

Clustering bipartite graphs in terms of approximate formal concepts and sub-contexts

Bruno Gaume, Emmanuel Navarro, Henri Prade

▶ **To cite this version:**

Bruno Gaume, Emmanuel Navarro, Henri Prade. Clustering bipartite graphs in terms of approximate formal concepts and sub-contexts. International Journal of Computational Intelligence Systems, 2013, 6 (6), pp.1125-1142. 10.1080/18756891.2013.819179 . hal-00992119v2

HAL Id: hal-00992119

https://hal.science/hal-00992119v2

Submitted on 9 Mar 2015

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License



Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 12874

To link to this article : DOI :10.1080/18756891.2013.819179
URL : http://dx.doi.org/10.1080/18756891.2013.819179

To cite this version : Gaume, Bruno and Navarro, Emmanuel and Prade, Henri *Clustering bipartite graphs in terms of approximate formal concepts and sub-contexts*. (2013) International Journal of Computational Intelligence Systems, vol. 6 (n° 6). pp. 1125-1142. ISSN 1875-6891

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Clustering bipartite graphs in terms of approximate formal concepts and sub-contexts

Bruno Gaume [1] Emmanuel Navarro [2] Henri Prade [2]

[1] *CLLE-ERSS, Université de Toulouse II,*
5, allées Antonio Machado; 31058 Toulouse Cedex 9, France
E-mail: gaume@univ-tlse2.fr
[2] *IRIT, Université de Toulouse III,*
118 Route de Narbonne; 31062 Toulouse Cedex 9, France
E-mail: navarro@irit.fr, prade@irit.fr

Abstract

The paper first offers a parallel between two approaches to conceptual clustering, namely formal concept analysis (augmented with the introduction of new operators) and bipartite graph analysis. It is shown that a formal concept (as defined in formal concept analysis) corresponds to the idea of a maximal bi-clique, while sub-contexts, which correspond to independent "conceptual worlds" that can be characterized by means of the new operators introduced, are disconnected sub-graphs in a bipartite graph. The parallel between formal concept analysis and bipartite graph analysis is further exploited by considering "approximation" methods on both sides. It leads to suggest new ideas for providing simplified views of datasets, taking also inspiration from the search for approximate itemsets in data mining (with relaxed requirements), and the detection of communities in hierarchical small worlds.*

Keywords: formal concept analysis (FCA), bipartite graph, small world, clustering, possibility theory.

1. Introduction

The human mind tries to make sense of a complex set of data usually by conceptualizing it by some means. Roughly speaking, it generally amounts to putting labels on subsets of data that are judged to be similar enough. Formal concept analysis[25,24] offers a theoretical setting for defining the notion of a formal concept as a pair made of (i) the set of objects that constitutes the extension of the concept and of (ii) the set of properties shared by these objects and that characterize these objects as a whole. This set of properties defines the intention of the concept. Thus, particular subsets of objects are associated with conjunctions of properties that identify them in a bi-univoque way. This provides a formal basis for data mining algorithms[46].

Formal concept analysis exploits a relation that links objects with properties. Such a relation can be viewed as well as a bi-graph (or bipartite graph), i.e. a graph having two kinds of vertices, and whose links are only between vertices of dif-

*This paper is a fully revised and expanded version of a conference paper[28]. In particular, Sections 4 and 5 are new.

ferent kinds.

Besides, the discovery that real-world complex networks from many different domains (linguistics, biology, sociology, computer science, ...) are sharing some non-trivial characteristics has raised a considerable interest[53,3,42,26]. These networks are indeed sparse, highly clustered, and the average length of shortest paths is rather small with regard to the graph size[53], hence their name of "small worlds". Moreover, most of parameters, and in particular their vertices degree, follow a power-law distribution[4,42], which acknowledges a hierarchical organization. One of the most active fields of this new *network science* concerns the problem of graph clustering[48,23]. This problem is often called "community detection" in the literature due to its application to social networks.

Intuitively speaking, a cluster (or community) corresponds to a group of vertices with a high density of internal links and only a few links with external vertices. Nevertheless there is no universally accepted formal definition of a cluster[23] and making a parallel with formal concept analysis may provide some relevant views for defining graph clusters. Many real-world large networks are bipartite and it has been shown that such networks also share properties similar to the above-mentioned ones[38]. While clustering is usually done on projected graphs, some authors address the problem of community detection directly on bipartite networks[5,39]. Besides, techniques inspired from formal concept analysis have been also used for detecting human communities in social bipartite networks[50].

The purpose of this paper is first to investigate the parallel between formal concept analysis and the graph-based detection of communities in bipartite graphs. In fact, we do not restrict ourselves here to standard formal concept analysis, but we rather consider an enlarged setting with new operators[17,18]. This setting includes the classical Galois connection that is at the basis of the definition of formal concepts, but also another connection that characterizes independent sub-contexts. This is the graph counterpart of this enlarged setting that is discussed here from a bi-partite graph point of view. Moreover, extensions of this setting which allows various forms of approximations of formal concepts and sub-contexts are then paralleled and compared with methods used in bi-graph clustering.

The paper is organized as follows, the basic elements of formal concept analysis are first restated and the other operators are introduced in Section 2. Then, after a short background on graphs, it is shown in Section 3 that a formal concept corresponds to a maximal bi-clique in a bi-graph, while conceptual worlds (i.e., independent sub-contexts), obtained by the second connection, correspond to disconnected sub-parts in the graph. Then different ways of introducing various types of approximation, or gradualness, in formal concept analysis, data mining, or in community detection are reviewed in Section 4, before discussing and illustrating their counterpart in the bi-graph setting by proposing a two step clustering procedure in Section 5.

2. Extended formal concept analysis

Let R be a *binary relation* between a set $\mathbf{O}$ of objects and a set $\mathbf{P}$ of Boolean properties. We note $\mathscr{R} = (\mathbf{O}, \mathbf{P}, R)$ the tuple formed by these objects and properties sets and the binary relation. It is called a *formal context*[25]. The notation $(x,y) \in R$ means that object x has property y. Let $R(x) = \{y \in \mathbf{P} | (x,y) \in R\}$ be the set of properties of object x. Similarly, $R^{-1}(y) = \{x \in \mathbf{O} | (x,y) \in R\}$ is the set of objects having property y.

Formal concept analysis[25] defines two set operators, here denoted $(.)^{\Delta}$, $(.)^{-1\Delta}$, called *intent* and *extent* operators respectively, s.t. $\forall Y \subseteq \mathbf{P}$ and $\forall X \subseteq \mathbf{O}$:

$$X^{\Delta} = \{y \in \mathbf{P} | \forall x \in X, (x,y) \in R\} \quad (1)$$

$$Y^{-1\Delta} = \{x \in \mathbf{O} | \forall y \in Y, (x,y) \in R\} \quad (2)$$

X^{Δ} is the set of properties possessed by all objects in X. $Y^{-1\Delta}$ is the set of objects having all properties in Y. These two operators induce an antitone Galois connection between $2^{\mathbf{O}}$ and $2^{\mathbf{P}}$. This means that the following property holds

$$X \subseteq Y^{-1\Delta} \Leftrightarrow Y \subseteq X^{\Delta}.$$

A pair such that $X^{\Delta} = Y$ and $Y^{-1\Delta} = X$ is called a *formal concept*[25]. X is its extent and Y its intent. In other words, a formal concept is a pair (X,Y) such that X is the set of objects having all properties in Y and Y is the set of properties shared by all objects in X. It can be shown that formal concepts correspond to *maximal pairs* (X,Y) such that

$$X \times Y \subseteq R.$$

A recent parallel between formal concept analysis and possibility theory[17] has led to emphasize the interest of three other remarkable set operators $(.)^{\Pi}$, $(.)^{N}$ and $(.)^{\nabla}$. These three operators and the already defined intent operator can be written as follows, $\forall X \subset \mathbf{O}$:

$$X^{\Pi} = \{y \in \mathbf{P} | R^{-1}(y) \cap X \neq \emptyset\} \quad (3)$$

$$X^{N} = \{y \in \mathbf{P} | R^{-1}(y) \subseteq X\} \quad (4)$$

$$X^{\Delta} = \{y \in \mathbf{P} | R^{-1}(y) \supseteq X\} \quad (5)$$

$$X^{\nabla} = \{y \in \mathbf{P} | R^{-1}(y) \cup X \neq \mathbf{O}\} \quad (6)$$

Note that (5) is equivalent to the definition of operator $(.)^{\Delta}$ in (1). X^{Π} is the set of properties that are possessed by at least *one* object in X. X^{N} is the set of properties such that any object that satisfies one of them is necessarily in X. X^{Δ} is the set of properties shared by all objects in X. X^{∇} is the set of properties that some object outside X misses.

