Journal of Universal Computer Science, vol. 10, no. 8 (2004), 934-954
submitted: 22/3/04, accepted: 28/6/04, appeared: 28/8/04 © *J.UCS*

Breadth First Search Graph Partitions and Concept Lattices

James Abello
(DIMACS, Rutgers University, USA
abello@dimacs.rutgers.edu)

Alex J. Pogel
(New Mexico State University, USA
Alex.Pogel@psl.nmsu.edu)

Lance Miller
(University of Connecticut, USA
lmiller@math.uconn.edu)

Abstract: We apply the graph decomposition method known as *rooted level aware breadth first search* to partition graph-connected formal contexts and examine some of the consequences for the corresponding concept lattices. In graph-theoretic terms, this lattice can be viewed as the lattice of maximal bicliques of the bipartite graph obtained by symmetrizing the object-attribute pairs of the input formal context. We find that a rooted breadth-first search decomposition of a graph-connected formal context leads to a closely related partition of the concept lattice, and we provide some details of this relationship. The main result is used to describe how the concept lattice can be unfolded, according to the information gathered during the breadth first search. We discuss potential uses of the results in data mining applications that employ concept lattices, specifically those involving association rules.

Key Words: Formal Concept Analysis, Bipartite Graph, Breadth First Search
Category: G.2.2, G.2.3

1 Introduction

In order to decompose lattices that appear in a variety of data analysis applications, we examine a graph-theoretic decomposition method, namely level-aware breadth first search, and determine some of its connections with lattices. The key step in connecting this graph-theoretic method with lattices is to relate each (lattice-generating) binary relation to an undirected bipartite graph. When the obtained bipartite graph is a connected graph, its lattice of maximal bicliques can be interpreted as the concept lattice (cf. [Ganter and Wille (99)]) of the binary relation. This lattice amounts to an organization of the tabular data, which is used for Knowledge Discovery in Databases [Wille 01, FCA URL], e.g., involving the examination of concepts, implications and association rules present in the data. Given the computational complexity and time requirements of association rule mining [Agrawal et al. 93] and the connection of association rules with the concept lattice [Zaki et al. 98], it is imperative to design efficient algorithms that can focus their search into potentially interesting regions of the lattice.

An important step of lattice-centered data analysis involves viewing a lattice on a computer monitor, as hand-calculations are only reasonable in the smallest of examples – see [Freese, ConExp] for some automated lattice drawing tools. However, the size of the lattices can grow quickly and this presents two challenges. The first issue is whether the automated drawing tools can provide a diagram at all, since the underlying algorithms are often at least quadratic in the number of concepts. Assuming that a diagram can be created in a reasonable amount of time, the second issue arises from the observation that hundreds, or thousands, of concepts in a diagram immediately tax the human viewer's ability to absorb information from the diagram. More generally, the concept lattice computation simply generates too large a number of concepts to easily manage, regardless of whether the concepts are viewed or not. These complexity concerns create a need for some control to be exercised with regard to how many concepts are computed at a time, a problem discussed in [Stumme et al. 02] and [Berry and Sigayret 2-02]. These issues have led us to examine various decomposition methods, along the lines of [Abello and Korn 02], to apply to the given input binary relation that is the usual initial datum for the concept lattice construction. In this paper we discuss the Level-Aware Breadth First Search through the binary relation. We consider some of the theoretical aspects related to the use of this graph-theoretic decomposition method, specifically to induce partitions of the corresponding concept lattice. In the sense that we apply graph-theoretic methods to concept analysis, this work is similar to [Berry and Sigayret 1-02].

An important aspect of this method is that it determines an inexpensively computed decomposition of the input data (the binary relation), which we foresee will at times help to organize the computation and search of the concept lattice. This strategy should be understood in contrast with approaches that use general properties of the structure of the input data. For example, [Stumme et al. 02] incorporates the usual support thresholding used in association rule mining into an algorithm for computing generators (key sets) for frequent closed itemsets, using a pruning method [Agrawal et al. 93]; the general property used there is that the set of all key sets is an order ideal of the power set of the attribute set. In summary, we see the Level-Aware Breadth First Search as one possible method of offering glimpses of the full concept lattice, by allowing portions of the lattice to be viewed independently. This paper presents initial structural results, which we expect will provide support for further advances in this direction.

