HAL open science

Totally Balanced Formal Context Representation

François Brucker, Pascal Préa

▶ To cite this version:

François Brucker, Pascal Préa. Totally Balanced Formal Context Representation. ICFCA, 2015, Malaga, Spain. hal-02435802

HAL Id: hal-02435802
https://hal.science/hal-02435802v1

Submitted on 11 Jan 2020

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.

Totally Balanced Formal Context Representation

François Brucker and Pascal Préa

École Centrale Marseille – LIF, CNRS UMR 7279
{francois.brucker, pascal.prea}@centrale-marseille.fr,

Abstract. We show in this paper that doubly lexical orders of totally balanced formal context matrices yield a unique graphical representation binding formal contexts, associated concepts and underlying lattice directed cover graphs. Moreover this representation can be done linearly in the size of the formal context matrix.

Keywords: formal concept analysis · totally balanced matrices · dismantlable lattices · graphical representation

1 Introduction

We will focus in this paper on a special case of formal contexts, those associated with dismantlable lattices. These formal contexts can be helpful in practice because they only have a polynomial number of concepts that are easily computable and, as we shall show in this paper, admit a convenient graphical representation.

Indeed, given a formal context matrix $\mathcal{M}$ whose associated concept lattice is dismantlable, we present a procedure which associates each formal concept to an element of the formal context matrix. This allows to superpose the associated directed cover graph onto the data (Figure 2). Moreover, since the procedure is linear in the size of the matrix and the graphical representation involves non-overlapping ordered 2-dimensional boxes (see Figures 5 and 6), it can be used to locally see the interactions between concepts or to explore areas of interest in very large data-sets (through approximation of the original data, which can also be done linearly in the size of the context matrix).

Finally, because dismantlable lattices generalize several models used in clustering (hierarchical clustering and seriation clustering for instance) this procedure can be used in many fields of applications. For instance they are well suited for phylogenetic problems because co-atomic dismantlable lattices are in bijection with strongly chordal graphs (which are a sub-class of chordal graphs — graphs whose every cycle $(x_1, x_2, \ldots, x_n, x_1)$ with $n \geq 3$ contains an edge $x_i x_j$ with $i < j + 1$ — used in perfect phylogeny) and the clusters generated by phylogenetic trees (X-trees) form a dismantlable lattice (see [6] or [5]).

In a theoretical point of view, dismantlable lattices are the "trees" for lattices and there is an ongoing work to see if one can decompose a given lattice into a sum of dismantlable lattices.

Formally speaking, a *formal context* [8] $K = (G, M, I)$ is a triple where G is a set of objects, M a set of attributes and $I \subseteq G \times M$ a binary relation. For subsets $A \subseteq G$ and $B \subseteq M$ two operators (so called *derivation operators*) are defined: $A^{\uparrow} = \{y \mid xIy, \forall x \in A\}$ and $B^{\downarrow} = \{x \mid xIy, \forall y \in B\}$. A *formal concept* associated with K is a pair (A, B) where:

- $A \subseteq G$ and $B \subseteq M$,
- $A = B^{\downarrow}$,
- $B = A^{\uparrow}$.

We shall assume that all sets in this paper are finite. Thus if we label the objets and the attributes of a formal context $K = (G, M, I)$ such that $G = \{l_1, \ldots, l_n\}$ and $M = \{c_1, \ldots, c_m\}$, K is equivalent to a $n \times m$ binary matrix $\mathcal{M}$ (called *formal context matrix*) such that $\mathcal{M}_{i,j} = 1$ if $l_i I c_j$ and 0 otherwise. Formal contexts or its associated formal context matrix can be represented by a cross table (see Table 1).

Table 1. Example of cross table (left) and its associated formal context matrix (right).

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
l_1			×	×	×			
l_2						×	×	×
l_3		×	×	×		×		×
l_4							×	×
l_5	×	×	×					

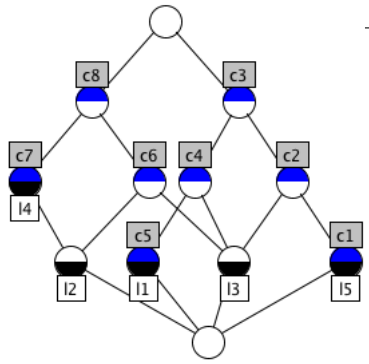
	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
l_1	0	0	1	1	1	0	0	0
l_2	0	0	0	0	0	1	1	1
l_3	0	1	1	1	0	1	0	1
l_4	0	0	0	0	0	0	1	1
l_5	1	1	1	0	0	0	0	0

The formal concepts associated with $K = (G, M, I)$ are exactly the pairs $(A, A^{\uparrow})$ where $A^{\uparrow\downarrow} = A$ (equivalently the pairs $(B^{\downarrow}, B)$ where $B^{\downarrow\uparrow} = B$). It is well known that the set $\mathcal{B}(G, M, I)$ of formal concepts associated with the order $\leq$ defined as $(A_1, B_1) \leq (A_2, B_2)$ whenever $A_1 \subseteq A_2$ (equivalently $B_2 \subseteq B_1$) forms a formal lattice that one can represent by its order diagram (see Figure 1). The order diagram is a graphical representation of the *directed cover graph* associated with the formal lattice $(\mathcal{B}(G, M, I), \leq)$ which is the directed graph $\mathcal{G} = (\mathcal{B}(G, M, I), E)$ where $uv \in E$ whenever $u \prec v$.

Finally, for any $n \times m$ binary matrix $\mathcal{M}$, one can define the formal context $K_{\mathcal{M}} = (\{1, \ldots, n\}, \{1 \ldots, m\}, I_{\mathcal{M}})$ such that $i I_{\mathcal{M}} j$ whenever $\mathcal{M}_{i,j} = 1$. The *column j of $\mathcal{M}$* is equal to $C_j = \{i \mid \mathcal{M}_{i,j} = 1\}$ and the *line i of $\mathcal{M}$* is equal to $L_i = \{j \mid \mathcal{M}_{i,j} = 1\}$. According to these definitions, it is known [4] that:

- the closure under intersection of $\{C_1, \ldots, C_m\} \cup \{\{1, \ldots, n\}\}$, noted $\mathcal{C}$, is equal to the set $\{A \mid A^{\uparrow\downarrow} = A\}$,
- the closure under intersection of $\{L_1, \ldots, L_n\}$, noted $\mathcal{L}$, is equal to the set $\{B \mid B^{\downarrow\uparrow} = B\}$.