It is usually assumed that the relation R is such that $R^{-1}(y) \neq \emptyset$ and $R^{-1}(y) \neq \mathbf{O}$ ("bi-normalization"), which respectively means that there is no property y that is possessed by no object, or by all objects. It guarantees that $X^{N} \subseteq X^{\Pi}$ and $X^{\Delta} \subseteq X^{\nabla}$ hold, as expected.

Operators $(.)^{-1\Pi}$, $(.)^{-1N}$, $(.)^{-1\Delta}$ and $(.)^{-1\nabla}$ are defined similarly on a set Y of properties by substituting R^{-1} to R and by inverting $\mathbf{O}$ and $\mathbf{P}$. $(Y)^{-1\Pi}$, $(Y)^{-1N}$, $(Y)^{-1\Delta}$ and $(Y)^{-1\nabla}$ are respectively, i) the set of objects having at least one property in Y, ii) the set of objects whose properties are all in Y, iii) the set of objects that have all the properties in Y, and iv) the set of objects that are missing at least one property outside Y. Moreover, we also assume the bi-normalization of R for objects, namely $R(x) \neq \emptyset$ and $R(x) \neq \mathbf{P}$, i.e., no object misses all properties or has all properties.

These new operators lead to consider a new connection[20,16] that corresponds to pairs (X,Y) such that $X^{\Pi} = Y$ and $Y^{-1\Pi} = X$ (or equivalently) such that $X^{N} = Y$ and $Y^{-1N} = X$, while $(.)^{\nabla}$ and $(.)^{\Delta}$ lead to the same remarkable pairs which define formal concepts. But pairs (X,Y) such that $X^{\Pi} = Y$ and $Y^{-1\Pi} = X$ do not define formal concepts, but rather *independent sub-contexts*. Indeed, it has been recently shown[16,20] that pairs (X,Y) of sets exchanged through the new connection operators, are subsets such that

$$(X \times Y) \cup (\overline{X} \times \overline{Y}) \supseteq R,$$

just as formal concepts correspond to maximal pairs (X,Y) such that

$$X \times Y \subseteq R.$$

For instance, the pairs $(\{1,2,3,4\},\{g,h,i\})$ and $(\{5,6,7,8\},\{a,b,c,d,e,f\})$ in Figure 1 are two independent sub-contexts, whereas pairs $(\{1,2,3,4\},\{g,h\})$, $(\{5,6\},\{a,b,c,d,f\})$ and $(\{5,6,7,8\},\{a,c,d\})$ are examples of formal concepts. However, note that in general, it might be the case that an independent sub-context in a binary relation R can still be further decomposed into smaller sub-contexts.

Thus, in the setting of formal concept analysis, by means of two companion connections, two key aspects of the idea of clustering are at work. On the one hand, independent sub-contexts are characterized, and on the other hand inside each sub-context, formal concepts (X,Y) are identified where *each* pair (x,y) such that $x \in X, y \in Y$ are in relation (while *no* pair (x,y) such that $x \in \overline{X'}, y \in Y'$ or $x \in X', y \in \overline{Y'}$ are in relation if (X',Y') and $(\overline{X'},\overline{Y'})$ two independent subcontexts). In particular, two formal concepts belonging to two different sub-contexts are clearly well-separated. A recent discussion paper[19] has indeed emphasized a parallel between the characterizations of formal concepts and sub-contexts in formal concept analysis and the characterization of fuzzy clusters in the setting of the extensional fuzzy set approach[34]. The relation with clustering is made still clearer in the next section by providing a bipartite graph reading of formal concept analysis.

	1	2	3	4	5	6	7	8
a					×	×	×	×
b					×	×		
c					×	×	×	×
d					×	×	×	×
e							×	
f					×	×		×
g	×	×	×	×				
h	×	×	×	×				
i				×				

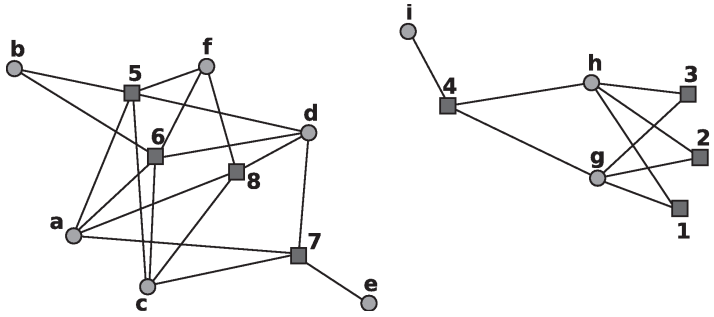


Figure 1: A formal context R and the corresponding bipartite graph.

3. Graph reading of formal concept analysis

Let us start by restating some graph theory definitions. A *graph* is a pair of sets $\mathscr{G} = (V,E)$, where V is a set of *vertices* and E a set of *edges*. In the paper only *undirected graphs* will be considered, it means that edges are unordered pairs of vertices. A graph is *bipartite* if the vertex set V can be split into two sets A and B such that there is no edge between vertices of the same set (in other words for every edge $\{u,v\}$ either $u \in A$ and $v \in B$ or $u \in B$ and $v \in A$). We note $\mathscr{G} = (A,B,E)$ such a graph where A and B constitute two *classes* of vertices.

A vertex v is a *neighbour* of a vertex u if $\{v,u\} \in E$, we say that u and v are *adjacent*. $\Gamma(u)$ is the set of neighbours of a given vertex u, it is called *neighbourhood* of u. An ordinary graph is *complete* if every couple of vertices from $V \times V$ are adjacent. A bi-graph is *complete* if every couple of vertices from $A \times B$ are adjacent.

An *induced subgraph* on the graph $\mathscr{G}$ by a set of vertices S is a graph composed of a vertex set $S \subseteq V$, and an edge set $E(S)$ that contains all vertices of E that bind vertices of S ($\forall u,v \in S$, $\{u,v\} \in E \Leftrightarrow \{u,v\} \in E(S)$). A set of vertices S that induces a complete subgraph is called a *clique*. If no vertex could be added to this induced subgraph without loosing the clique property then the clique is *maximal*. It is straightforward that every subgraph of a bi-graph is still bipartite, every vertex keeping the same class. A set of vertices S that induces a complete subgraph (in a bipartite sense) on a bi-graph $\mathscr{G}$ is called a *bi-clique* and if no vertex could be added without loosing this bi-clique property then the bi-clique is *maximal*.

A path from a vertex u to a vertex v is a sequence of vertices starting with u and ending with v and such that from each of its vertices there exists an edge to the next vertex in the sequence. The *length* of a path is the length of this vertices sequence minus one (it is to say the number of edges that run along the path). Two vertices are *connected* if there is a path between them. We note S^k the set of vertices connected to at least one vertex of S with a path of length smaller or equal to k. By definition $S^0 = S$. One can observe that $\forall k, S^k \subseteq S^{k+1}$. S^* is the set of vertices connected to at least one vertex of S with a path of any length, we have $S^* = \bigcup_{k \geqslant 0} S^k$. Two vertices are *disconnected* if there is no path between them. Two subsets A,B of vertices are disconnected if every vertex of A is disconnected from any vertex of B. A subset of vertices S is *connected* if there is a path between every pair of vertices of S, An induced subgraph that is connected is called a *connected component*. If no vertex could be added to this induced subgraph without loosing the property of connectedness then the connected component is *maximal*. Note that often "connected component" is used for speaking of a "maximal connected component".

3.1. From formal context to bi-graph

For every formal context $\mathscr{R} = (\mathbf{O},\mathbf{P},R)$, we can build an undirected bi-graph $\mathscr{G} = (V_o,V_p,E)$ s.t. there is a direct correspondence between: the set of objects $\mathbf{O}$ and a set V_o of "o-vertices", the set of properties $\mathbf{P}$ and a set V_p of "p-vertices", and between the binary relation R and a set of edges E. In

other words, there is one o-vertex for each object, one p-vertex for each property, and one edge between an o-vertex and a p-vertex if and only if the corresponding object possesses the corresponding property (according to R).