The outline of the paper is as follows. After introducing some notions from graph theory and Formal Concept Analysis, we prove the main results regarding the relationship between the Level-Aware Breadth First Search decomposition of a formal context and the concept lattice of that formal context. This is followed by a description of the manner in which the concept lattice can be computed and visualized in steps. The final section suggests directions for further research.

2 Definitions

Although some of the definitions appearing throughout this section do not require that the sets involved be finite, we make a standing assumption that all sets under consideration are finite.

2.1 Graph Theoretical Notions

In this subsection we introduce the necessary graph theoretical terminology.

Definition: A (loopless) *graph* $\mathfrak{G}$ is a pair (V, E) such that V is a nonempty set and E is a subset of $\mathcal{P}_2(V)$, the set of all two-element subsets of V; elements of V are called *vertices*, while elements of E are called *edges*, and the two vertices associated with a particular edge are called the *endpoints* of the edge.

Definition: If x is a vertex in a graph $\mathfrak{G} = (V, E)$, its *neigborhood* is

$$N(x) = \{y \in V : \{x, y\} \in E\}.$$

The *subgraph induced* by a subset S of V is the graph $\mathfrak{G}_S$ whose edge set E_S consists of those edges with both endpoints in S.

Definition: A *bipartite graph (bigraph)* is a graph $\mathfrak{G} = (V, E)$ for which there exists a non-trivial partition $\{V_L, V_R\}$ of V such that for each $e \in E$,

$$e \cap V_L \neq \emptyset \quad \textit{and} \quad e \cap V_R \neq \emptyset.$$

In words, V is partitioned into two independent nonempty sets: each edge connects an element in one block of the partition to an element in the other block. We call V_L the set of *left vertices*, and we call V_R the set of *right vertices.* If E contains all possible edges between V_L and V_R, $\mathfrak{G}$ is called a complete bipartite graph. A complete bipartite graph $\mathfrak{G}_S$ that is an induced subgraph of $\mathfrak{G}$ is called a *biclique* of $\mathfrak{G}$. A biclique is *maximal* if it is not contained in a larger biclique.

2.2 Formal Concept Analysis Notions

We follow the definitions introduced in [Ganter and Wille (99)], and repeat a few here, especially when there are graph-theoretical interpretations of interest.

Definition: A *formal context* is a triple $\mathbb{K} = (O, A, E)$ of nonempty sets, where $E \subseteq O \times A$. If a formal context $\mathbb{K}$ satisfies $O \cap A = \emptyset$, then we say it is *bigraph inducing*. The bipartite graph $\mathfrak{G}_K$ of a bigraph inducing formal context $\mathbb{K} = (O, A, E)$ is

$$\mathfrak{G}_K = (O \cup A, \{ \{o, a\} : oEa \}) .$$

The edge set of the bigraph $\mathfrak{G}_K$ is called the *symmetrization* of the binary relation E.

Definition: Given a formal context (O, A, E), for each subset P of O and for each $H \subseteq O \times A$, we define the operator $(\ )^H$ on $\mathcal{P}(O)$ and $\mathcal{P}(A)$, as follows: for $P \subseteq O$, let

$$P^H = \{a \in A : \forall p \in P, pHa \}$$

and dually, for $B \subseteq A$, let

$$B^H = \{o \in O : \forall b \in B, oHb \}.$$

When a formal context $\mathbb{K} = (O, A, E)$ is fixed in a discussion, we write P' in place of P^E, and B' in place of B^E. Also, P'' is shorthand for $(P')'$.

Suppose $\mathbb{K}$ is bigraph inducing. Then, using graph terminology regarding the graph $\mathfrak{G}_K$, we see that $P' = \bigcap_{p \in P} N(p)$, i.e. P' is the intersection of the $\mathfrak{G}_K$-neigbourhoods of all the elements in P. Further, if both P and P' are nonempty, then $P \cup P'$ is the vertex set of a biclique in the bipartite graph $\mathfrak{G}_K$. If both P'' and P' are nonempty, then the union $P'' \cup P'$ is the vertex set of a maximal biclique of $\mathfrak{G}_K$ and every maximal biclique of $\mathfrak{G}_K$ arises this way, for some subset P of O. Dually, $B' = \bigcap_{a \in B} N(a)$ is the intersection of the $\mathfrak{G}_K$-neigbourhoods of all the elements in B. Similar comments regarding cliques and bicliques of the bipartite graph $\mathfrak{G}_K$ apply here as well.