The above equivalence gives a way to link known *hypergraph* classes to known lattice models (an hypergraph is a pair $\mathcal{H} = (X, H)$ where $H \subseteq 2^X$). Indeed it is



number	formal concept
0	$(\emptyset, \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8\})$
1	$(\{\mathbf{l_2}\}, \{c_6, c_7, c_8\})$
2	$(\{\mathbf{l_1}\}, \{c_3, c_4, \mathbf{c_5}\})$
3	$(\{\mathbf{l_3}\}, \{c_2, c_3, c_4, c_6, c_8\})$
4	$(\{\mathbf{l_5}\}, \{\mathbf{c_1}, c_2, c_3\})$
5	$(\{l_2, \mathbf{l_4}\}, \{\mathbf{c_7}, c_8\})$
6	$(\{l_2, l_3\}, \{\mathbf{c_6}, c_8\})$
7	$(\{l_1, l_3\}, \{c_3, \mathbf{c_4}\})$
8	$(\{l_3, l_5\}, \{\mathbf{c_2}, c_3\})$
9	$(\{l_2, l_3, l_4\}, \{\mathbf{c_8}\})$
10	$(\{l_1, l_3, l_5\}, \{\mathbf{c_3}\})$
11	$(\{l_1, l_2, l_3, l_4, l_5\}, \emptyset)$

Fig. 1. Order diagram (left) and formal concepts (right) associated with the formal context matrix of Table 1 (numbered from bottom to top and left to right).

clear that by labeling X as $\{x_1, \ldots, x_n\}$ and H as $\{h_1, \ldots, h_m\}$, each hypergraph $\mathcal{H}$ is equivalent to a $n \times m$ binary matrix $\mathcal{M}(\mathcal{H})$ where $\mathcal{M}(\mathcal{H})_{i,j} = 1$ whenever $x_i \in h_j$. Conversely, a $n \times m$ binary matrix $\mathcal{M}$ is equivalent to a hypergraph $\mathcal{H}(\mathcal{M}) = (\{1, \ldots, n\}, \{C_1, \ldots, C_m\})$.

We will in this paper focus on *totally balanced hypergraphs*. A hypergraph $\mathcal{H} = (X, H)$ is totally balanced [2] if there is no cycle $(v_1, e_1, \ldots, v_k, e_k)$ with $k \geq 3$ such that:

- $v_i \in e_i \cap e_{i-1}$ (for $i > 1$) and $v_1 \in e_k$,
- $v_i \notin e_j$ for $j \notin \{i, i-1\}$ and $(i, j) \neq (1, k)$.

This class has very nice combinatorial properties (several NP-hard problems become polynomial when focusing on this class) and a clear cluster interpretation when dealing with real data (clusters are connected parts of some tree and restrictions of a totally balanced hypergraph remains totally balanced [9]).

This class is in correspondence with *dismantlable lattices* [6]. Dismantlable lattices where defined recursively by Rival [10] as lattices L for which there is a *doubly irreducible element* x in L (a doubly irreducible element is such that there is at most one element x^- such that $x^- \prec x$ and at most one element x^+ such that $x \prec x^+$) such that $L \backslash \{x\}$ is also dismantlable. For instance, the order diagram of Figure 1 represents a dismantlable lattice (elements 2, 4 and 5 are doubly irreducible).

Considering a formal context $K = (G, M, I)$, a formal concept (A, B) is doubly irreducible for the associated lattice $(\mathcal{B}(G, M, I), \leq)$ if there exists $(g, m) \in (A, B)$ such that for any formal concept $(U, V) \in \mathcal{B}(G, M, I)$, if $g \in U$, then $A \subseteq U$ and if $m \in V$ then $B \subseteq V$. Thus, a *formal context is totally balanced*

(*ie.* its associated lattice is dismantlable) if and only if there is a decomposition order such that [8]:

- $\mathcal{K}_0 = \mathcal{B}(G, M, I)$,
- there is $(A_i, B_i) \in \mathcal{K}_i$ such that $\exists (g, m) \in (A_i, B_i)$ for which $\forall (U, V) \in \mathcal{K}_i$, $g \in U \Rightarrow A_i \subseteq U$, and $m \in V \Rightarrow B_i \subseteq V$,
- $\mathcal{K}_{i+1} = \mathcal{K}_i \backslash \{(A_i, B_i)\}$
- $\mathcal{K}_{|\mathcal{H}|} = \emptyset$.

Looking at Figure 1, the formal context admits $2, 7, 4, 8, 10, 5, 1, 9, 6, 3, 0, 11$ as a decomposition order (among many other).

The paper is organized as follows. We will first (Section 2) recall some properties of totally balanced matrices and use them in Sections 3 and 4 to show that there is a one-to-one correspondence between proper formal concepts and some elements of the associated formal concept matrix.

2 Doubly lexical ordering of totally balanced matrices

We recall here some properties of totally balanced hypergraphs and show the implication to totally balanced formal concepts.

2.1 Formal concepts of totally balanced formal context matrices

A $n \times m$ binary matrix $\mathcal{M}$ is said to be totally balanced if its associated hypergraph $\mathcal{H}(\mathcal{M})$ is totally balanced. The result from which all the results of this paper will follow is given by Theorem 1 and is linked with gamma-free matrices.

A $n \times m$ binary matrix $\mathcal{M}$ is said to be *gamma-free* whenever for any $1 \leq i < i' \leq n$ and any $1 \leq j < j' \leq m$: $\mathcal{M}_{i,j} = \mathcal{M}_{i,j'} = \mathcal{M}_{i',j} = 1$ implies $\mathcal{M}_{i',j'} = 1$.