The four operators $(.)^{\Pi}$, $(.)^{N}$, $(.)^{\Delta}$ and $(.)^{\nabla}$ can be redefined for a set of vertices in this graph framework by replacing, in equations (3) to (6), $\mathbf{O}$ by V_o, $\mathbf{P}$ by V_p and $R^{-1}(y)$ by $\Gamma(y)$. Operators $(.)^{\Pi}$ and $(.)^{\Delta}$ can also be rewritten in the following way:

$$X^{\Pi} = \cup_{x \in X} \Gamma(x) \tag{7}$$

$$X^{\Delta} = \cap_{x \in X} \Gamma(x) \tag{8}$$

These notations are interesting since only the neighborhood of vertices of X is involved. It permits to immediately understand operators $(.)^{\Pi}$ and $(.)^{\Delta}$ in terms of neighborhood in the bi-graph: X^{Π} is the union of neighbors of vertices of X whereas X^{Δ} is the intersection of these neighbors. Note that with this writing and interpretation there is no difference between $(.)^{\Pi}$ and $(.)^{-1\Pi}$ neither between $(.)^{\Delta}$ and $(.)^{-1\Delta}$.

Graph interpretations of $(.)^{N}$ and $(.)^{\nabla}$ are less straightforward, nevertheless X^{N} can be understood as the union of neighbors of vertices of X that have no neighbors outside of X. In other words it is the set of vertices exclusively connected with vertices of X (but not necessarily all). Whereas X^{∇} is the set of p-vertices that are not connected to all o-vertices of $\overline{X}$.

3.2. *Two views of graph clusters in terms of connections*

The connections induced by $(.)^{\Delta}$ and $(.)^{\Pi}$ can also be understood in the graph setting framework. On the bi-graph $\mathscr{G} = (V_o, V_p, E)$, with $X \subseteq V_o$ and $Y \subseteq V_p$:

Proposition 1 *$X = Y^{-1\Delta}$ and $Y = X^{\Delta}$, iff $X \cup Y$ is a maximal bi-clique.*

Proof. Let (X,Y) be a pair such that $X = Y^{-1\Delta}$ and $Y = X^{\Delta}$. For all $x \in X$ and $y \in Y$, as $Y = \cap_{x \in X} \Gamma(x)$ we have $y \in \Gamma(x)$ thus $\{x,y\} \in E$. It means that the subgraph induced by $X \cup Y$ is complete. Moreover there is no vertex that are adjacent to all vertices of X (resp. Y) which are not in X^{Δ} (resp. $Y^{-1\Delta}$), therefore $X \cup Y$ is a maximal bi-clique.

If $X \cup Y$ is a maximal bi-clique, every vertex of X (resp. Y) is adjacent to any vertex of Y (resp. X) and there exists no vertex that is adjacent to all vertices of X (resp. Y) which are not in Y (resp. X), therefore it is straightforward that $Y = X^{\Delta}$ (resp. $X = Y^{-1\Delta}$). □

Proposition 2 *For a pair (X,Y) the two following propositions are equivalent:*

1. $X = Y^{-1\Pi}$ *and* $Y = X^{\Pi}$.
2. *(a)* $(X \cup Y)^{*} = (X \cup Y)$ *and*
   (b) $\forall v \in (X \cup Y), \Gamma(v) \neq \emptyset$.

Proof. $1 \Rightarrow 2$. By definition $(X \cup Y) \subseteq (X \cup Y)^{*}$. We show by recurrence that $(X \cup Y)^{*} \subseteq (X \cup Y)$. $(X \cup Y)^{0} \subseteq (X \cup Y)$ is given by definition. We then assume that it exists k such that $(X \cup Y)^{k} \subseteq (X \cup Y)$. We can notice that $(X \cup Y)^{k+1} \subseteq ((X \cup Y)^{k})^{1}$, by considering that a $k+1$ long path is a path of length k followed by a one edge step. So $(X \cup Y)^{k+1} \subseteq (X \cup Y)^{1}$. But as $X = Y^{-1\Pi}$ and $Y = X^{\Pi}$ all vertices connected to $X \cup Y$ with a path of length 1 are in $X \cup Y$. So $(X \cup Y)^{k+1} \subseteq (X \cup Y)$. This implies by recurrence that $\forall k \geqslant 0, (X \cup Y)^{k} \subseteq (X \cup Y)$. Thus $(X \cup Y)^{*} = \bigcup_{k \geqslant 0} (X \cup Y)^{k} \subseteq (X \cup Y)$. We still have to show that any vertex v of $X \cup Y$ has at least one neighbour, which is straightforward if we consider that either $v \in X^{\Pi}$ or $v \in Y^{-1\Pi}$.

$2 \Rightarrow 1$. We show that $X = Y^{-1\Pi}$, the proof is exactly the same for $Y = X^{\Pi}$. $Y^{-1\Pi}$ is the set of vertices adjacent to one vertex of Y, so $Y^{-1\Pi} \subset Y^{*}$ and then $Y^{-1\Pi} \subset (X \cup Y)^{*}$. That means that $Y^{-1\Pi} \subset (X \cup Y)$, but as the graph is bipartite: $Y^{-1\Pi} \subset X$. Let x be a vertex of X, by 2.b[†] x has at least one neighbour v, v is in X^{*} and therefore in $(X \cup Y)^{*}$, so $v \in X \cup Y$, but the graph is bipartite, so $v \in Y$. It's then straightforward that $X \subset Y^{-1\Pi}$ and therefore $X = Y^{-1\Pi}$. □

[†]Note that condition 2.b is not required if we assume that the relation (i.e. the graph) is bi-normalized.

A set S such that $S^* = S$ is not exactly a maximal connected component but it is a set of vertices disconnected from the rest of the graph. So if there is no strict subset S' of S satisfying $S'^* = S'$ it means that there is no subset of S disconnected from other vertices of S. In other words, S is connected and then S is a maximal connected component. Therefore, the following property:

Proposition 3 *For a pair (X,Y) the two following propositions are equivalent:*

1. *$X = Y^{-1\Pi}$ and $Y = X^{\Pi}$ and there is no strict subset $X' \subset X$ and $Y' \subset Y$ such that $X' = Y'^{-1\Pi}, Y' = X'^{\Pi}$.*
2. *$X \cup Y$ is a maximal connected component (which has at least two vertices).*

According to Propositions 1-3, it is worth noting that the two Galois connections correspond to extreme definitions of what a cluster (or a community) could be:

1. a group of vertices with **no link missing inside**.
2. a group of vertices with **no link with outside**.

One the one hand a maximal bi-clique is a maximal subset of vertices with a maximal edge density. Vertices cannot be moved closer, and in that sense one can not build a stronger cluster. On the other hand, a set of vertices disconnected from the rest of the graph can not be more clearly separated from other vertices. It corresponds to another type of cluster. In fact, only the smallest of such sets are really interesting, and they are nothing else than maximal connected components. This two extreme definitions were already pointed out for clusters in unipartite graphs[51].

4. Approximate conceptual structure: comparative state of the art

Formal concepts correspond to maximal bi-cliques, while independent sub-context correspond to disconnected subparts. These two notions may need to be relaxed for various reasons. Motivations are mainly twofold: first data may be noisy (some links in the graph may be missing or wrongly present), secondly one may need to have a simplified view of the set of concepts at work in the data. The first motivation is the one which is the most frequently emphasized in the literature. However in some sense any exceptional piece of data, even if there is no doubt about its validity, may be considered as contributing some "noise" that blurs the picture and prevents to have a simplified image of the data. This suggests to forget some "details" in order to summarize the information more easily. For instance, one may forget an edge because it simplifies the view by disconnecting weakly connected parts (for example the link $(4,d)$ in Figure 2), or introduce some missing edges in order to reinforce the connectedness inside a potential cluster (missing links $(1,h)$ and $(5,c)$ for example in Figure 2) and lay bare a simpler and more general concept.

Such ideas are encountered in formal concept analysis, when looking for relevant, or for approximate / pseudo formal concepts, but have also counterparts in other areas such as frequent itemset mining, or in graph clustering (also now known as "community detection" problem). In this section, we provide an overview of the existing literature in these different areas, starting with formal concept analysis. We end the section by considering different weighted extensions of formal concept analysis, where the weights bearing on the links between objects and properties may provide an indication of the importance of the link for deciding if we keep it, or we cut it in an approximation process.