The elements of the concept lattice associated with $\mathbb{K} = (O, A, E)$ are the pairs $(P, B) \in \mathcal{P}(O) \times \mathcal{P}(A)$ such that $P' = B$ and $B' = P$; such pairs are called *concepts* of the formal context $\mathbb{K}$, and P is called the *extent* and B the *intent* of the concept. In graph terminology, a formal concept of a bigraph inducing $\mathbb{K}$ with nonempty intent B and nonempty extent P will generate a maximal biclique with vertex set $P \cup B$ in the bigraph $\mathfrak{G}_K$. Let $\mathfrak{B}(\mathbb{K})$ be the set of all concepts of the formal context $\mathbb{K}$, ordered by inclusion in the first coordinate, i.e. $(P, B) \leq (Q, C)$ if and only if $P \subseteq Q$. This ordering makes $(\mathfrak{B}(\mathbb{K}), \leq)$ a complete lattice, i.e. a partially ordered set $(L, \leq)$ in which every subset of L has a least upper bound and a greatest lower bound in L. Thus we call the ordered set $(\mathfrak{B}(\mathbb{K}), \leq)$ *the concept lattice* of $\mathbb{K} = (O, A, E)$, and we usually omit the ordering $\leq$ and denote the lattice simply by $\mathfrak{B}(\mathbb{K})$.

Given a formal context $\mathbb{K} = (O, A, E)$, if $P \subseteq O$ and $B \subseteq A$, then the formal context $\mathbb{K}_{P,B} = (P, B, E \cap (P \times B))$ is called a *subcontext* of $\mathbb{K}$. Each restriction of either the domain or codomain to a proper subset induces maps between concept lattices. In particular, Propositions 31 and 32 of [Ganter and Wille (99)] state that subsets $P \subseteq O$ and $B \subseteq A$ induce order embeddings

$$\mathfrak{B}(\mathbb{K}_{P,A}) \to \mathfrak{B}(\mathbb{K}), \quad \mathfrak{B}(\mathbb{K}_{O,B}) \to \mathfrak{B}(\mathbb{K}) \quad \textit{and} \quad \mathfrak{B}(\mathbb{K}_{P,B}) \to \mathfrak{B}(\mathbb{K})$$

such that the first map is $\bigvee$-preserving, the second is $\bigwedge$-preserving, and the third could be given by either mapping,

$$(X, Y) \mapsto (X'', X') \quad \textit{or} \quad (X, Y) \mapsto (Y', Y'')$$

but these need be neither $\bigvee$- nor $\bigwedge$-preserving. Since our key purpose is to decompose lattices in practice, a technical goal of this paper is to provide conditions on concepts of subcontexts (determined by the Level-Aware Breadth First Search) which imply that these operators are simply the identity map on such concepts. Stated in more intuitive terms, we seek conditions that are sufficient to ensure that concepts of the subcontext are “real”, i.e. that they are concepts of the full context $\mathbb{K}$.

2.3 The Undirected Bigraph of a Formal Context

Definition: A formal context $\mathbb{K} = (O, A, E)$ is *graph-connected* provided it is bigraph inducing and the binary relation $E \cup E^{-1}$ on $O \cup A$ is connected in the usual sense, i.e. for all $x, y \in O \cup A$, there exists a path from x to y using ordered pairs from $E \cup E^{-1}$.

We observe that if a bigraph inducing formal context $\mathbb{K} = (O, A, E)$ is graph-connected, then the induced bipartite graph $\mathfrak{G}_{\mathbb{K}}$ is connected. If a bipartite graph is connected, then the partition of its vertex set is unique, so in the case of a graph-connected formal context $\mathbb{K} = (O, A, E)$ such that $|O| \neq |A|$, we can recover $\mathbb{K}$ from $\mathfrak{G}_{\mathbb{K}}$. The important point is that if we induce a bigraph from a formal context, then various standard graph decomposition methods immediately come to mind and can be considered as a way to work around the complexity problem that is inherent to the formation of concepts.