For instance, the formal context matrix from Table 1 is not gamma-free because $\mathcal{M}_{1,3} = \mathcal{M}_{1,5} = \mathcal{M}_{3,3} = 1$ and $\mathcal{M}_{3,5} = 0$.

Theorem 1 ([1]). *Let $\mathcal{M}$ be a $n \times m$ binary matrix. $\mathcal{M}$ is totally balanced if and only if there is a gamma-free ordering of its line and columns.*

Even though the formal context matrix from Table 1 is not gamma-free, it admits a gamma-free reordering. See for instance Table 2 which is gamma-free.

We will use a special gamma-free ordering in section 2.2 which also characterizes totally balanced matrices, but before that, we just state another main property of totally balanced hypergraphs and precise it for totally balanced matrices.

Hypergraphs for which the intersection of three clusters is the intersection of two of them are called *weak hierarchies* [3] and are a very popular model in classification theory because they generalize the well known hierarchical model (the intersection of two clusters is either empty or is one of them). Moreover Proposition 1 shows that the closure under intersection which is usually an expensive operation can be done easily for weak hierarchies.

Table 2. Cross table associated with a gamma-free ordering of the formal context matrix from Table 1.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
l_5	×	×	×					
l_4							×	×
l_3		×	×	×		×		×
l_2						×	×	×
l_1			×	×	×			

Proposition 1 ([6]). *Let $\mathcal{H} = (X, H)$ be a totally balanced hypergraph and $A, B, C \in H$. We have: $A \cap B \cap C \in \{A \cap B, A \cap C, B \cap C\}$.*

Indeed, for a weak hierarchical hypergraph $\mathcal{H} = (X, H)$ its closure is simply equal to the clusters $\{A \cap B \mid A, B \in H\}$, which can be done in $\mathcal{O}(|X|^3)$ operations. Moreover, $|H|$ is bound by $|X|^2$ (it derives directly from the closure property), thus:

- the number of formal concepts associated with a totally balanced formal context $K = (G, M, I)$ is bound by $\min(|G|, |M|)^2$,
- the formal concepts associated with a totally balanced formal context $n \times m$ matrix $\mathcal{M}$ are $(C_i \cap C_j, (C_i \cap C_j)^{\uparrow})$ with $1 \leq i, j \leq m$.

The fact that the intersections of two columns are sufficient to compute all the formal concepts, combined with the doubly lexical ordering of Section 2.2 will be the key of all the demonstrations given in Section 3.

2.2 Doubly lexical ordering of totally balanced Matrices

A doubly lexical ordering of an $n \times m$ binary matrix is an ordering such that if the rows and columns are viewed as n or m digit numbers read from right to left for lines and from bottom to top for columns, both rows and columns occur in increasing order. Clearly:

Proposition 2. *A n by m binary matrix $\mathcal{M}$ is* doubly lexically ordered *if the two following assertions are satisfied:*

- *for any $1 \leq j \leq m$ if there exist $1 \leq i < i' \leq n$ such that $\mathcal{M}_{i,j} = 1$ and $\mathcal{M}_{i',j} = 0$ then there exists $j' > j$ with $\mathcal{M}_{i,j'} = 0$ and $\mathcal{M}_{i',j'} = 1$,*
- *for any $1 \leq i \leq n$ if there exist $1 \leq j < j' \leq m$ such that $\mathcal{M}_{i,j} = 1$ and $\mathcal{M}_{i,j'} = 0$ then there exists $i' > i$ with $\mathcal{M}_{i',j} = 0$ and $\mathcal{M}_{i',j'} = 1$.*

Any $n \times m$ binary matrix can be doubly lexically ordered. Several polynomial algorithms to perform such an ordering exist, see for instance Spinrad [11] for a $\mathcal{O}(nm)$ algorithm (linear in the size of $\mathcal{M}$). Table 3 is a doubly lexical ordering of Table 1.

This ordering allows another characterization of totally balanced matrices:

Table 3. Cross table associated with a doubly lexical ordering of the formal context matrix from Table 1.

	c_5	c_1	c_4	c_2	c_3	c_6	c_7	c_8
l_1	×		×		×			
l_5		×		×	×			
l_3			×	×	×	×		×
l_4							×	×
l_2						×	×	×

Theorem 2 ([1]). *The three following assertions are equivalent:*

1. *$\mathcal{M}$ is a totally balanced binary matrix,*
2. *there is a doubly lexical ordering of $\mathcal{M}$ which is gamma-free,*
3. *every doubly lexical ordering of $\mathcal{M}$ is gamma-free.*

Since the formal context from Table 1 is totally balanced, its ordering given in Table 3 is gamma-free. Note that a gamma-free ordering is not necessarily doubly lexically ordered. For instance Table 2 is gamma-free but not doubly lexically ordered. In fact the doubly lexical order "packs" the formal concepts together as we shall show it hereafter.

3 Formal concepts of totally balanced formal context matrices

We shall prove here that, given a totally balanced formal context $K = (G, M, I)$, one can associate to each formal concept (A, B) an element (i, j) of its doubly lexically ordered formal context matrix $\mathcal{M}$.

We will assume without any loss of generality that the formal context K is such that:

- $G = \{1, \ldots, n\}$,
- $M = \{1, \ldots, m\}$,
- its associated formal context matrix $\mathcal{M}$ such that $\mathcal{M}_{i,j} = 1$ whenever iIj is doubly lexically ordered.