4.1. Relevant or approximate formal concepts

A first line of research for simplifying a set of formal concepts which tend to be large (especially when data are noisy) is to select the relevant concepts only by means of appropriate measures[35]. The *stability* measure[36,32] is the most commonly used. The stability measure is all the greater for a concept (X,Y) as more subsets of X are such that $X^{\Delta} = Y$. This means in practice that the objects have not in general common properties outside Y, or in other words the concept is likely to

	1	2	3	4	5	6	7	8
a					×	×	×	×
b					×	×		
c						×	×	×
d				×	×	×	×	×
e							×	
f					×	×		×
g	×	×	×	×				
h		×	×	×				
i				×				

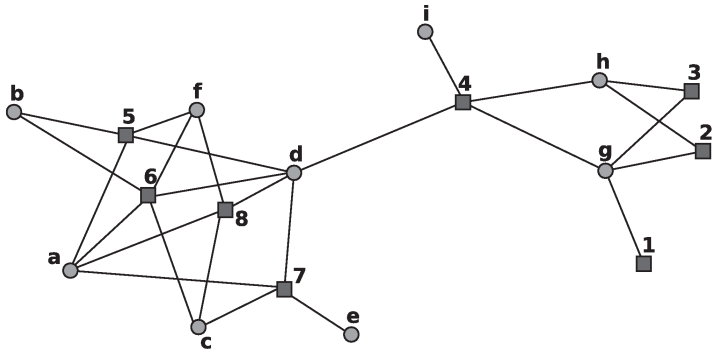


Figure 2: R': Relation R modified and the corresponding bi-graph.

be "stable" if one remove some objects. These approaches only select relevant concepts among the whole set of existing ones. So they cannot produce larger approximate concepts that would cover concepts that are distinct due to some missing links. One may think of doing that either by defining approximate formal concepts, or by using fuzzy formal concepts.

Defining an approximate formal concept can be done in a rather straightforward manner. Indeed the definition of X^{Δ} may be softened by looking for the set of properties shared by "most" objects in X rather than all. It leads to define[20] an operator $X^{\Delta,k}$ which allows for at most k exceptions among objects (provided that X has more than k elements). Namely,

$$X^{\Delta,k} = \left\{ y \in \mathbf{P} \,|\, \exists K \subseteq X, |K| = k, (X - K) \subseteq R^{-1}(y) \right\}$$

Likewise, we can define $Y^{\Delta,j} = \{x \in \mathbf{O} | \exists K' \subseteq Y, |K'| = j, (Y - K') \subseteq R(x)\}$. Then, an approximate formal concept is a pair (X,Y) such that $Y = X^{\Delta,k}$ and $X = Y^{\Delta,j}$. While a formal concept (X,Y) corresponds to the Cartesian product $X \times Y$, an approximate formal concept thus may have at most k holes by column and at most j holes by line. This idea is at work in data mining when looking for error-tolerant (closed) itemsets, see the next Section 4.2. Similarly one may think of defining approximately independent subcontexts (X',Y') and $(\overline{X'},\overline{Y'})$ by tolerating a limited number of elements in $(X \times \overline{Y'}) \cup (\overline{X'} \times Y')$.

In a related spirit, another approach consists in looking for pseudo concepts[44]. Pseudo concepts are a way to enlarge formal concepts: Indeed a set of objects X is associated with a set of primary properties Y, and with a secondary set of properties such that a majority of the objects are associated with the properties in this latter set.

4.2. Looking for error tolerant itemsets

Methods for mining association rules[1,52] look for sets of items that are frequently present together in a transaction database. The search for itemset can be related to formal concept analysis[46]. In that perspective, the transaction database is viewed as a formal context where each transaction stands for an object and each item in this transaction corresponds to a property satisfied by the object. Then the intent of each formal concept corresponds to a closed itemset. Then frequent itemset can be found by pruning the lattice of the formal concepts. In particular the rule: $Y_0 \Rightarrow Y \setminus Y_0$ has no exception iff $((Y_0)^{-1\Delta})^{\Delta} = Y$ where $Y_0 \subset Y$.

The presence of noise in data has led to the development of different ways of finding *error-tolerant itemsets*[30]. Roughly speaking, the idea is to no longer require that every item in a frequent itemset appears in each supporting transaction. This idea is very close of the idea of tolerating "holes" in format concept (which are maximal rectangles included in the formal context). As already said for approximate formal concept, it is not only the proportion of holes in the rectangle which matters but also their relative positions. Similar issues can be found with error-

tolerant itemsets where one distinguishes weak error-tolerant itemsets and strong error-tolerant itemsets[54]. In the former tolerance is global, while in the latter the number of possibly missing items (of the considered itemset) is limited inside each supporting transaction and the number of supporting transaction missing an item (of the considered itemset) is also limited. Moreover strong error tolerant itemsets may be further constrained by requiring that a given proportion of supporting transactions include all items (of the considered itemset)[10].

Besides, once the error-tolerant itemsets are obtained, one may think of merging them (by taking their union) on a similarity basis in order to reduce their number. Similarity may be judged by requiring that the result of the merging is a strong error-tolerant itemset with respect to some tolerance thresholds[9].

4.3. Community detection in hierarchical small world

There is a large amount of literature about graph clustering[51,23], especially since the emergence of so-called *community detection* problem[42]. Most of these works concern unipartite graphs. However such graphs are often obtained from bipartite ones (for example co-authoring graph between authors are based on the author-paper bipartite graph). Some authors prefer to address the clustering problem on bipartite graphs rather than on unipartite graph projections (where a part of the information is lost). In the following we briefly review existing graph clustering methods that have been adapted for bipartite graphs.

The classical graph partitioning problem[33] consists in splitting the vertex set in a given number of nearly equally-sized subsets such that the number of edges binding two vertices belonging to two different groups (i.e. the *cut size*) is minimal. This view has traditional applications in problems such as electronic circuit partitioning or load balancing in parallel computing. However it has also been applied to finding conceptual clusters in both unipartite and bipartite graphs[14]. The main drawback of this approach is that the number of clusters has to be known a priori, which makes sense for example in parallel computing when the number of resources is known. However when clustering communities, the number of clusters is usually unknown. In addition, as mentioned by Newman[43] the minimal cut argument is maybe a too naive approach to find "natural" structures in networks (some edges may be more significant than other according to the graph structure).

The modularity is a more complex quality measure of graph vertices partitioning. It is equal to the number of edges contained inside each cluster minus the same number for a graph built by rewiring vertices randomly but preserving their incidence degree. One looks for a partition that maximizes this modularity measure. This also amounts to minimize the number of edges between cluster minus the same number in the same randomly rewired graph. It was initially introduced by Newman and Girvan[42] as an argument to select a particular cut in a dendrogram (resulting from a hierarchical clustering algorithm). It has then been used as an objective function in various optimizing algorithm[11,7]. Different bipartite adaptations of this quality measure has been proposed recently[5,29,40].

Another common approach for graph clustering consists in using random walks. The intuitive idea is that random walk reveal the graph structure. Indeed more precisely, a random walker tends to be trapped into clusters since once inside a cluster the probability of getting out is low. This is because there are only few paths that would enable the walker to go out, and many paths that remain inside the cluster. This has been used for computing similarity degrees between graph vertices[26,27] which are then used for building clusters[47] (this algorithm has been adapted to bipartite graph clustering[41]). This same idea may also been used in a different way for defining quality measures of a partition of graph vertices[13,49]. Such quality measure is then used as an objective function in optimizing algorithms for obtaining a partition of the graph into clusters. The quality measure proposed by Rosvall and Bergstrom[49] seems especially promising[37], but has not be ap-

plied to bipartite graphs yet.