2.4 Association Rules: Confidence and Support

We conclude this section with two important functions used in data mining activity involving association rules [Agrawal et al. 93]. As in Formal Concept Analysis, the input data involves binary attribute values assigned to a set of objects, i.e. a formal context. Given a set A of attributes, an association rule is a pair (X, Y) (often written $X \to Y$) with X, Y subsets of A, interpreted to say “in (some) cases where X holds, Y also holds” (near implication), or “in the event of X, event Y also occurs” (conditional event).

Two functions used to formulate evaluation criteria for association rules, and to control the size of sets of association rules that are created during data mining activity based on a formal context $\mathbb{K} = (O, A, E)$, are $conf_{\mathbb{K}}(-)$ (confidence) and $supp_{\mathbb{K}}(-)$ (support), given by

$$conf_{\mathbb{K}}(\ (X, Y)\ ) = \frac{|X' \cap Y'|}{|X'|} \qquad and \qquad supp_{\mathbb{K}}(\ (X, Y)\ ) = \frac{|X' \cap Y'|}{|O|}$$

Support outputs the percent of overall evidence in the formal context for which the rule is positively witnessed, while confidence outputs the percent of those instances in the formal context where the hypothesis holds for which the conclusion also holds (with the appropriate qualifications, this is clearly conditional probability).

Suppose we consider a grocery shopping context $\mathbb{G}$, where O is the set of shopping carts observed at checkout and A is the set of items the carts contained (e.g., cart #141 may have contained beer, diapers, pretzels and milk). If we consider the rule "beer → pretzels" and find that

$$supp_{\mathbb{G}}(beer \rightarrow pretzels) = 0.22 \qquad and \qquad conf_{\mathbb{G}}(beer \rightarrow pretzels) = 0.84$$

then of all the observed carts, 22% bought both beer and pretzels, and of those carts that contained beer, 84% of them also contained pretzels.

We define a function $Csupp_{\mathbb{K}} : \mathfrak{B}(\mathbb{K}) \rightarrow [0, 1]$, called the *concept support* function, by assigning to each concept $D = (P, B)$ of $\mathbb{K} = (O, A, E)$ the value $Csupp_{\mathbb{K}}(D) = \frac{|P|}{|O|}$. Then the support of an association rule (X, Y) is the concept support of the concept generated by $X \cup Y$, that is

$$supp_{\mathbb{K}}(\ (X, Y)\ ) = Csupp_{\mathbb{K}}((X', X'') \wedge (Y', Y'')) = Csupp_{\mathbb{K}}(\ (X' \cap Y', (X' \cap Y')')\ )\ .$$

Also note that the support of a valid implication $X \rightarrow Y$ (i.e. an association rule (X, Y) with 100% confidence) is the concept support of the concept (X', X'') generated by the premise X.

The connection between the computation of concepts and the derivation of association rules has been observed by many authors, e.g. [Stumme et al. 02, Zaki et al. 98], and we will provide some observations regarding the connection of graph decompositions with association rules in a later section.

3 Distance Partitions and Concept Lattices

Formal Concept Analysis includes a variety of decomposition and construction methods. Many of these constructions are lattice-theoretic or universal-algebraic in nature and origin. In this section, we discuss a traditional graph-theoretic decomposition method that has been successfully applied to provide an overview of sparse massive data sets [Abello et al. (02)]. After the Breadth First Search decomposition method on bigraphs is presented, it is then extended to a given formal context (O, A, E) by applying it to the symmetrization of the binary relation E. Finally, we present theorems regarding the relationship between the decomposition of the formal context and the decomposition of the concept lattice.