If $\mathcal{B}(G, M, I)$ is the set of all the formal concepts, we denote by $\mathring{\mathcal{B}}(G, M, I)$ the set of all the *proper formal concepts*: $\mathring{\mathcal{B}}(G, M, I) = \mathcal{B}(G, M, I) \backslash \{(\emptyset, G), (M, \emptyset)\}$. We define $f : \mathring{\mathcal{B}}(G, M, I) \to G \times M$ and $g : G \times M \to 2^G \times 2^M$ by:

$$f((A, B)) = (\min(A), \min(B))$$
$$g((i, j)) = (\{i' \mid \mathcal{M}_{i',j} = 1, i' \geq i\}, \{j' \mid \mathcal{M}_{i,j'} = 1, j' \geq j\})$$

Consider for instance the doubly lexically ordered formal context of Table 3. In order to fit the above definitions, we have to write its associated formal concepts (see Figure 1) according to the lines and the columns indices, thus the formal concept $(\{l_2, l_3\}, \{c_6, c_8\})$ (number 6 in Figure 1) is here equal to $((3, 5), (6, 8))$. We have then:

- $f(((3,5),(6,8))) = (3,6)$,
- $g((3,6)) = ((3,5),(6,8))$

Let $\mathcal{M}$ be an $n \times m$ doubly lexically ordered totally balanced formal context matrix. We denote by $\mathcal{B}(\mathcal{M})$ the set of all formal representative of $\mathcal{M}$. A pair (i,j) of $G \times M$ is a *formal representative* if:

- $\mathcal{M}_{i,j} = 1$,
- either $i = 1$ or there exists $j' \geq j$ such that $\mathcal{M}_{i,j'} = 1$ and $\mathcal{M}_{i-1,j'} = 0$,
- either $j = 1$ or there exists $i' \geq i$ such that $\mathcal{M}_{i',j} = 1$ $\mathcal{M}_{i',j-1} = 0$.

The goal of this section is to establish links between formal representatives and formal concepts. We first need some Lemmas. Note that Lemmas 1 and 2 only suppose that the formal context matrix is gamma-free, only Lemma 3 requires that $\mathcal{M}$ sould be doubly lexically ordered.

Lemma 1. *Let $\mathcal{M}$ be a $n \times m$ gamma-free ordered formal context matrix associated with a formal context $F = (G, M, I)$ and $\mathring{\mathcal{B}}(G, M, I)$ the set of its associated proper formal concepts. For all $(A, B) \in \mathring{\mathcal{B}}(G, M, I)$, $g(f((A, B))) = (A, B)$.*

Proof. We will first prove that there exist j_A and i_B such that $A = \{i \mid \mathcal{M}_{i,j_A} = 1, i \geq min(A)\}$ and $B = \{j \mid \mathcal{M}_{i_B j} = 1, j \geq min(B)\}$.

According to Proposition 1, A is either a column $C_j = \{\mathcal{M}_{i,j} = 1 \mid i \geq 1\}$ or an intersection of two columns $C_j \cap C_{j'}$ and one can assume that $j < j'$. If $A = C_j$ then $A = \{i' \mid \mathcal{M}_{i',j} = 1 i' \geq 1\} = \{i' \mid \mathcal{M}_{i',j} = 1, i' \geq min(A)\}$. If $A = C_j \cap C_{j'}$ then $\min(A)$ is the smallest line number i for which $\mathcal{M}_{i,j} = \mathcal{M}_{i,j'} = 1$ and since $\mathcal{M}$ is gamma-free, $\mathcal{M}_{i',j} = 1$ implies $\mathcal{M}_{i',j'} = 1$ for all $i' \geq i$: $A = C_j \cap C_{j'} = \{i' \mid \mathcal{M}_{i',j} = 1, i' \geq min(A)\}$.

Since ${}^t\mathcal{M}$ is also gamma-free, the same proof leads to the fact that there exists i such that $B = \{j' \mid \mathcal{M}_{i,j'} = 1, j' \geq min(B)\}$.

Now, since $\mathcal{M}_{i,j_A} = 1$ for any $i \in A$ we have that $j_A \in B$ (because $A^{\uparrow} = B$) thus $\min(B) \leq j_A$. Moreover, $\mathcal{M}_{i,\min(B)} = 1$ for all $i \in A$ (because $B^{\downarrow} = A$) thus $\{i \mid \mathcal{M}_{i,\min(B)} = 1, i \geq min(A)\} = \{i \mid \mathcal{M}_{i,j_A} = 1, i \geq min(A)\}$ (because $\mathcal{M}$ is gamma-free). This proves that $A = \{i \mid \mathcal{M}_{i,\min(B)} = 1, i \geq min(A)\}$.

The same kind of proof can be done to prove that $B = \{j \mid \mathcal{M}_{\min(A),j} = 1, j \geq min(B)\}$ which concludes the proof $\square$

Lemma 1 shows that f is an injection and that $g = f^{-1}$ for its image. The following two Lemmas will characterize the image of f.

Lemma 2. *Let $\mathcal{M}$ be a $n \times m$ gamma-free ordered formal context matrix associated with a formal context $F = (G, M, I)$ and $\mathring{\mathcal{B}}(G, M, I)$ the set of its associated proper formal concepts. For any $(A, B) \in \mathring{\mathcal{B}}(G, M, I)$:*

- *either* $\min(B) = 1$ *or there exists* $j \geq min(B)$ *such that* $\mathcal{M}_{\min(A),j-1} = 0$,
- *either* $\min(A) = 1$ *or there exists* $i \geq min(A)$ *such that* $\mathcal{M}_{i-1,\min(B)} = 0$.

Proof. We will only prove the first assertion, the second one follows by using ${}^t\mathcal{M}$ instead of $\mathcal{M}$.

Lemma 1 shows that $A = \{i \mid \mathcal{M}_{i,\min(B)} = 1, i \geq \min(A)\}$ and according to Proposition 1, A is either a column $C_j = \{\mathcal{M}_{i,j} = 1 \mid i \geq 1\}$ or an intersection of two columns $C_j \cap C_{j'}$ and one can assume that $j < j'$.

Columns C_j are equal to $\{i \mid \mathcal{M}_{i,j} = 1, i \geq i(C_j)\}$ with $i(C_j) = \min\{i \mid \mathcal{M}_{i,j} = 1\}$. Thus either $i(C_j) = 1$ or $\mathcal{M}_{i(C_j)-1,j} = 0$.

