tous les ensembles ordonnés<sup>7</sup> sont finis.

DÉFINITION 2. — *Soit  $E$  un ensemble ordonné et  $f : E \rightarrow E$  une application telle que pour tout  $x, y \in E$ ,  $f$  est croissante, c'est-à-dire  $x \leq y \implies f(x) \leq f(y)$  et idempotente, c'est-à-dire  $f(f(x)) = f(x)$ , alors :*

- Si  $f$  est extensive, c'est-à-dire  $f(x) \geq x$ ,  $f$  est appelé un opérateur de fermeture.
- Si  $f$  est intensive, c'est-à-dire  $f(x) \leq x$ ,  $f$  est appelé un opérateur d'ouverture ou opérateur intérieur, ou encore une projection.

*Dans le premier cas,  $f(x)$  est appelé le fermé de  $x$ , et un élément tel que  $x = f(x)$  est appelé un élément fermé. Dans le second,  $f(x)$  est appelé l'intérieur de  $x$ .*

L'image d'un opérateur intérieur sur un treillis est appelée ici une *abstraction* et est caractérisée par la proposition suivante:

PROPOSITION 3. — (Soldano, Ventos, 2011) *Un sous-ensemble  $A$  d'un treillis  $X$  est l'image  $p[X]$  d'un opérateur intérieur si et seulement si la borne supérieure de tout sous-ensemble de  $A$  appartient aussi à  $A$ , y compris le minimum  $\perp$  de  $X$ . On a alors :*

$$p(x) = \vee_{\{a \in A \mid a \leq x\}} a.$$

Si  $p$  est l'opérateur intérieur associé à une abstraction  $A$  alors  $p(x)$  est le plus grand élément de  $A$  inclus dans  $x$ . L'analyse des motifs fermés a récemment été étendue à l'analyse des motifs fermés abstraits en remarquant qu'un opérateur intérieur sur un treillis préserve la notion de correspondance de Galois. En particulier en appliquant un tel opérateur sur l'espace extensionnel  $2^O$  on obtient encore un opérateur de fermeture sur le langage de motifs (Pernelle *et al.*, 2002 ; Soldano, Ventos, 2011) :

7. Tout ensemble muni d'un ordre partiel est appelé ici ensemble ordonné

PROPOSITION 4. — Soit  $X = 2^O$  et  $L = 2^I$  liés par la correspondance de Galois  $(\text{int}, \text{ext})$ ,  $p$  un opérateur intérieur sur  $2^O$ , et soit  $A = p[X]$  l'abstraction associée, alors  $(\text{int}, p \circ \text{ext})$  est une correspondance de Galois sur  $(A, L)$ , et donc

$$f = \text{int} \circ p \circ \text{ext} \text{ est un opérateur de fermeture sur } L$$

L'extension abstraite du motif  $q$  est alors  $\text{ext}_A(q) = p \circ \text{ext}(q)$ . Il y a donc un unique motif abstrait support fermé c'est-à-dire un unique motif maximale spécifié  $f(q) = \text{intext}_A(q)$  dans la classe d'équivalence de tous les motifs ayant comme extension abstraite  $\text{ext}_A(q)$ .  $f(q)$  est appelé un motif fermé abstrait. On obtient alors un treillis de concepts abstraits :

COROLLAIRE 5. — (Soldano, Ventos, 2011). L'ensemble des paires (extension abstraite, motif fermé abstrait)  $(e = \text{ext}_A(c), c = \text{int}(e))$ , ordonné selon l'abstraction  $A$  forme un treillis appelé un treillis de concepts abstraits.

Remarquons que,  $p$  étant par définition monotone, lorsque  $\text{ext}(q) \subseteq \text{ext}(w)$ , c'est-à-dire  $q \rightarrow w$  est valide, alors  $\text{ext}_A(p) \subseteq \text{ext}_A(w)$ , et nous dirons que l'implication abstraite  $\Box_A q \rightarrow \Box_A w$ , est aussi valide. Cette notation remplace l'abstraction comme une modalité de type nécessité, et la logique modale associée est présentée en (Soldano, 2011).

Nous obtenons de cette manière une base min-max abstraite en reprenant la définition de la section 3.1 dans laquelle  $\text{ext}_A$  remplace  $\text{ext}$ , et en reliant des éléments minimaux de ces nouvelles classe d'équivalence, les  $A$ -générateurs, à l'élément maximum, le motif fermé abstrait, ou  $A$ -fermé :

DÉFINITION 6. — La base min-max abstraite  $m_A$  des implications abstraites valides sur  $A$  est définie par :

$$m_A = \{\Box_A g \rightarrow \Box_A f \mid f \text{ est un motif } A\text{-fermé, } g \text{ est un } A\text{-générateur, } f \neq g, \text{ext}_A(g) = \text{ext}_A(f)\}$$

Une implication abstraite traduit donc l'inclusion de l'extension abstraite de la prémisse dans celle de la conclusion. Si on appelle groupe abstrait un élément d'une abstraction, alors une implication abstraite  $\Box_A g \rightarrow \Box_A f$  s'interprète de la manière suivante : si un groupe abstrait partage le motif  $g$  alors il partage aussi le motif  $f$ . L'utilisation d'une telle implication pour un objet donné  $o$  nécessite donc de trouver un groupe abstrait contenant  $o$  et partageant le motif  $g$ , ce qui permet de déduire que le groupe et donc  $o$  satisfont le motif  $f$ . Remarquons que pour un objet donné seuls les groupes abstraits minimaux qui le contiennent sont à considérer. Dans les abstractions de graphe définies ci-après, ces groupes seront par exemple des  $k$ -cliques ou autres structures topologiques.

### 3.2. Abstractions de graphe

Ces idées ont été appliquées aux graphes attribués en définissant les abstractions de graphe (Soldano, Santini, 2014). L'ensemble d'objets  $O$  est alors l'ensemble de

sommets d'un graphe  $G = (O, E)$  et chaque sommet  $o$  est étiqueté par un motif  $d(o) \in 2^I$ .

On définit une abstraction de graphe à travers une propriété caractéristique  $P(x, e)$  qui requiert une forme de connectivité minimale du sommet  $x$  dans le sous-graphe  $G_e$  induit par le sous-ensemble de sommets  $e$

LEMME 7. — Soit  $P$  tel que :

- $P(x, e)$  implique  $x \in e$  et
- $e \subseteq e'$  et  $P(x, e)$  implique  $P(x, e')$ ,

et soit la fonction  $q$  définie par  $q(e) = \{x \in e \mid P(x, e)\}$ , alors la fonction  $p$  définie par  $p(e) = \text{pointFixe}(q, e)$  est un opérateur intérieur sur  $2^O$ .

$p(e)$  représente le plus grand sous-ensemble de sommets dans  $e$  dont les sommets satisfont la propriété caractéristique  $P$  dans le sous-graphe qu'ils induisent  $G_{p(e)}$ . Voici ci-dessous quelques exemples d'abstractions de graphe illustrées figure 3. On les représente par la propriété  $P(x, e)$  que le sommet  $x$  doit satisfaire dans  $e$ .

1.  $\text{degré} \geq k$ . Un  $k$ -core est un sous-graphe maximal dont tous les sommets ont un degré au moins  $k$ . En appliquant l'opérateur intérieur correspondant à  $\text{degré} \geq k$  à un ensemble de sommets on obtient l'union des  $k$ -cores (parfois aussi appelé le  $k$ -core) du sous-graphe induit par ces sommets. Il s'agit d'une notion classique d'analyse de la structure d'un réseau social et de ses communautés (Seidman, 1983).

2.  $k\text{-club} \geq s$ :  $x$  doit appartenir à un  $k$ -club  $c$  de taille au moins  $s$  dans  $G_e$ . Il s'agit d'un affaiblissement de la notion de clique (Balasundaram *et al.*, 2005) : il existe une chaîne de longueur au plus  $k$  entre toute paire de sommets dans  $G_e$ . Un triangle, ou 3-clique, est un 1-club de taille 3 (figure 3-a). La figure 3-b représente un 2-club de taille 6.