Another family of approaches views cliques as kernels of potential communities. A first idea is to look for adjacent cliques. It is illustrated by the well known method CFinder[45] which defines clusters as chains of adjacent k-cliques (a k-clique is a complete subgraph of size k, and two k-cliques are adjacent iff they have -at least- $k-1$ vertices in common). This method has been adapted to bipartite graphs[39]: clusters become chains of adjacent $K_{a,b}$-bicliques. A $K_{a,b}$-biclique is a complete subgraph of a o-vertex and b p-vertex and two $K_{a,b}$-bicliques are adjacent if they share $a-1$ o-vertex and $b-1$ p-vertex. It is worth noting that all maximal bicliques that count more than a o-vertices and b p-vertices are made of adjacent $K_{a,b}$-bicliques. In other words, each (big enough) maximal bi-clique is contained in a cluster (while maximal bi-cliques that are too small are ignored). Other approaches based on cliques work in more global way. Some authors have proposed to perform an agglomerative hierarchical clustering algorithm over edges (ie. 2-cliques) of uni-partite graphs[2]. It has been also proposed to transform a uni-partite graph into the graph of its k-cliques and then to compute a partitioning of the optain graph[22,21].

Apparently ignored by the above proposals, an older method is worth mentioning[31]. It starts with a bipartite graph for which one computes all the maximal cliques, then a bipartite graph is built by associating these maximal cliques to one of the two vertex sets. Finally this new graph is partitioned by minimizing the cut size.

In all these clique-based methods, the clusters of vertices that are obtained are allowed to overlap (since it is expected that individuals may belong to several communities). As emphasized by Hu et al.[31] the methods working with cliques better preserve the maximal cliques which constitute the core of communities, while all the previously presented methods that focus on vertices may split these meaningful maximal cliques.

4.4. Weighted extensions of formal concept analysis

There may be several good reasons for having a weighted formal context, i.e. a context where the links between objects and properties are weighted. Indeed, the weights may be understood in two different ways. First they may reflect the idea of allowing the properties to be a matter of degree. Another type of weights correspond to the situation where the properties remains binary but are pervaded with uncertainty. In both cases, this supposes that the information about the gradedness, or the uncertainty, is available. The first way, which has been the most investigated until now, amounts to consider that objects may have properties only to a degree. Such fuzzy formal concept analysis[6] is based on the operator :

$$X^{\Delta}(y) = \bigwedge_{x \in \mathbf{O}} (X(x) \to R(x,y)) \qquad (9)$$

where now R is a fuzzy relation, and X and X^{Δ} are fuzzy sets of objects and properties respectively, and $\bigwedge$ denotes the min conjunction operator and $\to$ an implication operator. A suitable choice of connective (the residuated Gödel implication: $a \to b = 1$ if $a \leqslant b$, and $a \to b = b$ if $a > b$) still enables us to see a fuzzy formal concept in terms of its level cuts X_α, Y_α such that $(X_\alpha \times Y_\alpha) \subseteq R_\alpha$ where $X_\alpha \times Y_\alpha$ are maximal, with $R_\alpha = \{(x,y) | R(x,y) \geqslant \alpha\}$, $X_\alpha = \{x \in \mathbf{O} | X(x) \geqslant \alpha\}$, $Y_\alpha = \{y \in \mathbf{P} | Y(y) \geqslant \alpha\}$.

Another way[16,15] is related to the idea of uncertainty. The possibilistic manner of representing uncertainty here is to associate with each link (x,y) a pair of number (α,β) such as $\alpha,\beta \in [0,1]$ and $\min(\alpha,\beta) = 0$ expressing respectively to what extent it is certain that the link exists (α), and does not exists (β). A link in a classical formal context corresponds to a pair $(1,0)$, the absence of a link to the pair $(0,1)$, and the pair $(0,0)$ models complete ignorance on the existence or not of a link. On this basis a link may be all the more easily added (resp. deleted) as α (resp. β) is larger.

One may also consider that some properties are less important, or that some objects are more typical[15]. Then weights are no longer put on links

or edges, but rather on the nodes. Thus forgetting a non compulsory property (e.g. the ability to fly for a bird) may help building a larger concept (e.g. birdness, although typical birds fly). Forgetting an object or a property also suppresses links, which may also help obtaining disconnected subparts.

These three views require different kinds of additional knowledge which are not often available, especially in large data sets. Moreover the two last views have been only recently introduce and have not been yet investigated in detail. However such weights may provide a help for building larger formal concepts and smaller sub-contexts. We shall see a way of producing weights by exploiting the structure of a binary relation through a diffusion process; see Section 5.2.

5. Looking for an approximate conceptual view of data

We have reviewed different trends of research aiming either at defining approximate concepts, subcontexts, closed itemsets, or at clustering communities in bipartite graphs. Among these methods, some focus on the identification of what may be retrospectively understood as approximate independent sub-contexts, while others rather look for approximate (formal) concepts. Indeed some methods partition the bipartite graph into non-overlapping subgraphs, while the others use formal concepts (or bi-cliques) for building larger groups (which are still allowed to overlap). We recognize here the two views for characterizing clusters (no link with outside vs. no missing link inside, see section 3.2). The partitioning methods suppress (or forget) links that are judged unimportant, whereas approximate concept methods tend to add (or to compensate) missing links.

It is worth noticing that the reviewed methods adopt one point of view (either looking for subcontexts that are as much as possible independent, or approximate concepts that are as dense as possible). But none of these methods may both add and suppress links (depending on the situation), at least in the initial context. However, there exists a two step approach that first looks for the concepts in the initial context, and then try to partition the set of concepts obtained. This may be done by considering the meta-context built from the links relating concepts and the objects of their extents and looking for approximatively independent subcontext in this meta-context, or doing the same in terms of bipartite graphs[31]. Moreover from a formal concept analysis point of view, it would seem natural to look first for approximate independent sub-contexts and then inside each of these subcontexts to look for the concepts.

Generally speaking, one may think of two types of approaches for identifying meaningful clusters (or communities), namely the ones that try to modify the initial context (by suppressing or adding links) in order to simplify the resulting clusters, and the ones that rather start from the set of concepts associated to the initial context and then try to simplify this set (for instance, by gathering or selecting relevant concepts). The first ones require some evaluation of the links in order to be able to decide to add or suppress them, while the second type of methods need some measure of the goodness of what is produced.

In the following, we present a two-step procedure that aims at providing a simplified, structured view of a set of data. These two steps correspond respectively to the two types of approaches mentioned above. The first step uses a random walk method for transforming the initial context into a graded one. This graded context is in turn reduced to a simplified binary context that is expected to have a smaller number of formal concepts. The second step then tries to merge together sets of concepts that overlap sufficiently. Before discussing this two-step procedure, we introduce the use of random walks, which also leads to a worth-noticing parallel with extended versions of formal concept analysis.

5.1. Random walks and formal concept analysis

A large panel of approaches developed within community detection literature use random walks for identifying communities. As already mentioned in Section 4.3, the underlying idea is that

random walkers tend to be trapped inside communities. Let us consider a random walk[8] on a bipartite-graph. In the following we continue to use the formal concept analysis notations taking advantage of the strict parallel with bipartite graphs established in section 3.1. The relation R is now replaced by a probabilistic transition matrix for going from a vertex x to a vertex y, or conversely. The probability is generally equally shared between the edges directly connected to the starting vertex. Let $P_{x\to y}$ be the probability for going from a vertex $x \in \mathbf{O}$ to a vertex $y \in \mathbf{P}$, formally defined as follows:

$$P_{x\to y} = \begin{cases} \frac{1}{|R(x)|} & \text{if} \quad (x,y) \in R, \\ 0 & \text{elsewhere,} \end{cases} \tag{10}$$

where $R(x)$ is the set of properties verified by x. Let $X(x)$ be the probability for a random walker to be in vertex $x \in \mathbf{O}$ at a given step; the probability $X^P(y)$ to reach a vertex $y \in \mathbf{P}$ at the next step is given by:

$$X^P(y) = \sum_{x\in\mathbf{O}} X(x).P_{x\to y} \tag{11}$$

Similarly when going from a property vertex to an object vertex we have the following equations:

$$P_{y\to x} = \begin{cases} \frac{1}{|R^{-1}(y)|} & \text{if} \quad (x,y) \in R, \\ 0 & \text{elsewhere,} \end{cases} \tag{12}$$

$$Y^P(x) = \sum_{y\in\mathbf{P}} Y(y).P_{y\to x} \tag{13}$$

where $Y(y)$ is the probability for a random walker to be in vertex $y \in \mathbf{P}$ at a given step.