Intersection $C_j \cap C_{j'}$ with $1 \leq j < j' \leq m$. Let $i(C_j \cap C_{j'}) = \min\{i \mid \mathcal{M}_{i,j} = \mathcal{M}_{i,j'} = 1\}$. Since $\mathcal{M}$ is gamma-free $\mathcal{M}_{i',j} = 1$ implies $\mathcal{M}_{i',j'} = 1$ for all $i' \geq i(C_j \cap C_{j'})$, thus $C_j \cap C_{j'} = \{i \mid \mathcal{M}_{i,j} = 1, i \geq i(C_j \cap C_{j'})\}$. Note that if $i(C_j \cap C_{j'}) > 1$ then either $\mathcal{M}_{i(C_j \cap C_{j'})-1,j} = 0$ or $\mathcal{M}_{i(C_j \cap C_{j'})-1,j'} = 0$. The above two cases show that if $\min(B) > 1$ and $\mathcal{M}_{\min(A),j-1} = 1$ for all $j \geq \min(B)$ then $A = \{i \mid \mathcal{M}_{i,\min(B)} = 1, i \geq \min(A)\}$ cannot be an intersection of columns thus $(A, A^{\uparrow})$ is not a formal concept, which is impossible. □

Lemma 3. *Let $\mathcal{M}$ be an $n \times m$ doubly lexically ordered totally balanced formal context matrix associated with a formal context $F = (G, M, I)$. Let (i, j) a pair such that:*

- $\mathcal{M}_{i,j} = 1$,
- *either $i = 1$ or there exists $j' \geq j$ such that $\mathcal{M}_{i,j'} = 1$ and $\mathcal{M}_{i-1,j'} = 0$,*
- *either $j = 1$ or there exists $i' \geq i$ such that $\mathcal{M}_{i',j} = 1$ $\mathcal{M}_{i',j-1} = 0$.*

We have $g((i, j)) \in \mathring{\mathcal{B}}(G, M, I)$.

Proof. Let $\mathcal{C}$ be the closure under intersection of the matrix columns $C_j = \{i \mid \mathcal{M}_{i,j} = 1\}$. We will prove by induction on the column number that if $\mathcal{M}_{i,j} = 1$ and $\mathcal{M}_{i-1,j} = 0$ then $\{i' \mid \mathcal{M}_{i',j} = 1, i' \geq i\} \in \mathcal{C}$. First consider the last column. Since $\mathcal{M}$ is doubly lexically ordered one cannot have $\mathcal{M}_{i,m} = 1$ and $\mathcal{M}_{i',m} = 0$ with $1 \leq i < i' \leq n$. The last column is then either empty or there exists i_m for which $\mathcal{M}_{i,m} = 1$ if and only if $i \geq i_m$ thus the only possible pair (i, m) for which $\mathcal{M}_{i,m} = 1$ and $\mathcal{M}_{i-1,m} = 0$ is (i_m, m) and $\{i' \mid \mathcal{M}_{i',m} = 1, i' \geq i_m\} = C_m \in \mathcal{C}$. Suppose the property true for any $j' > j_0$ and let j be the largest column smaller or equal to j_0 such that there exists i for which $\mathcal{M}_{i,j} = 1$ and $\mathcal{M}_{i-1,j} = 0$. two cases may occur:

- $\mathcal{M}_{i',j} = 0$ for all $i' < i$. In this case $\{i' \mid \mathcal{M}_{i',j} = 1, i' \geq i\} = C_j \in \mathcal{C}$,
- otherwise let $i_2 < i$ be the largest index such that $\mathcal{M}_{i_2,j} = 1$. Since $\mathcal{M}$ is doubly lexically ordered, there exists $j_2 > j$ such that $\mathcal{M}_{i_2,j_2} = 0$ and $\mathcal{M}_{i-1,j_2} = 1$. Two sub-cases are possible:
  - $\mathcal{M}_{i,j_2} = 0$. Since $\mathcal{M}$ is doubly lexically ordered there exists $j_3 > j_2$ such that $\mathcal{M}_{i-1,j_3} = 0$ and $\mathcal{M}_{i,j_3} = 1$. The induction hypothesis holds, so $\{i' \mid \mathcal{M}_{i',j_3} = 1, i' \geq i\} \in \mathcal{C}$. Because $\mathcal{M}$ is gamma-free, $\{i' \mid \mathcal{M}_{i',j} = 1, i' \geq i\} = C_j \cap \{i' \mid \mathcal{M}_{i',j_3} = 1, i' \geq i\} \in \mathcal{C}$.

- $\mathcal{M}_{i,j_2} = 1$. There exists $i_3 < i$ for which $\mathcal{M}_{i',j_2} = 1$ for any $i_3 \leq i' \leq i$ and $\mathcal{M}_{i_3-1,j_2} = 0$. The induction hypothesis holds, so $\{i' \mid \mathcal{M}_{i',j_2} = 1, i' \geq i_3\} \in \mathcal{C}$. As $\mathcal{M}$ is gamma-free, $\{i' \mid \mathcal{M}_{i',j} = 1, i' \geq i\} = C_j \cap \{i' \mid \mathcal{M}_{i',j_2} = 1, i' \geq i_3\} \in \mathcal{C}$.

In all cases $\{i' \mid \mathcal{M}_{i',j} = 1, i' \geq i\} \in \mathcal{C}$. This concludes the proof by induction.

Let now (i, j) be a pair as stated in the Lemma. If $i = 1$ then $A = \{i' \mid \mathcal{M}_{i',j} = 1, i' \geq i\} \in \mathcal{C}$ and if $i > 1$ there exists $j' \geq j$ such that $\mathcal{M}_{i,j'} = 1$ and $\mathcal{M}_{i-1,j'} = 0$. The above induction demonstration states that $\{i' \mid \mathcal{M}_{i',j'} = 1, i' \geq i\} \in \mathcal{C}$ thus $A = \{i' \mid \mathcal{M}_{i',j} = 1, i' \geq i\} = C_j \cap \{i' \mid \mathcal{M}_{i',j'} = 1, i' \geq i\} \in \mathcal{C}$.