3.  $\text{nearStar}(k, d)$ :  $x$  doit avoir un degré au moins  $k$  ou il doit y avoir un chemin de longueur au plus  $d$  entre  $x$  et un sommet  $y$  de degré au moins  $k$ . Cette abstraction est utile lorsqu'on veut préserver les *hubs*, c'est-à-dire les sommets de haut degré et leurs voisins (Barabási, Albert, 1999) (voir figure 3-c).

4.  $cc \geq s$ :  $x$  doit appartenir à une composante connexe de taille au moins  $s$  dans  $G_e$  (voir figure 3-d).

On peut combiner deux abstractions  $A_1$  and  $A_2$  de deux manières pour construire une nouvelle abstraction de graphe  $A$ . On peut, par exemple, définir un sous-graphe abstrait dont les sommets ont à la fois un degré plus grand que 10 et appartiennent à une composante connexe de taille au moins 30. On peut au contraire définir un sous-graphe abstrait dans lequel les sommets satisfont au moins une des deux propriétés caractéristiques. Nous pourrions ainsi considérer les sommets qui ont un degré plus grand que 10 ou sont liés à un sommet de degré au moins 50. Dans le premier cas on obtient une abstraction plus forte, c'est-à-dire  $A$  est inclus dans  $A_1$  et dans  $A_2$ , dans l'autre une abstraction plus faible, c'est-à-dire  $A_1$  et  $A_2$  sont inclus dans  $A$ .

PROPOSITION 8. — Soit  $P_1$  et  $P_2$  deux propriétés caractéristiques et soient  $P_1 \wedge P_2$

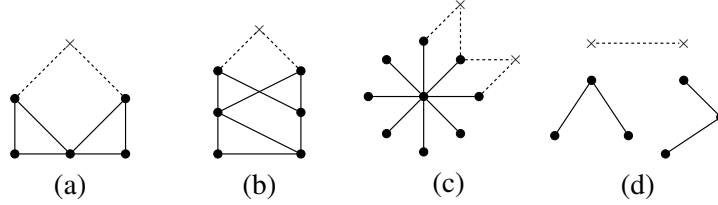


Figure 3. Abstractions de graphe. Dans chaque dessin les points et lignes pleines représentent un sous-graphe abstrait, les croix et les tiretés représentent la partie non abstraite. a)  $x$  appartient à un triangle, (b)  $x$  appartient à un 2-club de taille 6, (c)  $x$  a un degré au moins 6 ou est lié à un sommet de degré au moins 6. (d)  $x$  appartient à une composante connexe de taille au moins 3

et  $P_1 \vee P_2$  définies comme suit :

- $P_1 \wedge P_2(x, e) = P_1(x, e) \wedge P_2(x, e)$
- $P_1 \vee P_2(x, e) = P_1(x, e) \vee P_2(x, e)$

$P_1 \wedge P_2$  et  $P_1 \vee P_2$  sont deux propriétés caractéristiques.

### 3.3. Mesures d'intérêt des motifs et implications abstraites

Nous nous intéressons maintenant aux mesures d'intérêt de ces motifs abstraits et implications abstraites, en fonction de leur relation à la topologie des sous-graphes abstraits impliqués. Ces mesures ont été présentées dans (Soldano *et al.*, 2015a). Nous allons d'abord généraliser la *corrélation structurelle* introduite par A. Silva et co-auteurs (Silva *et al.*, 2012), pour mesurer la proportion d'un sous-graphe constituée de quasi-cliques :

DÉFINITION 9. — Soit un motif  $q$  et une abstraction  $A \subseteq 2^O$ , la spécificité de  $q$  relativement à  $A$  est :

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

Par exemple, si on considère une abstraction de triangles, lorsque  $s_A(q)$  est proche de 1, le sous-graphe de  $q$  est essentiellement constitué de triangles. Au contraire si  $s_A(q)$  est proche de 0, alors le motif  $q$  apparaît dans des sommets relativement isolés. Nous pouvons ainsi relier un motif à la manière dont il respecte la connectivité du graphe, selon une certaine propriété caractéristique.

EXEMPLE 10. — La figure 4 représente un graphe dont les sommets sont décrits par un sous-ensemble d'items. Nous avons alors :

- $\text{ext}(a) = e = \{1, 2, 3, 4, 5, 7\}$  induit le sous-graphe  $G(e)$  (bleu+noir).
- $\text{ext}_A(a) = \{1, 2, 3\}$  car 4, 5, 7 n'appartiennent à aucun triangle dans  $G(e)$ .
- $\text{int} \circ \text{ext}_A(a) = ab \cap ab \cap ab = ab$  est un motif fermé abstrait
- $s_A(a) = 1/2, s_A(ab) = 3/4$

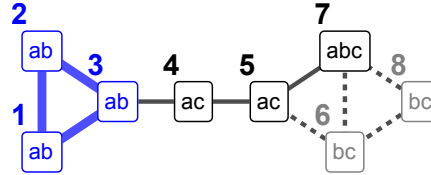


Figure 4. Le sous-graphe du motif  $a$  est en trait plein, le sous-graphe abstrait selon l'abstraction de triangle est en bleu et en gras. Le motif fermé abstrait associé est  $ab$

□

Nous nous intéressons aussi à l'informativité des implications abstraites, c'est-à-dire à ce qui est apporté de nouveau par l'abstraction. Pour cela nous considérons des implications abstraites dont les parties gauche et droite ont même extension mais des extensions abstraites différentes, comme dans la base min-max abstraite définie ci-dessus. Intuitivement, lorsque les parties gauche et droite d'une implication abstraite ont aussi même extension initiale, l'implication n'est pas informative. Nous mesurons simplement l'*informativité* d'une implication abstraite par la probabilité, au niveau non abstrait, de ne pas avoir la conclusion sachant la prémisse :

DÉFINITION 11. — Soit une abstraction  $A$  de  $2^O$ , l'informativité de l'implication abstraite valide  $r : \Box q \rightarrow \Box qw$  est définie par

$$I_A(r) = 1 - \frac{|\text{ext}(qw)|}{|\text{ext}(q)|}$$

Intuitivement l'informativité mesure ce que nous avons découvert lorsque nous avons constaté que  $q$  et  $qw$  ont même support abstrait.

EXEMPLE 12. — En reprenant l'exemple 10 illustré figure 4 et l'implication abstraite  $r : \Box a \rightarrow \Box ab$  dont le sens est « si les sommets d'un triangle de  $G$  ont le motif  $a$ , ils ont aussi le motif  $ab$  », on obtient comme informativité de cette implication abstraite :  $I_A(r) = 1 - 1/2 = 0.5$ .

□

### 3.4. Experiences

#### 3.4.1. S50

Des expériences menées sur les deux jeux de données de la section 2 ont été présentées dans (Soldano, Santini, 2014). Sur le jeu de données s50, une abstraction de triangle a abouti à 36 motifs fermés de taille au moins 3 et 15 implications min-max. Ainsi le motif  $\{C1, D123\}$  est présent chez 28 élèves pour un support abstrait de 16, ce qui correspond à une spécificité de 0,57 et donne lieu à une règle abstraite  $\{C1, D123\} \rightarrow T1$  : « les élèves d'un triangle d'amies qui n'ont jamais essayé le cannabis et ont une consommation d'alcool au plus modérée, sont également non fumeuses ». Cependant,

si on réduit l'ensemble d'élèves concernées en ajoutant une activité sportive régulière, S2, il reste 21 élèves mais l'extension abstraite devient vide, la spécificité tombant à 0. Comme on le voit dans la figure 5 ceci est dû au fait qu'ajouter S2 détruit les groupes abstraits d'élèves partageant {C1, D123}.