The equations (11) and (13) correspond to a one step walk between an object and a property. More generally, a walk with an odd number of steps links objects and properties. The probability to reach a property from an object (or conversely an object from a property) after an odd number of t steps can be computed by composing t times the operator $(.)^P$. For t being an odd number, let $P^t_{x\to y}$ (resp. $P^t_{y\to x}$) be the probability to reach a property y from an object x (resp. an object x from an property y) in t steps.

One can show[26,27] that when t tends to infinity, $P^t_{x\to y}$ no more depends on the starting vertex x. However, the dynamics of the convergence towards this limit clearly depends on the starting node. Indeed, the trajectory of the random walker is completely governed by the topology of the graph: after t steps, any vertex y located at a distance of t links can be reached. The probability of this event depends on the number of paths between x and y, and the degree (i.e. number of neighbours) of each vertex along those paths. The more interconnections between these vertices, the higher the probability of reaching y from x. Therefore, for a small t, $P^t_{x\to y}$ reveals "how far" is y from x. This idea is used in the next section to compute a weight for each pair of object and property.

With regard to formal concept analysis, there is a worth noticing parallel between the "diffusion" operator at the basis of random walk methods and graded extensions of (the possibility theory reading of) formal concept analysis operators. Indeed equations (11) and (13) can formally be paralleled with the formula defining the operator at the basis of the definition of a formal concept[6]:

$$X^\Delta(y) \quad = \quad \min_{x\in\mathbf{O}} X(x) \to R(x,y) \tag{14}$$

and with the formula of the operator inducing independent sub-contexts[18]:

$$X^\Pi(y) \quad = \quad \max_{x\in\mathbf{O}} X(x) * R(x,y) \tag{15}$$

where R may be graded, as well as X, X^Π and X^Δ and where an usual choice for $*$ is min, and a residuated implication for $\to$.

This parallel between operators may be further extended between definitions of concepts and communities as stable points for these operators. Some random walk approaches define clusters as sets of vertices almost stable in the sense that a random walker tends to stay inside them[13,49]. In formal concept analysis, a formal concept is also a stable set for the Galois connection operator ($X^{\Delta\Delta} = X$ and $Y^{-1\Delta-1\Delta} = Y$).

Nevertheless, we do not intend in the following to compute a probabilistic substitute to the notion of formal concept. We rather compute

a weighted counterpart of the considered formal context by using random walks for assessing the importance of links between objects and properties.

5.2. A two-step procedure

In the following, we propose a two-step procedure aiming at providing an approximate conceptual view of data. It takes advantages of several ideas coming from the previous comparative discussion of different research trends. Namely the procedure first uses a random walk approach for providing a weighted counterpart to the formal context which is the basis of a heuristic method for diminishing the number of formal concepts associated to the context. Then, in a second step, the procedure merges formal concepts which are sufficiently close.

5.2.1. Using random walks for assessing the importance of edges

We propose to use a short random walk to attribute a weight $W(x,y)$ on each pair $(x,y) \in \mathbf{O} \times \mathbf{P}$ of object and property:

$$W(x,y) = \frac{1}{2}\left(P^3_{x\to y} + P^3_{y\to x}\right) \qquad (16)$$

The choice of random walk of length three may be discussed. As seen in the previous section, we have to use short random walks anyway. Moreover the number of steps t has to be odd, and if $t = 1$ only pairs (x,y) that are in R will have a non null weight. It is why we use the smallest informative number of steps, which has obvious computational advantages.

So the result of this computation is to substitutes a weighted context W to the original one R, where the weights accounts for a form of closeness. It is clear that given a threshold s, one can extract from W a binary relation R':

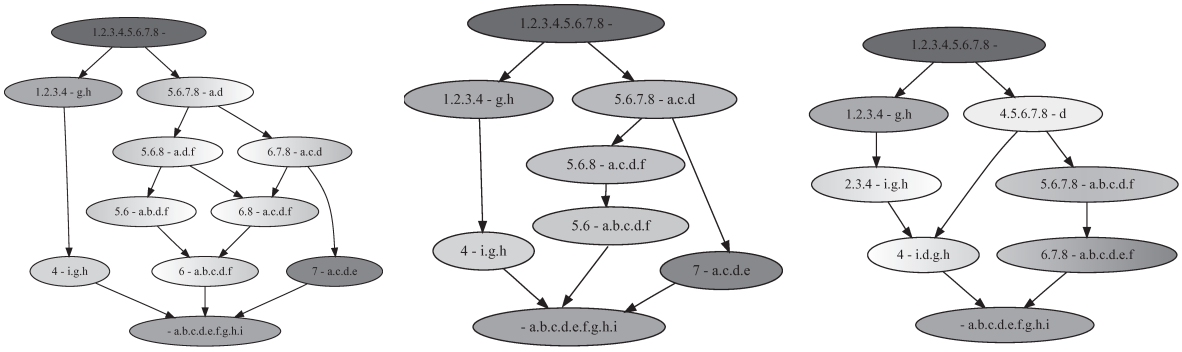
$$R' = \{(x,y) \in V_O \times V_P | W(x,y) > s\} \qquad (17)$$

This new formal context R' is associated with independent sub-contexts (if any) and with a lattice of formal concepts. Table 1(a) provides the weighted context W obtained from the example of Figure 2. As can be seen, an appropriate choice of s (here, e.g., $s = 0.16$) enables us to identify two approximately independent sub-contexts, since $(4,d)$ is removed. It has also the effect of introducing new links in the cells where the weight is sufficiently high. For example the link $(1,h)$ is added ; as a consequence concepts $(\{g\},\{1,2,3,4\})$ and $(\{g,h\},\{2,3,4\})$ are merged together into the new formal concept $(\{g,h\},\{1,2,3,4\})$. If we look at Figure 3 which shows the lattice of concepts extracted from the relation of Figure 2 on which the transformation from R' to W has been applied with different thresholds s, one observes that the number of concepts diminishes with s. However, there is *no monotonic decrease* in general. See for instance Figure 4 which gives the number of concepts in function of the threshold, where in the vicinity of 0.100 the number of concepts goes from 8 to 10 and then to 9. We also observe in Figure 3 that when the threshold is decreased too much, the two independent sub-contexts no longer exist (since for $s = 0.10$, $(4,d)$ is no longer removed). Note that independent sub-contexts are easy to recognize in the lattice of concepts, since they correspond to families of paths from the top and the bottom formal concepts (with empty sets of properties and objects respectively), which are fully separated from the others.

In order to explore what values of the threshold may be of interest for looking for approximately independent sub-contexts, or for diminishing the number of formal concepts, two landmark values are of interest: let m be the weight of the worse of the edges and M be the weight of the best of the "non-edges" in the formal context. Namely,

$$m = \min(\{W(x,y), \forall (x,y) \in R\}) \qquad (18)$$
$$M = \max(\{W(x,y), \forall (x,y) \in \overline{R}\}) \qquad (19)$$

One can note that if all edges have a larger weight than all "non-edges" (i.e. $M \leqslant m$) then the relation stays the same for any value of the threshold in $[M,m]$. When $m \leqslant M$, the idea it to remove some edges that have a weight lower than the best "non-edges", and to add some "non-edges" that have a better score than the worst edges, by choos-



(a) $s = 0.20$ (b) $s = 0.16$ (c) $s = 0.10$

Figure 3: Lattices of formal concepts for different values of the threshold (example of Figure 2, with W given in Table 1(a).

ing a threshold $s \in [m,M]$. In practice, one may blindly take $s = \frac{m+M}{2}$, or explore different values. However, a better strategy seems to first look for approximately independent sub-contexts by taking a value close to M (see Table 1(b)), and then *in each of the sub-contexts* to diminish the threshold progressively in order to increase the number of edges and checking the corresponding number of formal concepts.

As a matter of illustration, Table 2 gives the result of the transformation on a relation that corresponds to the *Southern Women Data Set*[12], a bipartite graph between 18 women and 14 events that is a standard example in social network analysis literature and in community detection literature. The crosses in Table 2 (as well as in Table 1(b)) represent the new relation R' computed from W, and the cells in gray are the ones that have been modified by the transformation (to get back to the original relation, one has just to swap cross and blank in each of these gray cells). In Table 2, there are 67 formal concepts in the original relation and only 22 with the new relation.

5.2.2. *Merging concepts*

At the end of the previous step, it is expected to have a simplified view of the original context in terms of sub-contexts (when possible) and formal concepts. However the number of formal concepts may remain still too high with respect to user needs. Indeed in the example of Table 2 after the first step there is still 22 formal concepts. In order to further simplify the view one may complete the previous procedure with a second step aiming at merging "close" formal concepts.