Since $\mathcal{M}$ is gamma-free, $\mathcal{M}_{i,j'} = 1$ with $j' \geq j$ implies that $\mathcal{M}_{i',j'} = 1$ for all $i' \in A$ thus $B = \{j' \mid \mathcal{M}_{i,j'} = 1, j' \geq j\} \subseteq A^{\uparrow}$. If $j = 1$ we clearly have equality. If $j > 1$ let i' be the largest line larger than i for which $\mathcal{M}_{i',j} = 1$ $\mathcal{M}_{i',j-1} = 0$. If there exists $j' < j$ such that $j' \in A^{\uparrow}$, then $\mathcal{M}_{i',j'} = 1$. Since $\mathcal{M}$ is doubly lexically ordered and $\mathcal{M}_{i',j'} = 1$ and $\mathcal{M}_{i',j-1} = 0$ there exists $i'' > i'$ for which $\mathcal{M}_{i'',j'} = 0$ and $\mathcal{M}_{i'',j-1} = 1$ which implies $\mathcal{M}_{i'',j} = 0$ because $j' \in A^{\uparrow}$. The doubly lexical ordering of $\mathcal{M}$ then states that there exists $i''' > i''$ for which $\mathcal{M}_{i''',j-1} = 0$ and $\mathcal{M}_{i''',j} = 1$ which is impossible by maximality of i' so such a j' does not exist.

We finally have that $A = \{i' \mid \mathcal{M}_{i',j} = 1, i' \geq i\} \in \mathcal{C}$ and $A^{\uparrow} = \{j' \mid \mathcal{M}_{i,j'} = 1, j' \geq j\}$ which concludes the proof. □

The hereunder Proposition 3 directly follows from Lemmas 1, 2 and 3.

Proposition 3. *The sets $\mathring{\mathcal{B}}(G, M, I)$ and $\mathcal{B}(\mathcal{M})$ are in one-to-one correspondence:*

- *for all $(A, B) \in \mathring{\mathcal{B}}(G, M, I)$: $f((A, B)) = (\min(A), \min(B)) \in \mathcal{B}(\mathcal{M})$,*
- *for all $(i, j) \in \mathcal{B}(\mathcal{M})$: $g((i, j)) = (\{i' \mid \mathcal{M}_{i',j} = 1, i' \geq i\}, \{j' \mid \mathcal{M}_{i,j'} = 1, j' \geq j\}) \in \mathring{\mathcal{B}}(G, M, I)$,*
- *$g \circ f((A, B)) = (A, B)$,*
- *$f \circ g((i, j)) = (i, j)$.*

Table 4 shows the proper formal concepts of the formal context of Table 1 with the reordering of Table 3. Note that it is easy to find all the pairs of $\mathcal{B}(\mathcal{M})$ in $\mathcal{O}(nm)$ (the size of the formal context matrix) operations.

Proposition 3 and the fact that the formal context matrix is gamma-free allow us to state Proposition 4.

Proposition 4.

- *$(A, B) \leq (A', B')$ implies that $\min(A) \leq \min(A')$ and $\min(B) \leq \min(B')$: the formal concepts larger that a given (A, B) are associated with a pair in the top-right corner of $f((A, B))$,*
- *let $(i, j), (i', j') \in \mathcal{B}(\mathcal{M})$. $g((i, j)) \leq g((i', j'))$ is equivalent to $\mathcal{M}_{i,j'} = 1$,*
- *let $(i, j), (i', j') \in \mathcal{B}(\mathcal{M})$. $g((i, j)) \geq g((i', j'))$ is equivalent to $\mathcal{M}_{i',j} = 1$,*

Table 4. Formal concepts of Figure 1 represented by their associated pair of the doubly lexical ordering of the formal context from Table 3.

	c_5	c_1	c_4	c_2	c_3	c_6	c_7	c_8
l_1	(2)		(7)		(10)			
l_5		(4)		(8)	$\times$			
l_3			(3)	$\times$	$\times$	(6)		(9)
l_4							(5)	$\times$
l_2						(1)	$\times$	$\times$

0	$(\emptyset, \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8\})$
1	$(\{\mathbf{l_2}\}, \{c_6, c_7, c_8\})$
2	$(\{\mathbf{l_1}\}, \{c_3, c_4, \mathbf{c_5}\})$
3	$(\{\mathbf{l_3}\}, \{c_2, c_3, c_4, c_6, c_8\})$
4	$(\{\mathbf{l_5}\}, \{\mathbf{c_1}, c_2, c_3\})$
5	$(\{l_2, \mathbf{l_4}\}, \{\mathbf{c_7}, c_8\})$
6	$(\{l_2, l_3\}, \{\mathbf{c_6}, c_8\})$
7	$(\{l_1, l_3\}, \{c_3, \mathbf{c_4}\})$
8	$(\{l_3, l_5\}, \{\mathbf{c_2}, c_3\})$
9	$(\{l_2, l_3, l_4\}, \{\mathbf{c_8}\})$
10	$(\{l_1, l_3, l_5\}, \{\mathbf{c_3}\})$
11	$(\{l_1, l_2, l_3, l_4, l_5\}, \emptyset)$

Proposition 4 gives a way to show all the formal concepts of a given totally balanced formal context and their order relationship on a special ordering of its associated formal context matrix. We shall extend this property in the next section by showing that the whole formal concept lattice can be embedded into the matrix.

4 Cover graphs and graphical representation

In this section, we show that the formal representatives can be associated to non-overlapping boxes of the formal context matrix. This also represents the directed cover graph.

Let $\mathcal{M}$ be a $n \times m$ doubly lexically ordered totally balanced formal context matrix associated with a formal context $F = (G, M, I)$, $\mathring{\mathcal{B}}(G, M, I)$ the set of all its proper formal concepts and $\mathcal{B}(\mathcal{M})$ the set of formal representatives.

We associate the pair $(n+1, 1)$ to the formal concept $(\emptyset, \{1, \ldots, m\})$ and the pair $(1, m+1)$ to the formal concept $(\{1, \ldots, n\}, \emptyset)$. So we extend maps f and g to all the formal concepts $\mathcal{B}(G, M, I)$. Thus we can draw the directed cover graph directly on $\mathcal{M}$ as shown in Figure 2. For the rest of this section, we shall then consider that $g((n+1, 1)) = (\emptyset, \{1, \ldots, m\})$ and that $g((1, m+1)) = (\{1, \ldots, n\}, \emptyset)$.