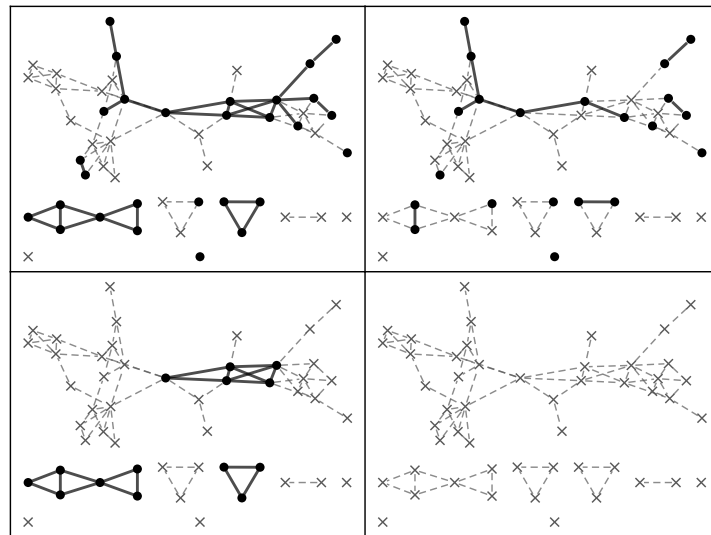


Figure 5. Sous-graphe d'élèves partageant un motif. Sommets et arêtes de chaque sous-graphe sont en noir et en gras. A gauche le sous-graphe (haut) et le sous-graphe abstrait degré  $\geq 2$  (bas) du motif {C1,D123,T1}. A droite, on observe qu'en ajoutant la pratique régulière du sport (S2) on préserve une partie de l'extension (haut) mais on détruit complètement l'extension abstraite (bas)

Pour le même jeu de données nous donnons dans le tableau 1 les 10 motifs fermés abstraits de plus haute spécificité.

Tableau 1. 10 meilleures spécificités de motifs fermés 3-clique abstraits sur s50

Motif fermé abstrait	Support abstrait	Spécificité
"C12"	33	0,79
$\emptyset$	39	0,78
"D345"	22	0,76
"C12", "D345"	15	0,71
"C12", "T1"	24	0,67
"D2345"	30	0,67
"D345", "S2"	14	0,64
"C1"	23	0,64
"C12", "D123", "T1"	18	0,6
"C1", "T1"	20	0,61

### 3.4.2. DBLP

Sur le jeu de données DBLP, nous avons appliqué l'abstraction  $\text{degré} \geq k$  avec des valeurs de  $k$  croissantes. Nous nous focalisons ici sur l'abstraction  $\text{degré} \geq 16$ , ce qui correspond à une abstraction particulièrement forte : dans un sous-graphe abstrait chaque auteur doit être lié à 16 autres par un lien de co-publication. Parmi les motifs fermés abstraits nous trouvons en particulier VLDBJ, ICDE, SIGMOD, VLDB et l'implication abstraite  $\square \text{VLDBJ} \rightarrow \square \text{ICDE, SIGMOD, VLDB}$ . VLDBJ a un support abstrait de 38 pour un support total de 1276 auteurs. Le sous-graphe abstrait correspondant est présenté figure 6.

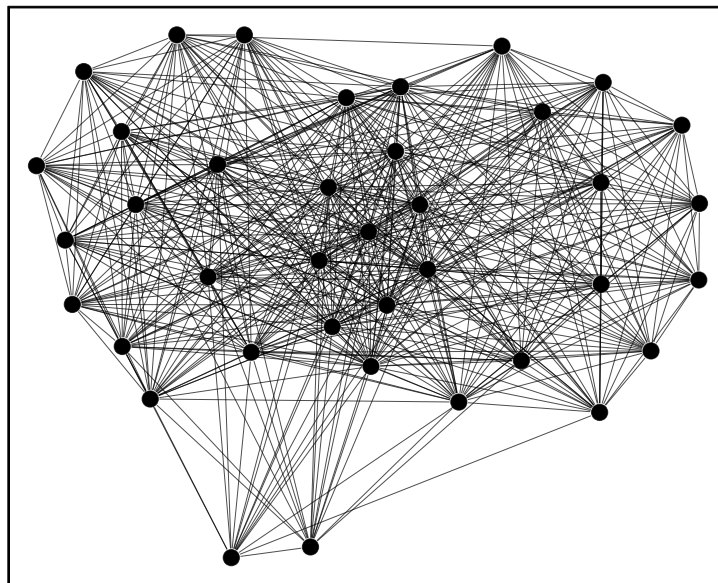


Figure 6. Le sous-graphe  $\text{degré} \geq 16$ -abstrait du motif VLDBJ dans le graphe de co-auteurs DBLP

Lorsqu'on recherche sur DBLP les publications de ces auteurs, on trouve en particulier un article dont le résumé commence par les mots suivants :

A group of senior database researchers gathers every few years to assess the state of database research ...

avec la référence suivante :

- [56]    Serge Abiteboul, Rakesh Agrawal, Philip A. Bernstein, Michael J. Carey, Stefano Ceri, W. Bruce Croft, David J. DeWitt, Michael J. Franklin, Hector Garcia-Molina, Dieter Gawlick, Jim Gray, Laura M. Haas, Alon Y. Halevy, Joseph M. Hellerstein, Yannis E. Ioannidis, Martin L. Kersten, Michael J. Pazzani, Michael Lesk, David Maier, Jeffrey F. Naughton, Hans-Jörg Schek, Timos K. Sellis, Avi Silberschatz, Michael Stonebraker, Richard T. Snodgrass, Jeffrey D. Ullman, Gerhard Weikum, Jennifer Widom, Stanley B. Zdonik: **The Lowell database research self-assessment**. Commun. ACM 48(5): 111-118 (2005)

C'est précisément le rôle de la fouille de données de distinguer des motifs cachés dans de grands jeux de données et de les interpréter. L'explication de l'observation d'un sous-graphe si dense semble donc assez directe, mais, si les chercheurs du domaine sont évidemment conscients de ces rencontres, ce n'est pas forcément le cas de ceux qui fouillent ce jeu de données.

Lorsqu'on applique une abstraction plus faible, précisément ici *degré*  $\geq 4$ , nous observons plus de motifs abstraits parfois constitués de plusieurs composantes connexes. Ainsi, la figure 1 représente le sous-graphe abstrait du motif DMKD, IDArev. Le sous-graphe est constitué de deux composantes connexes (en rouge), celle figurant à droite est constituée de 10 sommets. Nous nous posons alors la question suivante : il y a-t-il un motif plus spécifique que DMKD, IDArev qui serait partagé par les sommets de cette composante connexe ? Pour répondre à ce type de question il faut analyser le graphe à un niveau plus local, ce qui est le propos de la section suivante.

#### 4. Motifs fermés locaux

Dans (Soldano, 2015b) nous avons introduit la notion de localité en analyse formelle de concept, avec pour objectif de réduire le langage de motifs. Les notions sont approfondies dans (Soldano, 2015a) de manière à permettre de réduire le domaine extensionnel, généralisant ainsi la notion d'abstraction extensionnelle. Nous exploitons et expérimentons ces notions pour la fouille de graphes attribués dans (Soldano *et al.*, 2015b). Dans ce qui suit nous exposons ces derniers travaux en donnant les définitions et résultats nécessaires.

Nous devons d'abord définir deux structures d'ordre partiel. Les *pré-confluences*, qui généralisent les treillis, et les *confluences* de treillis qui généralisent les abstractions (Soldano, Santini, 2014 ; Soldano, 2015a). Les confluences d'un treillis des parties, en particulier, sont proches, mais différentes, des familles confluentes définies par M. Boley et co-auteurs (Boley *et al.*, 2010).

Dans ce qui suit nous noterons  $E^x$  la section finissante  $\{y \in E | y \geq x\}$  d'un ensemble ordonné  $E$ ,  $E_x$  la section commençante  $\{y \in E | y \leq x\}$ , et  $\min(E)$  l'ensemble des éléments minimaux de  $E$ . Comme dans le reste de cet article, tous les ensembles ordonnés sont de cardinal fini.

Nous définissons d'abord la notion de pré-confluence :



DÉFINITION 13. —  $F$  est une pré-confluence si et seulement si pour tout  $m \in \min(F)$ ,  $F^m = \{x \in F \mid x \geq m\}$  est un treillis.