3.1 Level-Aware Breadth First Search, for a Bigraph

The *rooted level-aware breadth first search* (abbreviation: LABFS) decomposition of a connected bigraph fixes a vertex as a root and partitions the vertex set by graph-theoretic distance from the root. Given a connected bigraph $\mathfrak{G} = (V, F)$, we fix some $r \in V$ and consider the function

$$d_r : V \to \mathbb{N} \cup \{0\},$$

where $d_r(v)$ is the graph-theoretic distance from v to r (the minimum path length from v to r). Now we partition V by setting $L_i(r) = d_r^{-1}(\{i\})$, for each $i \in \mathbb{N} \cup \{0\}$. Clearly $V = \bigsqcup_{i \in N \cup \{0\}} L_i(r)$, where $\sqcup$ denotes disjoint union. The element r is called *the root* of the LABFS decomposition, and in general the induced partition of V will depend on r. While it is always true that $L_0(r) = \{r\}$, beyond that we cannot say much more about the partition. There are bigraphs with choices of r such that $V = L_0(r) \sqcup L_1(r)$ and there are bigraphs with choices of r such that

$$V = \bigsqcup_{i=0,1,2,\ldots,|V|-1} L_i(r) \ ,$$

where each $L_i(r)$ is nonempty. Note that all the partition blocks (and later, subrelations) that we consider in the sequel are dependent on the choice of r, so we will write expressions such as $L_1(r)$ as L_1.

The following statement is easy to prove:

Lemma 1. *Let $\mathfrak{G} = (V, F)$ be a connected bigraph, and let a root $r \in V$ be given. For $i \in \mathbb{N} \cup \{0\}$, let $L_i = d_r^{-1}(\{i\})$ and let $F_i = \{\{u, v\} : u \in L_i, v \in L_{i+1}\} \cap F$. Then*

1.

$$V = \bigsqcup_{i \in \mathbb{N} \cup \{0\}} L_i \quad \text{and} \quad F = \bigsqcup_{i \in \mathbb{N} \cup \{0\}} F_i,$$

2. If $\mathfrak{G} = (V, F)$ is a connected bigraph, say, with root r in the right vertex set V_R, then $\{L_{2j}\}_{j \in \mathbb{N} \cup \{0\}}$ is a partition of V_R and $\{L_{2j+1}\}_{j \in \mathbb{N} \cup \{0\}}$ is a partition of V_L.

3.2 LABFS for a Formal Context

From the LABFS partition of the induced connected bigraph $\mathfrak{G}_{\mathbb{K}}$ of a graph-connected formal context $\mathbb{K} = (O, A, E)$, we construct partitions of O and A and a covering of the binary relation E which will be used to understand large lattices by suitable smaller lattices.

For the rest of this section, we define *r-rooted formal context* $\mathbb{K}$ to mean that $\mathbb{K} = (O, A, E)$ is a bigraph inducing, graph-connected formal context with distinguished element $r \in A$, and we let $\mathfrak{G}_{\mathbb{K}}$ be the induced (connected) bipartite graph, specifically with left vertex set $V_L = O$ and right vertex set $V_R = A$, and with the ordered pairs in E converted to unordered pairs in $F = \{ \{o, a\} : oEa \}$. Given an r-rooted formal context $\mathbb{K}$, the partitions in Lemma 1 induce corresponding partitions of O and A, (via intersection with O and A respectively), all depending on the fixed choice of root r in A,

$$O = L_1 \sqcup L_3 \sqcup L_5 \sqcup ... \quad and \quad A = L_0 \sqcup L_2 \sqcup L_4 \sqcup ... \ ,$$

and Lemma 1 also implies that the relation E can be expressed via subrelations of E that are between blocks of the partitions of O and A.

Now, in place of the disjoint edge sets F_i in Lemma 1, we clearly have disjoint sets E_i of ordered pairs, where

$$E_i = \{ (o, a) : o \in O, a \in A, \{o, a\} \in F_i \}$$

but because of our interest in the concept lattice, we want to further define subrelations S_i of E, for $i \in \mathbb{N} \cup \{0\}$, by setting

$$S_i = (L_{i+1} \times (L_i \cup L_{i+2})) \ \cap \ E \qquad and \qquad S_i = ((L_i \cup L_{i+2}) \times L_{i+1}) \ \cap \ E$$

for i even and i odd, respectively. Note that these subrelations will not be disjoint. We call S_i the i^{th} *LABFS subrelation* of $\mathbb{K} = (O, A, E)$.