For two concepts $a = (X_a, Y_a)$ and $b = (X_b, Y_b)$, the inclusion value of a in b can be defined as follows:

$$\mathrm{inc}(a,b) = \frac{|X_a \cap X_b||Y_a \cap Y_b|}{|X_a||Y_a|} \tag{20}$$

This can be extended to a set C of n concepts

$$s(C) = \max(\{\min(\{\mathrm{inc}(a,b), \forall a \in C\}), \forall b \in C\}), \tag{21}$$

where $s(C) > \theta$ means that there is a concept $b \in C$ such that, for all concepts $a \in C$, $\theta\%$ of a is included in b. In other words, there exists a concept in the set C that almost includes all concepts of the set C.

In order to merge formal concepts, one may apply the following standard agglomerative algorithm :

1. each concept is in its own cluster,
2. for each pair of clusters, the score of their union is computed,
3. if this maximal score is larger than the threshold:
   (a) a pair having maximal score is selected and merged,
   (b) loop to 2.

Table 1: Modification of the relation permit to go from 13 to 11 concepts

(a) weighted relation W

	1	2	3	4	5	6	7	8
a				.048	.221	.252	.228	.224
b				.034	.217	.225	.101	.157
c				.041	.165	.213	.233	.204
d	.037	.051	.051	.143	.211	.234	.216	.212
e				.031	.070	.117	.279	.122
f				.041	.226	.246	.136	.204
g	.352	.367	.367	.320	.013	.011	.013	.013
h	.208	.304	.304	.300	.015	.013	.015	.015
i	.062	.109	.109	.279	.031	.030	.031	.031

(b) R' computed from W with $s = 0.20$

	1	2	3	4	5	6	7	8
a					×	×	×	×
b					×	×		
c						×	×	×
d					×	×	×	×
e							×	
f					×	×		×
g	×	×	×	×				
h	×	×	×	×				
i				×				

Table 2: New relation from southern woman bipartite network after applying random walk transformation. Cells in gray are the ones affected by the transformation.

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
Ev.	×	×	×	×	×	×	×	×						
La.	×	×	×	×	×	×	×	×						
Th.	×	×	×	×	×	×	×	×						
Br.	×	×	×	×	×	×	×	×						
Ch.			×	×	×		×							
Fr.			×		×	×		×						
El.					×	×	×	×						
Pe.								×	×					
Ru.							×	×	×					
Ve.								×	×					
My.								×	×	×		×		
Ka.								×	×	×		×	×	×
Sy.							×	×	×	×		×	×	×
No.							×	×	×	×	×	×	×	×
He.								×	×	×	×	×	×	×
Do.								×	×	×		×		
Ol.									×		×			
Fl.									×		×			

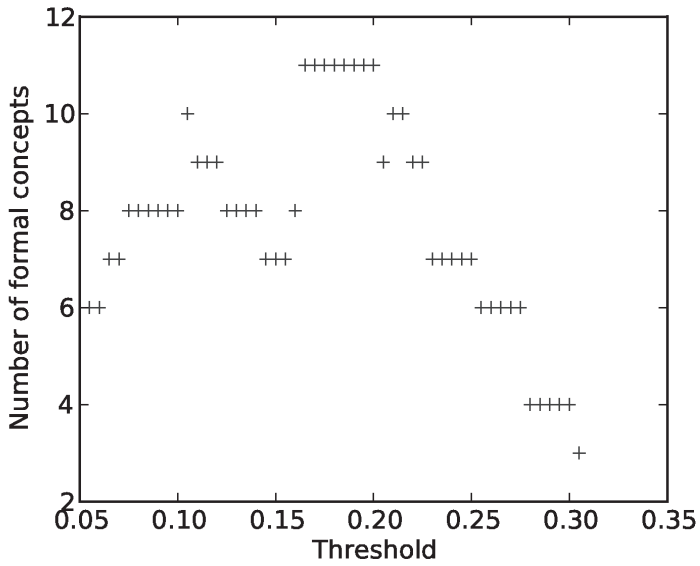


Figure 4: Number of formal concepts (including the bottom and top trivial ones) in function of the threshold applied to the weighted relation W of the formal context in Figure 2.

However, note that it does not necessarily lead to a partition of the set of concepts that is optimal in the sense that it does not necessarily maximize $\min_i(s(C_i))$ among all the n-partitions $\{C_1, C_2, \ldots, C_n\}$ of the set of formal concepts.

Table 3 provides the result of this second step on the women social network example. Finally, 6 non trivial clusters are retained. Clearly, a threshold θ too small may lead to an oversimplification of the formal context, as in the initial example, see Table 4, where each of the two (approximately) independent sub-context becomes a unique cluster for $\theta = 0.4$. It is worth noticing that this second step still preserves a structured view, since one may keep the benefit of the lattice of concepts restricted to the "father" concepts of each group of concepts (which approximately include the other concepts in their group). Besides, other criteria (e.g., the relative size) may be used for a further selection in the set of formal concepts, if too many formal concepts remain in the result of this second step.

6. Conclusion

Starting with a view of a formal context as a bi-graph, the paper has shown that formal concepts correspond to the idea of maximal bicliques, whereas independent sub-contexts, obtained thanks to the introduction of another connection, correspond to disconnected subsets of vertices. Noticeably enough, these two constructs reflect two ideal views of the idea of graph cluster, namely a set of vertices with no link missing inside and a group of vertices with no link with outside. The last part of the paper, after a review of different ways of getting approximate structured views of a formal context, or equivalently of clustering a bipartite graph, has outlined a two-step procedure for laying approximately independent sub-contexts if any, and simplifying the lattice of the formal concepts. The first step of the procedure takes advantage of random walks methods for introducing a closeness estimation of the vertices (or equivalently of the pairs (object, property) in the formal context), the second step merges formal concepts that are approximately included. Clearly,

Table 3: $\theta = 0.6$, Final clusters for *Southern Women Data Set*:

X	Y
E9.E11	No.He.Ol.Fl.
E3.E4.E5.E7	Ev.La.Th.Br.Ch.Fr.El.
E9.E7.E8	Ev.La.Th.Br.Ch.El.Ru.Sy.No.
E7.E8.E9.E10.E11.E12.E13.E14	My.Ka.Sy.No.He.Do.
E1.E2.E3.E4.E5.E6.E7.E8	Ev.La.Th.Br.Fr.El.
E9.E8	Ev.La.Th.Br.Fr.El.Pe.Ru.Ve.My.Ka.Sy.No.He.Do.Ol.Fl.

Table 4: $\theta = 0.4$, Final clusters of the academic example of Figure 2:

X	Y
1.2.3.4	igh
5.6.7.8	abcdef

by bridging different areas, this overview paper has discussed several ideas that are worth of further investigation, and the procedure that has been outlined for clustering bipartite graphs may be still refined and improved after a proper evaluation.

References

1. R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. of the 1993 ACM SIGMOD Inter. Conf. on management of data*, pages 207–216, New York, 1993.
2. Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466:761–764, August 2010.
3. R. Albert and A. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, January 2002.
4. A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.
5. M. J. Barber. Modularity and community detection in bipartite networks. *Phys. Rev. E*, 76(6), December 2007.
6. R. Belohlavek. *Fuzzy Relational Systems: Foundations and Principles*. Kluwer Academic Publishers, 2002.
7. V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *J. of Statistical Mechanics: Theory and Experiment*, 10:P10008, October 2008.
8. B. Bollobas. *Modern Graph Theory*. Springer, 2002.
9. L. Cerf, P.N. Mougel, and J.F. Boulicaut. Agglomerating local patterns hierarchically with alpha. In *Proc. of the 18th ACM Conf. on Information and Knowledge Management*, CIKM'09, pages 1753–1756, New York, 2009.
10. H. Cheng, P. S. Yu, and J. Han. Ac-close: Efficiently mining approximate closed itemsets by core pattern recovery. In *Proc. of the 6th Inter. Conf. on Data Mining*, ICDM'06, pages 839–844, Washington, 2006.
11. A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. *Phys. Rev. E*, 70(6):066111, December 2004.
12. A. Davis, B. B. Gardner, and M. R. Gardner. *Deep South*. Chicago: The University of Chicago Press, 1941.
13. J. C. Delvenne, S. N. Yaliraki, and M. Barahona. Stability of graph communities across time scales. *Proc. of the National Academy of Sciences of the USA*, 107(29):12755–12760, 2010.
14. I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proc. of the seventh ACM SIGKDD Inter. Conf. on Knowledge Discovery and Data mining*, pages 269–274, San Francisco, 2001.
15. Y. Djouadi, D. Dubois, and H. Prade. Graduality, uncertainty and typicality in formal concept analysis. In C. Cornelis, G. Deschrijver, M. Nachtegael, S. Schockaert, and Y. Shi, editors, *35 years of Fuzzy Set Theory*, pages 127–147. Springer, 2010.
16. Y. Djouadi, D. Dubois, and H. Prade. Pos-