Un treillis fini est donc une pré-confluence ayant un élément minimum. Une pré-confluence a plusieurs opérateurs de borne inférieure  $\wedge_m$ , mais les opérateurs de borne supérieure  $\vee_m$  coïncident :

LEMME 14. — Pour tout  $x, y \in F^m$  la borne supérieure ne dépend que de  $x$  et de  $y$  mais pas de  $m$  :

$$- x \vee_F y \text{ est le plus petit élément de } F^x \cap F^y$$

Un cas important est celui où  $F$  est inclus dans un treillis hôte dont il conserve l'opérateur de borne supérieure :

DÉFINITION 15. — Soit  $T$  un treillis et  $F \subseteq T$  une pré-confluence telle que  $\vee_F = \vee_T$ ,  $F$  est appelée une confluence de  $T$ .

Une abstraction de  $T$  est donc une confluence de  $T$  ayant  $\perp_T$  comme minimum.

PROPOSITION 16. — Soit un treillis  $X$ ,  $F \subseteq X$  est une confluence de  $X$  si et seulement si pour tout  $x, y \in F^m$  avec  $m \in \min(F)$ ,  $x \vee y$  appartient à  $F$ .

Cette propriété signifie que toute partie  $F^m$  est une abstraction du sous-treillis de  $X$  d'élément minimum  $m$ , et est donc associé à un opérateur intérieur (voir Proposition 3), ce qui nous permettra, comme dans le cas des motifs fermés abstraits, de définir des opérateurs de fermeture :

PROPOSITION 17. — Soit  $F$  une confluence de  $X$ , et  $m \in \min(F)$ ,

$$- p_m : X^m \rightarrow X^m, \text{ tel que } p_m(x) = \vee_{q \in F^m \cap X_x} q, \text{ est un opérateur intérieur et } p_m[X^m] = F^m.$$

Nous nous intéressons ici aux confluences de  $X = 2^O$  (Soldano, 2015a) qui généralisent les abstractions et, en particulier, les abstractions de graphe. Soit alors  $x$  un sous-ensemble de sommets contenant un certain  $m$  de  $F$ , alors  $p_m(x)$  renvoie le plus grand sous-ensemble de  $x$  dans  $F$  qui contient  $m$ . Dans l'exemple ci-dessous nous définissons la plus simple des confluences de graphe en ne considérant que les sous-ensembles de sommets connexes, au sens où ils induisent un sous-graphe connexe du graphe  $G = (O, E)$ .

EXEMPLE 18. — Soit le graphe  $G = (O, E)$ , représenté au bas de la figure 7, avec  $O = \{1, 2, 3, 4\}$ . Soit  $F \subseteq 2^O$  la famille des sous-ensembles de sommets connexes. L'ensemble de ses éléments minimaux est  $\min(F) = \{\{1\}, \{2\}, \{3\}, \{4\}\}$ . L'union de sous-ensembles de sommets connexes ayant un sommet commun  $s$ , donc tous deux plus grands que  $\{s\}$ , est également connexe et donc  $F$  est une confluence de  $2^O$ . Par abus de langage nous écrirons par la suite,  $p_s$  et  $F^s$  pour  $p_{\{s\}}$  et  $F^{\{s\}}$ . L'opérateur intérieur  $p_s$  projette ainsi tout sous-ensemble de sommets  $e$  contenant le sommet  $s$  dans la composante connexe contenant  $s$  du sous-graphe induit  $G_e$ . La section finissante  $F^s$  est alors l'ensemble des sous-ensembles connexes contenant  $s$ . Cette confluence est le cas le plus simple de ce que nous appellerons par la suite des *cc-confluences*.

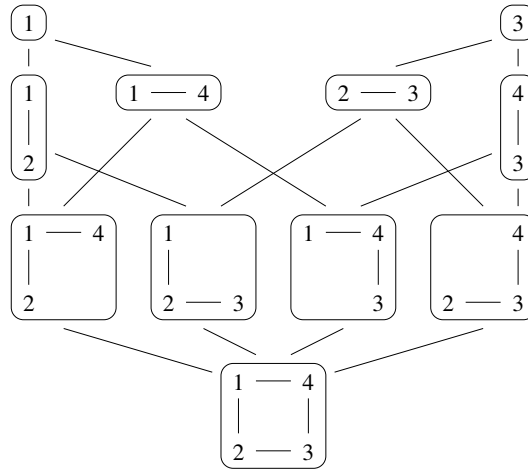


Figure 7. Un graphe carré (en bas) et le diagramme de la confluence  $F^{1+3}$  des sous-ensembles de sommets connexes contenant les sommets 1 ou 3

La partie  $F^{1+3} = F^1 \cup F^3$  représente l'ensemble des sous-ensembles de sommets connexes contenant les sommets 1 ou 3 et est aussi une confluence. La figure 7 présente le diagramme de  $F^{1+3}$ .  $\square$

L'extension  $e$  d'un motif  $q$  sera alors ainsi projetée par les opérateurs intérieurs  $p_i$  sur les extensions locales  $e_i$ , correspondant dans l'exemple ci-dessus aux composantes connexes du sous-graphe induit  $G_e$ . Ces opérateurs intérieurs conduisent à des opérateurs de fermetures locaux (Soldano, 2015a):

PROPOSITION 19. — Soient  $F$  une confluence de  $X = 2^O$ ,  $m$  un élément minimal de  $F$  et  $L_{\text{int}(m)}$  la section commençante du treillis de motifs  $L$  dont les éléments  $q$  sont tels que  $q \geq \text{int}(m)$ , on a alors:

$$f_m = \text{int} \circ p_m \circ \text{ext} \text{ est un opérateur de fermeture sur } L_{\text{int}(m)}.$$

Dans la confluence de graphe de l'exemple ci-dessus, soit  $e = \text{ext}(q)$ ,  $p_s(e)$  est la composante connexe de  $G_e$  contenant  $s$ . On a  $p_s(e) = p_v(e)$  pour tout sommet de cette composante.  $f_s(q)$  est alors le motif fermé local associé à tout sommet de cette composante, c'est-à-dire le motif le plus spécifique partagé par les éléments de cette composante.

L'ensemble de ces extensions locales est elle-même ordonnée en une pré-confluence :

THÉORÈME 20. — La fonction  $h : F \rightarrow F : h(e) = p_m \circ \text{ext} \circ \text{int}(e)$  pour  $m \leq e$

est un opérateur de fermeture sur  $F$  et  $E = h[F]$  est une pré-confluence.

$h(e)$  est l'extension locale de  $\text{int}(e)$  qui contient tout élément minimal  $m \leq e$ .  $h[F]$  est une pré-confluence isomorphe à l'ensemble  $P$  des concepts locaux définis comme suit :

DÉFINITION 21. — *L'ensemble des paires  $P = \{(e, l) \mid e = p_m \circ \text{ext}(l), l = \text{int}(e), m \leq e\}$  est appelé une pre-confluence de concepts.*

Pour résumer, nous avons donc défini des concepts locaux comme des paires (extension locale, motif fermé local) et montré qu'ils étaient organisés en une structure ordonnée, ayant en général plusieurs éléments minimaux, chacun étant le minimum d'un treillis, généralisant ainsi la définition de treillis de concepts. Dans le cas de la confluence de graphe illustrée ci-dessus, les extensions locales sont simplement les composantes connexes des sous-graphes définis par les motifs. Nous définissons maintenant plus précisément la notion de cc-confluence.

#### 4.1. Cc-confluences

Nous remarquons d'abord que l'intersection de plusieurs confluences est encore une confluence :

PROPOSITION 22. — *Soit  $F_1$  et  $F_2$  des confluences d'un treillis  $X$ , alors  $F_1 \cap F_2$  est une confluence de  $X$ .*

Et les abstractions d'un treillis  $X$  étant des confluences de  $X$  contenant le minimum de  $X$  comme unique élément minimal, en faisant l'intersection d'une confluence et d'une abstraction nous obtenons encore une confluence.