By Proposition 3, for such a figure, the edges $((u_i, u_j), (v_i, v_j))$ of the directed cover graph are such that $u_i \geq v_i$ and $u_j \leq v_j$ and at least one of these inequalities is strict.

We shall precise this drawing by using all the 1s of the matrix. Let $(i, j) \in \mathcal{B}(\mathcal{M})$, we define:

- $n(i, j)$: the smallest column such that $n(i, j) > j$ and for which there exists $i' \geq j$ with $\mathcal{M}_{i', n(i,j)} = 1$ and $\mathcal{M}_{i', n(i,j)-1} = 0$. If i' does not exist, $n(i, j) = m + 1$.

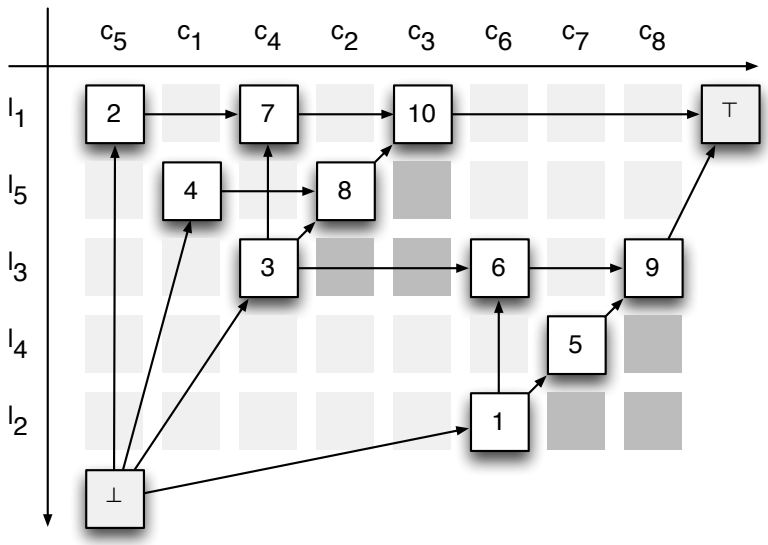


Fig. 2. Directed cover graph of the totally balanced formal context of Table 3. The unmarked 1 of the matrix are in dark grey.

– $w(i,j)$: the largest column such that $j \leq w(i,j) < n(i,j)$ and $\mathcal{M}_{i,j''} = 1$ for all $j \leq j'' \leq w(i,j)$. Note that $w(i,j)$ always exists (it can be equal to j).

It is clear that, given two formal representatives (i,j) and (i',j'), the intersection of the two intervals $[j, w(i,j)]$ and $[j', w(i',j')]$ is either empty or one of them. The set of all the intervals is a hierarchy. Moreover, if $[j, w(i,j)] \not\subset [j', w(i',j')]$ then $i < i'$.

Figure 3 shows the hierarchy formed by the formal concepts of the totally balanced formal context of Table 3. Note that this hierarchy can be directly drawn on the doubly lexically ordered context matrix and that it is a part of the associated directed cover graph.

Since ${}^t\mathcal{M}$ is also doubly lexically ordered and gamma-free, if we note ${}^tw(i,j)$ the element of $\mathcal{M}$ associated with $w(j,i)$ in ${}^t\mathcal{M}$, there is a also a hierarchy formed by the intervals $[i, {}^t w(i,j)]$. Figure 4 shows this hierarchy for the formal concepts of the totally balanced formal context of Table 3.

Note that the union of the two hierarchies forms the directed cover graph. Some of the 1s of the matrix ${}^t\mathcal{M}$ are not "used". The hereunder procedure combines the two drawings into a unique one. Given a formal representative (i,j), the box formed by the two intervals $[i, {}^t w(i,j)] \times [j, w(i,j)]$ is full of 1s in $\mathcal{M}$ and all the 1s of $\mathcal{M}$ are at least in one box. If we do not want to allow boxes overlapping we have to favor one dimension upon the other. In the hereunder construction we favor $[j, w(i,j)]$ upon $[i, {}^t w(i,j)]$.

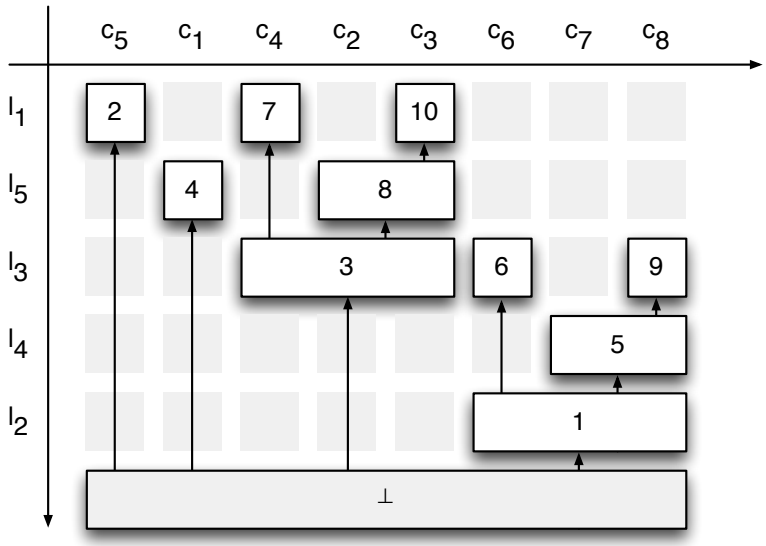


Fig. 3. Interval hierarchy of the columns of the formal concepts of the formal context of Table 3.

We associate to each $1 \leq i \leq n$, $1 \leq j \leq m$ such that $\mathcal{M}_{i,j} = 1$ a formal concept $g((i', j'))$:

- where $i' = i$ and j' is such that there exists $(i, j') \in \mathcal{B}(\mathcal{M})$ for which $j \leq j' < w(i, j')$,
- or where $i < i'$ and i' the largest element such that there exists $(i', j') \in \mathcal{B}(\mathcal{M})$ for which $j \leq j' < j + w(i', j')$.