sibility theory and formal concept analysis: Context decomposition and uncertainty handling. In E. Hüllermeier, R. Kruse, and F. Hoffmann, editors, *Computational Intelligence for Knowledge-Based Systems Design, Proc. 13th Inter. Conf. on Information Processing and Management of Uncertainty (IPMU 2010), Dortmund, June 28 - July 2*, volume 6178 of *LNAI*, pages 260–269. Springer, 2010.
17. D. Dubois, F. Dupin de Saint-Cyr, and H. Prade. A possibility theoretic view of formal concept analysis. *Fundamenta Informaticae*, 75(1):195–213, 2007.
18. D. Dubois and H. Prade. Possibility theory and formal concept analysis in information systems. In *Proc. 13th Inter. Fuzzy Systems Association World Congress IFSA-EUSFLAT*, Lisbon, July 2009.
19. D. Dubois and H. Prade. Bridging gaps between several frameworks for the idea of granulation. In *Proc. IEEE Symp. on Foundations of Computational Intelligence, 3rd Symp. Series on Computational Intelligence (IEEE SSCI'11), Paris, April 11-15*, 2011.
20. D. Dubois and H. Prade. Possibility theory and formal concept analysis: Characterizing independent sub-contexts. *Fuzzy Sets and Systems*, 2012, to appear.
21. T. S. Evans. Clique graphs and overlapping communities. *J. of Statistical Mechanics: Theory and Experiment*, 2010(12):P12037, 2010.
22. T. S. Evans and R. Lambiotte. Line graphs, link partitions, and overlapping communities. *Phys. Rev. E*, 80(1):016105, July 2009.
23. S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.
24. B. Ganter, G. Stumme, and R. Wille (eds.). *Formal Concept Analysis: Foundations and Applications*, volume 3626 of *LNAI*. Springer, 2005.
25. B. Ganter and R. Wille. *Formal Concept Analysis*. Springer-Verlag, 1999.
26. B. Gaume. Balades aléatoires dans les petits mondes lexicaux. *I3 Information Interaction Intelligence*, 4(2), 2004.
27. B. Gaume and F. Mathieu. PageRank induced topology for real-world networks. *Complex Systems*. (to appear).
28. B. Gaume, E. Navarro, and H. Prade. A parallel between extended formal concept analysis and bipartite graphs analysis. In E. Hüllermeier, R. Kruse, and F. Hoffmann, editors, *Computational Intelligence for Knowledge-Based Systems Design, Proc. 13th Inter. Conf. on Information Processing and Management of Uncertainty (IPMU 2010), Dortmund, June 28 - July 2*, volume 6178 of *LNAI*, pages 270–280. Springer, 2010.
29. R. Guimerà, M. Sales-Pardo, and L. N. Amaral. Module identification in bipartite and directed networks. *Phys. Rev. E*, 76:036102, September 2007.
30. R. Gupta, G. Fang, B. Field, M. Steinbach, and V. Kumar. Quantitative evaluation of approximate frequent pattern mining algorithms. In *Proc. of the 14th ACM SIGKDD Inter. Conf. on Knowledge discovery and data mining*, KDD '08, pages 301–309, New York, 2008.
31. T. Hu, C. Qu, C. L. Tan, S. Y. Sung, and W. Zhou. Preserving patterns in bipartite graph partitioning. In *Proc. of 18th IEEE Inter. Conf. on Tools with Artificial Intelligence. ICTAI'06.*, pages 489–496, November 2006.
32. N. Jay, F. K., and A. Napoli. Analysis of social communities with iceberg and stability-based concept lattices. In R. Medina and S. A. Obiedkov, editors, *Proc. 6th Inter. Conf. on Formal Concept Analysis (ICFCA'08), Montreal*, volume 4933 of *LNCS*, pages 258–272. Springer, 2008.
33. B. W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. *Bell System Technical J.*, 49:291, 1970.
34. F. Klawonn. Fuzzy points, fuzzy relations and fuzzy functions. In V. Novak and I. Perfilieva, editors, *Discovering the World with Fuzzy Logic*, pages 431–453. Physica-Verlag, Heidelberg, Germany, 2000.
35. M. Klimushkin, S. A. Obiedkov, and C. Roth. Approaches to the selection of relevant concepts in the case of noisy data. In L. Kwuida and B. Sertkaya, editors, *Proc. 8th Inter. Conf. on Formal Concept Analysis (ICFCA'10), Agadir*, volume 5986 of *LNCS*, pages 255–266. Springer, 2010.
36. S. O. Kuznetsov, S. A. Obiedkov, and C. Roth. Reducing the representation complexity of lattice-based taxonomies. In U. Priss, S. Polovina, and R. Hill, editors, *Conceptual Structures: Knowledge Architectures for Smart Applications, Proc. 15th Inter. Conf. on Conceptual Structures, Sheffield*, volume 4604 of *LNCS*, pages 241–254. Springer, 2007.
37. A. Lancichinetti and S. Fortunato. Community detection algorithms: A comparative analysis. *Phys. Rev. E*, 80(5):056117, November 2009.
38. M. Latapy, C. Magnien, and N. Del Vecchio. Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1):31–48, 2008.
39. S. Lehmann, M. Schwartz, and Lars Kai

Hansen. Biclique communities. *Phys. Rev. E*, 78(1):016108–9, 2008.

40. T. Murata. Modularity for bipartite networks. In Nasrullah Memon, Jennifer Jie Jie Xu, David L. L. Hicks, and Hsinchun Chen, editors, *Data Mining for Social Network Data*, volume 12 of *Annals of Information Systems*, pages 109–123. Springer US, 2010.
41. E. Navarro, Y. Chudy, B. Gaume, G. Cabanac, and K. Pinel-Sauvagnat. Kodex ou comment organiser les résultats d'une recherche d'information par détection de communautés sur un graphe biparti ? In *Proc. of CORIA, Avignon*, pages 25–40, March 2011.
42. M. E. J. Newman. The structure and function of complex networks. *SIAM Review 45*, pages 167–256, March 2003.
43. M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74(3):036104–19, 2006.
44. Y. Okubo and M. Haraguchi. Finding Top-N pseudo formal concepts with core intents. In *Proc. of the 6th Inter. Conf. on Machine Learning and Data Mining in Pattern Recognition*, pages 479–493, Leipzig, Germany, 2009.
45. G. Palla, I. Derenyi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.
46. N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Efficient mining of association rules using closed itemset lattices. *Information Systems*, 24(1):25–46, 1999.
47. P. Pons and M. Latapy. Computing communities in large networks using random walks (long version). *J. of Graph Algorithms and Applications (JGAA)*, 10(2):191–218, 2006.
48. M. Porter, J. P. Onnela, and P. J. Mucha. Communities in networks. *Notices of the American Mathematical Society*, 56(9):1082–1097, 2009.
49. M. Rosvall and C. T. Bergstrom. Maps of random walks on complex networks reveal community structure. *Proc. of the National Academy of Sciences*, 105(4):1118–1123, 2008.
50. C. Roth and P. Bourgine. Epistemic communities: Description and hierarchic categorization. *Mathematical Population Studies*, 12:107–130, June 2005.
51. S. E. Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27–64, 2007.
52. P. N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison-Wesley, Boston, 2005.
53. D. Watts and S. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393:440–442, 1998.
54. C. Yang, U. Fayyad, and P. S. Bradley. Efficient discovery of error-tolerant frequent itemsets in high dimensions. In *Proc. of the seventh Inter. Conf. on Knowledge Discovery and Data mining*, pages 194–203, 2001.