Etant donné un graphe  $G = (O, E)$ , nous pouvons alors construire un ensemble de confluences, que nous appellerons les cc-confluences de  $G$ . Une cc-confluence est obtenue à partir de :

- La confluence  $F$  des sous-ensembles de sommets connexes de  $G$ .
- Une abstraction de graphe  $A$  de  $G$

$F_A = F \cap A$  est alors la cc-confluence de  $G$  associée à  $A$ .

Si on reprend l'exemple 18,  $F^{1+3}$  est obtenue en considérant l'abstraction  $A$  des sous-ensembles de sommets contenant 1 ou 3<sup>8</sup>.  $F^{1+3}$  est donc une cc-confluence de  $G$ . Si on considère l'abstraction  $k$ -clique, on obtient la cc-confluence des sous-ensembles de sommets connexes dont le graphe induit est constitué de  $k$ -cliques.

Une remarque importante concernant les cc-confluences est que lorsqu'on définit les opérateurs de fermeture locaux on peut se contenter de ne considérer que les sommets, plutôt que tous les éléments minimaux de la cc-confluence. Cela vient du fait que tous les éléments minimaux  $m$  d'une confluence  $F$  contenus dans  $e$  conduisent à une projection  $p_m(e)$  potentiellement différente et à un motif fermé local. Cependant dans une cc-confluence, un sous-ensemble de sommets  $e$  est projeté sur les composantes connexes du sous-graphe abstrait, on n'aura donc besoin que de la composante connexe du sous-graphe abstrait associé à chaque sommet  $s$  pour calculer les extensions locales et fermés locaux, ce qui amène à la notation  $p_s(e) = p_m(e)$  bien que  $s$

8. La propriété caractéristique de cette abstraction est :  $P(x, e)$  est vraie si  $x \in e$  et  $e$  contient 1 ou 3.

ne soit pas un élément de  $F$  mais un sommet. Ceci est important aussi bien du point de vue combinatoire que du point de vue du calcul : ainsi si on considère les sous-ensembles de sommets connexes de taille au moins  $l$  dans un graphe de  $n$  sommets, le nombre d'éléments minimaux peut-être beaucoup plus grand que  $n$ , il n'y aura pourtant pas plus de  $n$  fermés locaux à considérer pour un motif donné.

**4.2. Implications locales dans une cc-confluence**

L'inclusion des extensions locales conduit à la définition d'implications locales  $\square_m q \rightarrow \square_m w$  où  $m$  est un élément minimal de  $F_A$ , que l'on peut également écrire dans le cas des cc-confluences  $\square_s q \rightarrow \square_s w$ , où  $s$  est un sommet de  $m$ , Remarquons que, comme dans le cas des implications abstraites, les opérateurs intérieurs étant monotones, si l'implication abstraite  $\square q \rightarrow \square w$  est valide et  $m \subseteq \text{ext}_A(w)$  alors  $\square_s q \rightarrow \square_s w$  est aussi valide : l'implication locale peut-être déduite de l'implication abstraite, le contraire n'étant pas vrai. Intuitivement, cela traduit le fait que si on considère que les sommets d'un groupe abstrait ayant  $q$  ont aussi  $w$  alors cela reste vrai lorsque ces sommets sont dans une même composante connexe. Ainsi, si en se limitant aux  $k$ -cliques l'implication est valide, elle le reste dans chaque composante connexe du sous-graphe de  $q$ , abstrait selon les  $k$ -cliques, Nous donnons ci-dessous un exemple d'implication locale dans une cc-confluence de triangles.

EXEMPLE 23. — Considérons le graphe attribué de la figure 8. La cc-confluence associée à l'abstraction 3-clique a comme éléments minimaux  $\{123, 567, 678\}$ . Elle est constituée de 4 éléments :  $F_A = \{123, 567, 678, 5678\}$ . L'extension du motif  $b$  coïncide avec son extension abstraite 123678. Le fermé abstrait est également  $b$  : ces sommets ont seulement  $b$  en commun. Cependant le sous-graphe abstrait induit est constitué de deux composantes connexes 123 et 678 et la seconde a comme motif fermé local  $bc$ . On a ainsi une implication locale à 678 qui est valide :

$$- \square_{678}^A b \rightarrow \square_{678}^A bc$$

Dans une cc-confluence on peut interpréter la règle relativement aux sommets de  $s$  : un triangle dans la même composante connexe que 6, lorsqu'on se restreint au sous-graphe de  $b$ , a aussi  $c$ , ce que l'on écrit  $\square_6^A b \rightarrow \square_6^A bc$  Nous dirons plus simplement « un triangle près de 6 qui a  $b$  a aussi  $c$  ». Ici l'expression « près de » est relative à la prémisse de l'implication.

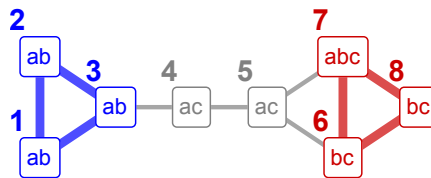


Figure 8. Le sous-graphe abstrait selon les triangles du motif  $b$  et ses deux composantes connexes en gras. Le sous-graphe en bleu à gauche est aussi le sous-graphe abstrait (ici connexe) du motif  $a$

□

Comme pour les implications abstraites, on cherchera à définir une base d'implications locales  $B$  associée à une  $cc$ -confluence, à partir de laquelle on pourra engendrer toutes les implications locales valides sur  $F_A$ . Pour cela nous devons maintenant considérer une base  $B_m$  chaque élément minimal de  $F_A$ , ou de manière équivalente pour chaque sommet. Nous obtiendrons la base complète  $B = \cup B_m$  en faisant l'union de ces bases.

Considérons un motif fermé abstrait  $c$  dont l'extension abstraite contient un groupe abstrait  $m$  de la  $cc$ -confluence  $F_A$  et soit  $l = f_m(c)$  le motif fermé associé, c'est-à-dire le motif le plus spécifique de la composante connexe contenant  $m$ . L'implication  $\square_m c \rightarrow \square_m l$  est alors valide. Parmi ces implications nous retiendrons dans  $B_m$  celles qui sont à la fois *informatives* ( $l \neq c$ ) et *non-redondantes* au sens où il n'existe pas d'autre implication  $\square_m c' \rightarrow \square_m l$  avec  $c'$  plus général que  $c$  dans l'ensemble de ces implications. De  $B_m$  nous pouvons déduire toutes les implications locales associées à  $m$  en appliquant les mêmes axiomes logiques utilisés pour définir la *base min-max* dans les cadres classiques ou abstraits. La base  $B = \cup B_m$  représente la connaissance locale dérivée de la réduction de l'espace extensionnel abstrait  $A$  à la  $cc$ -confluence  $F_A$  :

DÉFINITION 24 (Base locale min-max). —

$M_{F_A} = \{\square_m^A c \rightarrow \square_m^A l \mid \text{où } c \text{ est un fermé abstrait, } l \text{ est un fermé local, } c \neq l, \text{ext}_m^A(c) = \text{ext}_m^A(l), \text{ et pour tout } c' \subset c \text{ on a } \text{ext}_m^A(c') \neq \text{ext}_m^A(c)\}$

### 4.3. Mesures d'intérêt des motifs et implications locaux

Comme dans le cas abstrait nous mesurons ici les motifs fermés abstraits ainsi que les implications locales (Soldano *et al.*, 2015a). La démarche est analogue à celle utilisée dans le cas abstrait. Un motif fermé local étant associé à une extension locale, nous allons considérer le rapport entre l'extension locale et l'extension abstraite. S'il est élevé le sous-graphe induit est essentiellement constitué de cette extension locale, autrement dit le motif n'est pas dispersé entre plusieurs régions denses mais *spécifique* à la composante connexe considérée.