The above construction ensures that each formal concept is represented by two intervals forming a "box full of 1s" in $\mathcal{M}$ and that all these boxes never overlap. Note that the above association is unique and that each formal representative is self-associated. This association can be easily done in $\mathcal{O}(nm)$ operations.

In addition, one can prove that: $g((i, j)) \prec g((i', j'))$ if their associated boxes in $\mathcal{M}$ are neighbors. One can then draw the associated directed cover graph of the formal context on the matrix as shown in Figure 5.

5 Conclusion

This paper gives a way of representing, for a given totally balanced formal context, the context matrix, the directed cover graph and its concepts into a unique representation. This can be performed linearly in the size of the matrix allowing to address big data sets.

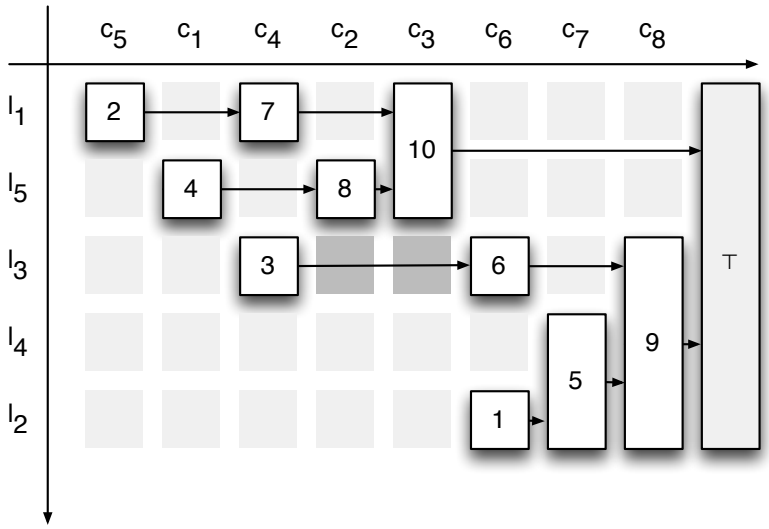


Fig. 4. Interval hierarchy of the lines of the formal concepts of the formal context of Table 3.

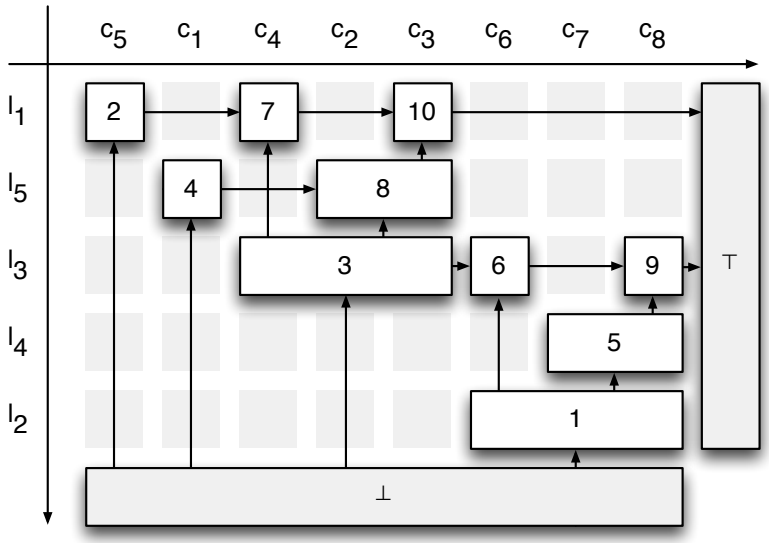


Fig. 5. Directed cover graph embedded into the formal context matrix of Table 4.

Spinrad [12] gives a $\mathcal{O}(nm)$ approximation scheme to transform a non-gamma-free $n \times m$ binary matrix into a gamma-free one. Since a gamma-free matrix admits a gamma-free doubly lexical order, one can linearly approximate a formal context matrix into a totally balanced one. This can be done for large datasets as a first step of a Data Analysis process. It allows to find potential areas of interest before searching for all the formal concepts of the original context in this area.

Figure 6 shows a way of presenting such data by coloring the different concepts. One can clearly see the concepts and their relationships despite the size of the matrix. All the algorithms have been implemented and there is a ongoing study to apply the on big datasets.

Fig. 6. Approximation of a 50×170 Matrix.

References

1. Richard Antsee and Martin Farber: Characterizations of totally balanced matrices. Journal of Algorithms 5, 215-230 (1984).
2. Claude Berge: Sur certains hypergraphes généralisant les graphes bipartites. Combinatorial Theory and its Applications I, Colloquium of Mathematical Society Janos Bolyai 4, 119–133 (1970)
3. Hans-Jürgen Bandelt and Andreas Dress: Weak hierarchies associated with similarity measures – an additive clustering technique. Bulletin of Mathematical Biology 51, 133-166, (1989).
4. Claude Barbut and Bernard Monjardet: *Ordre et classification, algèbre et combinatoire* Hachette, Paris (1970)
5. François Brucker and Alain Gély: Crown-free Lattices and Their Related Graphs. Order 28, 443-454 (2009)
6. François Brucker and Alain Gély: Parsimonious cluster systems. Advances in Data Analysis and Classification 3, 189-204 (2009)
7. Martin Farber: Characterizations of strongly chordal graphs. Discrete Mathematics 43, 173–189 (1983).

8. Bernhard Ganter and Rudolf Wille.: *Formal Concept Analysis*. Springer, Heidelberg (1999)
9. Jenö Lehel: A characterization of totally balanced hypergraphs. Discrete Mathematics 57, 59-65 (1985)
10. Ivan Rival: Lattices with doubly irreducible elements. Canadian Mathematical Bulletin 17(1), 91–95, (1974)
11. Jeremy Spinrad: Doubly Lexical Ordering of Dense 0/1 Matrices. Information Processing Letters 45, 229-235 (1993)
12. Jeremy Spinrad: *Efficient Graph representations* American Mathematical Society, 2003.