DÉFINITION 25. — Soit un motif  $q$ ,  $F$  une  $cc$ -confluence et  $m \in F$  tel que  $m \subseteq \text{ext}_A(q)$ , la *spécificité* de  $q$  près de  $m$  est définie par :

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

Comme dans le cas abstrait nous mesurons l'informativité d'une implication locale relativement à l'implication globale (ici abstraite) correspondante. L'idée ici est que dans une implication locale valide, les motifs gauche et droit ont même extension locale alors que leurs extensions abstraites sont différentes. On estimera donc la probabilité, au niveau abstrait, de ne pas avoir la partie droite lorsque la partie gauche est vraie, alors que cette probabilité est estimée à 0 au niveau local :

DÉFINITION 26 (*Informativité* d'une implication locale  $r : \square_m^A q \rightarrow \square_m^A qw$ ). —

$$I_F(r) = 1 - \frac{|\text{ext}_A(qw)|}{|\text{ext}_A(q)|}$$

Lorsque les parties gauche et droite ont même extension abstraite, on n'apprend rien de cette implication locale : elle était déjà valide au niveau abstrait. Si l'informativité est proche de 1 cela signifie que la probabilité d'avoir la partie droite au niveau abstrait est faible : le fait de déduire  $w$  de  $q$  est spécifique de la région du graphe où se trouve  $m$ .

EXEMPLE 27. — Nous continuons sur l'exemple 23 illustré figure 8. La spécificité locale à 678 du motif fermé local  $bc$  est  $s_F(bc, 678) = 3 \div 3 = 1$  : ce motif n'apparaît dans l'extension abstraite qu'en 678. L'informativité de l'implication  $\square_{678}^A b \rightarrow \square_{678}^A bc$  est de  $1 - (2 \div 6) = 0,5$  : l'extension abstraite de  $b$  est de 6 mais seulement 3 sommets ont le motif fermé local  $bc$ . En ce qui concerne le motif fermé local  $ab$ , il est aussi de spécificité 1 : on ne trouve  $ab$  que dans la composante connexe 123 de son sous-graphe abstrait, et donne lieu à une autre implication locale  $\square_{123}^A b \rightarrow \square_{123}^A ab$  d'informativité aussi égale à 0,5. On a ainsi des connaissances locales nouvelles associées au même motif fermé abstrait  $b$  mais portant sur des régions différentes du graphe.  $\square$

Nous revenons à la question posée lors de nos expériences sur le jeu de données DBLP de la section 3.4 et concernant le motif fermé abstrait DMKD, IDArev, relativement à l'abstraction  $\text{degré} \geq 4$ , dont l'extension abstraite est représentée en rouge figure 1. Si on considère la composante connexe sur la gauche de la figure, nous obtenons l'implication locale  $\text{DMKD, IDArev} \rightarrow_{268924} \text{DMgroup}$  spécifiant que dans la composante connexe du sous-graphe abstrait de DMKD, IDArev à laquelle appartient l'auteur 268924<sup>9</sup>, les auteurs ont aussi publié dans des conférences appartenant au groupe DMgroup.

Si on considère le jeux de données s50 dont le graphe est représenté figure 2, il est clair que les relations d'amitié sont organisées préférentiellement en triangles et toute abstraction plus forte, par exemple  $\text{degré} \geq 3$ , perdra une grande partie du graphe. Cependant comme mentionné dans l'introduction, nous observons, pour le motif vide, une grande composante connexe qu'il serait intéressant de décomposer en communautés chevauchantes (en couleur sur la figure). Dans ce qui suit nous montrons comment la méthodologie des extensions locales peut s'appliquer pour découvrir de telles sous-communautés dans un graphe attribué, en définissant les *cc*-confluences sur un graphe dérivé du graphe original.

9. Rappelons que dans une *cc*-confluence, pour un motif donné, on peut se référer pour une extension locale à un sommet  $s$  comme à un groupe connexe  $m$  de  $F_A$  qui le contient : une extension locale est une composante connexe et si elle contient  $s$  elle contient tout  $m$  dans lequel le motif apparaît.

## 5. Cc- confluences dans un graphe dérivé

Dans ce qui suit nous considérons une famille  $T \subseteq 2^O$  de sous-ensembles de sommets et un graphe  $G_T = (T, E_T)$  dérivé de  $G$ . La confluence de graphe  $F \subseteq 2^T$  est alors le nouvel espace extensionnel permettant de définir les motifs fermés locaux. Les extensions locales dans  $F$  seront retranscrites dans  $2^O$ .

Soit  $u : 2^T \rightarrow 2^O$  telle que  $u(e_T) = \cup_{t \in e_T} t$ .  $u(e_T)$  est appelé l'*aplatissement* de  $e_T$ . Nous considérons alors les deux fonctions  $\text{ext}_T$  et  $\text{int}_T$  définies comme suit :

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

$\text{ext}_T(q)$  est l'extension de  $q$  dans  $T$  quand on considère que  $q$  a une occurrence dans  $t$  lorsque  $q$  a une occurrence dans tous les sommets de  $t$ . Inversement  $\text{int}_T(e_T)$  est le motif le plus spécifique de  $L$  dont l'extension dans  $T$  contient  $e_T$ , c'est-à-dire dont l'extension dans  $O$  contient, en tant que sous-ensembles, les sommets dans  $G_T$  de  $e_T$ .

Considérons en particulier la famille  $T$  des  $k$ -cliques de  $G$  et posons que  $(t_1, t_2)$  forment une arête dans  $G_T$  lorsque  $t_1$  et  $t_2$  partagent  $k - 1$  sommets dans  $G$ . Une  $k$ -communauté de  $G$ , telle qu'elle est définie dans (Palla *et al.*, 2005), est alors l'aplatissement (au sens défini ci-dessus) d'une composante connexe de  $G_T$ . Nous allons nous intéresser aux sous-communautés induites par les extensions des motifs de  $L$ , autrement dit aux  $k$ -communautés des sous-graphes induits par ces motifs. Les motifs fermés locaux sont les motifs les plus spécifiques partagés au sein de ces  $k$ -sous-communautés. Nous obtenons ainsi une base d'implications associées à ces  $k$ -sous-communautés.

### 5.1. Expériences sur les 3-sous-communautés de jeu de données s50

Nous avons construit le graphe dérivé  $G_T$ , où  $T$  est l'ensemble des triangles du graphe d'amitiés entre élèves représenté figure 2. Nous représentons figures 9 et 10 l'ordre partiel des 3-sous-communautés de taille au moins 4<sup>10</sup>. Les plus petites 3-sous-communautés sont en bas dans les deux figures. Chaque 3-sous-communauté et son motif fermé local peuvent être associés à plusieurs implications locales. C'est ce qui se produit pour la 3-sous-communauté représentée en bas à droite de la figure 9 où les deux implications locales sont associées à des fermés abstraits différents (relativement à l'abstraction de triangle). Chaque cadre montre en rouge et gras la 3-sous-communauté, et en rouge ou noir et en trait plein l'extension abstraite associée à la partie gauche de l'implication locale. Dans la figure 10 nous avons une 3-communauté maximale unique en haut, et en dessous la hiérarchie de ses 3-sous-communautés.

10. Nous éliminons ainsi les sous-graphes connexes de  $G_T$  dont la taille, en nombre de sommets dans le graphe d'origine, est inférieure à 4.

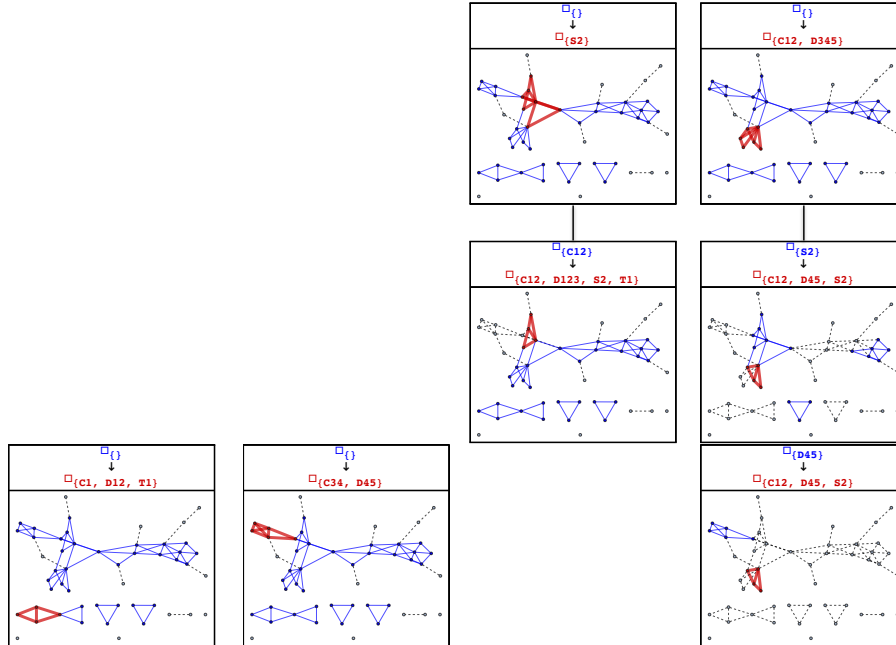


Figure 9. L'ordre partiel selon l'inclusion des 3-sous-communautés du graphe d'amitié S50 de taille au moins 4 (partie-I). Les 3-sous-communautés sont en rouge et en gras et ordonnées de haut en bas, des plus grandes aux plus petites. Les sous-graphes abstraits sont en trait plein. En bas à droite, une sous-communauté associée à deux règles locales

## 6. Implementation

Dans nos expériences sur les abstractions de graphes nous avons d'abord utilisé le logiciel CORON(Szathmary, Napoli, 2005) pour calculer les motifs fermés fréquents, puis appliqué un ensemble de fonctions PYTHON en post-traitement<sup>11</sup>. Le post-traitement permet de calculer pour chaque motif fermé fréquent  $c$  le sous-graphe induit par son extension abstraite, calculer les motifs fermés abstraits associés et la base min-max abstraite.

En ce qui concerne les fermés locaux et les bases d'implications locales, nous avons proposé et implémenté un algorithme efficace, utilisant une stratégie de partitionnement similaire à celle proposée dans (Boley *et al.*, 2010) et implémentée par B. Negrevergne et co-auteurs dans le programme PARAMINER (Negrevergne *et al.*, 2013). Une première version de l'algorithme, nommée ParaminerLC, est décrite et

11. Le logiciel correspondant est à l'adresse <https://lipn.univ-paris13.fr/~santini/>.



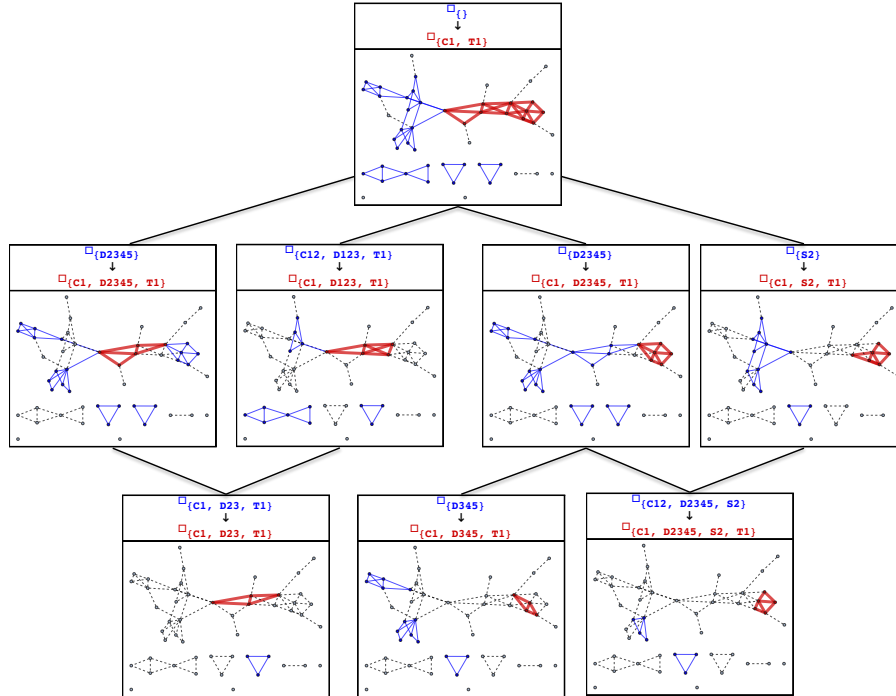


Figure 10. Les 3-sous-communautés du graphe d'amitié de taille au moins 4 (partie-II). Cette partie a un plus grand élément (en haut) qui se décompose, en descendant, en sous-communautés chevauchantes

expérimentée dans (Soldano *et al.*, 2015b). Elle permet le calcul efficace des fermés locaux et implications locales associées à l'ensemble des 3-sous-communautés<sup>12</sup>. La base des règles locales mini-max est calculée a posteriori. La figure 11 résulte de l'exécution de ParaminerLC sur le jeu de données DBLP. Elle présente le temps CPU  $t$  (à gauche) ainsi que le temps CPU par motif fermé abstrait  $\frac{t}{\#c}$ , en fonction du seuil de fréquence minimal, exprimé en nombre de sommets requis dans une 3-sous-communauté. On constate sur la courbe du temps CPU par motif fréquent que celui-ci décroît avec le seuil de fréquence, montrant l'efficacité de ParaminerLC lorsque il doit produire un grand nombre d'implications locales. Le nombre d'implications varie ici de 425, pour un seuil de fréquence de 200, à 31 067, pour un seuil de fréquence de 10.

Enfin, nous développons actuellement, sur le même schéma algorithmique, *minerLC*, un programme plus versatile, permettant d'avoir en entrée les abstractions de graphe et cc-confluences choisies, pour le calcul des fermés abstraits et locaux, et des implications locales associées, y compris dans le cas des 3-sous-communautés. Ce programme est entouré d'une suite de fonctions PYTHON permettant les pré-

12. ParaminerLC est disponible à l'adresse <https://lipn.univ-paris13.fr/~bouthinon/>.

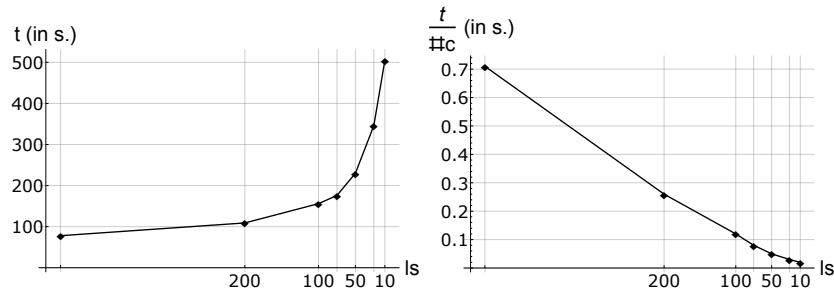


Figure 11. Temps CPU total (gauche) et temps CPU par motif fermé abstrait (droite) en fonction du seuil de fréquence

traitements et post-traitements, et en particulier le calcul des spécificités et informativités abstraites et locales.

## 7. Conclusion

Dans cet article nous proposons une méthodologie de fouille de graphes attribués étendant à ce contexte les notions de motifs fermés et d'ensembles ordonnés de paires (extension, motif fermé) proposées par l'analyse formelle de concept. L'adaptation aux graphes attribués est sous-tendue par l'intuition du rôle joué par les relations entre les objets dans le graphe : en s'intéressant à des parties denses du sous-graphe induit par un motif donné, on rend plus grossière la relations d'équivalence entre motifs de même extension. On diminue ainsi le nombre de paires (extension, motif fermé) et on fait apparaître des implications qui ne sont pas valides, si on considère l'extension indépendamment du graphe, mais qui le deviennent lorsqu'on restreint les extensions en les rendant abstraites (les sommets peu connectés sont éliminés des extensions) ou locales (la validité est alors restreinte à des régions du graphe).

La méthodologie distingue les aspects globaux et locaux des motifs et règles à extraire. Au niveau global il s'agit d'extraire les motifs fermés abstraits. Un tel motif est ici le plus spécifique commun aux sommets lorsque l'on réduit le sous-graphe induit par un motif à son noyau selon une certaine propriété topologique. Au niveau local, on s'intéresse aux motifs les plus spécifiques associés aux composantes connexes d'un sous-graphe réduit du graphe d'origine ou d'un graphe dérivé de celui-ci.

Nous avons décrit les adaptations de ces notions de fermeture, et d'implications, conduisant à des structures d'ordre partielles nouvelles en ce qui concerne les notions locales, proposé des mesures d'intérêt des motifs et implications abstraits et locaux, et illustré sur deux jeux de données la nature des connaissances nouvelles ainsi découvertes. En particulier nous avons montré comment, en jouant sur le niveau d'abstraction, et sur le graphe lui-même, on pouvait explorer de différentes manières un même graphe attribué. Ceci a conduit en particulier à ordonner l'ensemble des  $k$ -sous-communautés associées aux sous-graphes induits par les motifs, et à leur associer des motifs maximaux ainsi que des implications. Un des intérêts de cette formulation

reposant sur les ensembles ordonnés, est que l’algorithmique qui en découle est relativement simple et versatile, les mêmes schémas permettant d’énumérer les motifs et implications avec des points de vue et des niveaux d’abstraction assez différents. La contre-partie de cette flexibilité est que l’on multiplie les choix à faire pour examiner le graphe. On pourra cependant pour cela s’aider a posteriori des mesures d’intérêt décrites ci-dessus, et adapter a priori l’analyse à un point de vue privilégié. Ainsi, dans le cas des réseaux d’amitié ou d’affinités, comme le montre l’exemple du jeu de données S50, le triangle joue un rôle prépondérant. En revanche dans un réseau de co-auteurs on aura des structures denses à plusieurs échelles, ce que peut révéler une analyse en  $k$ -core, ou chercher des auteurs dominants au moyen d’abstractions de type  $\text{nearStar}(k, d)$ , que nous sommes en train d’examiner expérimentalement.

En résumé, nous espérons que cette nouvelle méthodologie permettra d’organiser la fouille de graphes attribués, en suivant ou se démarquant des notions de fermetures abstraites et locales proposées ici. Nous nous proposons de continuer cette exploration dans deux directions, d’abord en adaptant les notions d’abstraction et de localité aux graphes dirigés, ce qui permettra de traiter directement des graphes de citation ou de régulation, par exemple, puis en élargissant ces notions aux réseaux multiplexes, ces derniers étant constitués de plusieurs couches, chacune dédiée à un type de lien particulier, et exprimant ainsi un certain prédicat binaire (Berlingerio *et al.*, 2013).

## Bibliographie

- Balasundaram B., Butenko S., Trukhanov S. (2005). Novel approaches for analyzing biological networks. *Journal of Combinatorial Optimization*, vol. 10, n° 1, p. 23–39.
- Barabási A.-L., Albert R. (1999). Emergence of scaling in random networks. *Science*, vol. 286, n° 5439, p. 509-512.
- Batagelj V., Zaversnik M. (2011). Fast algorithms for determining (generalized) core groups in social networks. *Adv. Data Analysis and Classification*, vol. 5, n° 2, p. 129–145.
- Bechara Prado A., Plantevit M., Robardet C., Boulicaut J.-F. (2013). Mining Graph Topological Patterns: Finding Co-variations among Vertex Descriptors. *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, n° 9, p. 2090–2104.
- Berlingerio M., Coscia M., Giannotti F., Monreale A., Pedreschi D. (2013). Multidimensional networks: foundations of structural analysis. *World Wide Web*, vol. 16, n° 5-6, p. 567–593.
- Boley M., Horváth T., Pogné A., Wrobel S. (2010). Listing closed sets of strongly accessible set systems with applications to data mining. *Theor. Comput. Sci.*, vol. 411, n° 3, p. 691-700.
- Fortunato S. (2010). Community detection in graphs. *Physics Reports*, vol. 486, n° 3–5, p. 75 - 174.
- Galbrun E., Gionis A., Tatti N. (2014). Overlapping community detection in labeled graphs. *Data Min. Knowl. Discov.*, vol. 28, n° 5-6, p. 1586–1610.
- Moguel P.-N., Rigotti C., Gandrillon O. (2012). Finding collections of  $k$ -clique percolated components in attributed graphs. In *Pakdd(2), advances in knowledge discovery and data*

*mining - 16th pacific-asia conference, pakdd 2012, kuala lumpur, malaysia, may 29 - june 1, 2012*, vol. 7302, p. 181-192. Springer.

- Negrevergne B., Termier A., Rousset M.-C., Méhaut J.-F. (2013). Paraminer: a generic pattern mining algorithm for multi-core architectures. *Data Mining and Knowledge Discovery*, vol. 28, n° 3, p. 593-633.
- Palla G., Derenyi I., Farkas I., Vicsek T. (2005, Jun). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, vol. 435, n° 7043, p. 814–818.
- Pasquier N., Taouil R., Bastide Y., Stumme G., Lakhal L. (2005). Generating a condensed representation for association rules. *Journal Intelligent Information Systems (JIIS)*, vol. 24, n° 1, p. 29-60.
- Pernelle N., Rousset M.-C., Soldano H., Ventos V. (2002). Zoom: a nested Galois lattices-based system for conceptual clustering. *J. of Experimental and Theoretical Artificial Intelligence*, vol. 2/3, n° 14, p. 157–187.
- Seidman S. B. (1983). Network structure and minimum degree. *Social Networks*, vol. 5, p. 269–287.
- Silva A., Meira W., Jr., Zaki M. J. (2012, janvier). Mining attribute-structure correlated patterns in large attributed graphs. *Proc. VLDB Endow.*, vol. 5, n° 5, p. 466–477.
- Soldano H. (2011, July). A modal view on abstract learning and reasoning. In M. Genesereth, P. Z. Revesz (Eds.), *Ninth symposium on abstraction, reformulation, and approximation, sara 2011, cardona, catalonia, spain*, p. 99–106. AAAI Press.
- Soldano H. (2015a). Extensional confluences and local closure operators. In J. Baixeries, C. Sacarea, M. Ojeda-Aciego (Eds.), *Formal concept analysis - 13th international conference, ICFCA 2015, nerja, spain, june 23-26, 2015, proceedings*, vol. 9113, p. 128–144. Springer.
- Soldano H. (2015b). Motifs fermés et abstraction : au delà des treillis. *Revue d'Intelligence Artificielle*, vol. 29, n° 3-4, p. 321-348.
- Soldano H., Santini G. (2014). Graph abstraction for closed pattern mining in attributed network. In T. Schaub, G. Friedrich, B. O'Sullivan (Eds.), *European conference in artificial intelligence (ECAI)*, vol. 263, p. 849–854. IOS Press.
- Soldano H., Santini G., Bouthinon D. (2015a, October 21-23). Abstract and local rule learning in attributed networks. In F. Esposito, M.-S. Hacid, O. Pivert, Z. Ras (Eds.), *Foundations of intelligent systems 22nd international symposium, ismis*, vol. 9384, p. 313–323. Lyon, France, Springer.
- Soldano H., Santini G., Bouthinon D. (2015b, November 9-11). Local knowledge discovery in attributed graphs. In A. Esposito (Ed.), *27th IEEE international conference on tools with artificial intelligence, ICTAI, vietri sul mare, italy*, p. 250–257. IEEE Computer Society.
- Soldano H., Ventos V. (2011). Abstract Concept Lattices. In P. Valtchev, R. Jäschke (Eds.), *International conference on formal concept analysis (icfca)*, vol. 6628, p. 235–250. Springer, Heidelberg.
- Szathmary L., Napoli A. (2005). Coron: A framework for levelwise itemset mining algorithms. In B. Ganter, R. Godin, E. M. Nguifo (Eds.), *Third International Conference on Formal Concept Analysis (ICFCA'05), Lens, France, Supplementary Proceedings*, p. 110–113.