Consider the following graph-connected formal context $\mathbb{C}$ with 30 objects and 10 attributes, presented in tabular form at left in Figure 1. The object names have the form "o-i", for object i, and the attribute names have the form "a-j", for attribute j. If we choose root "a-1", and determine the sets corresponding to the various levels, and also shuffle the objects and attributes in the tabular presentation to reflect the levels, then the resulting tabular representation of $\mathbb{C}$ is shown at right in Figure 1.

	a-1	a-2	a-3	a-4	a-5	a-6	a-7	a-8	a-9	a-10
o-1	X			X			X			X
o-2				X						
o-3		X				X			X	
o-4	X									
o-5							X	X		
o-6			X				X			
o-7					X					
o-8				X					X	
o-9		X					X			
o-10		X	X	X	X			X		X
o-11	X									
o-12			X						X	
o-13						X				
o-14		X			X					
o-15				X					X	
o-16		X						X		
o-17				X						
o-18		X			X	X	X			X
o-19	X							X		X
o-20		X								
o-21			X			X				
o-22				X						
o-23		X					X			X
o-24	X									
o-25								X		
o-26		X			X					
o-27						X			X	
o-28			X	X					X	
o-29						X				X
o-30			X					X		

	a-1	a-7	a-4	a-10	a-8	a-3	a-2	a-5	a-6	a-9
o-1	X	X	X	X						
o-4	X									
o-11	X									
o-19	X			X	X					
o-24	X									
o-5		X			X					
o-6		X				X				
o-9		X					X			
o-18		X		X			X	X	X	
o-23		X		X			X			
o-2			X							
o-8			X							X
o-10			X	X	X	X	X	X		
o-15			X							X
o-17			X							
o-22			X							
o-28			X			X				X
o-29				X					X	
o-16					X		X			
o-25					X					
o-30					X	X				
o-12						X				X
o-21						X			X	
o-3							X		X	X
o-14							X	X		
o-20							X			
o-7								X		
o-13									X	

Figure 1. The formal context $\mathbb{C}$, in LABFS-specific tabular form at right.

At right, in the tabular representation corresponding to the LABFS decomposition we can read that there are levels L_0, L_2 and L_4 consisting of attributes and seen grouped in order at the column headings, and levels L_1, L_3 and L_5 consisting of objects, identified by the groupings at the row headings. For example, L_1 is the attribute set {a-7,a-4,a-10,a-8}. Further, the subrelations E_i and S_i are easy to read off from the diagram: the subrelations S_i appear as consecutive rectangles, placed successively from the top left of the diagram down toward the bottom right, each with a line splitting its interior, indicating that the local object set (alternatively, local attribute set, depending on parity) is a union of two levels, namely L_i and L_{i+2} for the appropriate value of i.

In the following Lemma, we summarize the partitions we have established for a given graph-connected formal context.

Lemma 2. *Let $\mathbb{K} = (O, A, E)$ be an r-rooted formal context. If, for $i \in \mathbb{N} \cup \{0\}$, we define L_i and S_i as above, then*

$$A = \bigsqcup_{i \in \mathbb{N} \cup \{0\}} L_{2i}, \quad O = \bigsqcup_{i \in \mathbb{N} \cup \{0\}} L_{2i+1}, \quad E = \bigsqcup_{i \in \mathbb{N} \cup \{0\}} E_i, \quad \textit{and} \quad E = \bigcup_{i \in \mathbb{N} \cup \{0\}} S_i.$$

3.3 Induced Decomposition of the Concept Lattice

Figure 1 indicates (by example) that we can find concepts of the full context by looking within three consecutive levels. This section makes this observation rigorous and provides conditions that are sufficient to compute locally by identifying concepts of the full context from the list of concepts of a subrelation. In summary, we establish a connection between the concept lattices of subrelations determined by the LABFS decomposition and the full concept lattice of the original relation.

Definition: Let $\mathbb{K} = (O, A, E)$ be an r-rooted formal context. We say concept $(P, B) \in \mathfrak{B}(\mathbb{K})$ is in *r-concept level i* provided

$$\forall j < i, \quad (P \cup B) \cap L_j = \emptyset \quad \textit{and} \quad (P \cup B) \cap L_i \neq \emptyset\,.$$

Thus (P, B) is in concept level i provided that L_i is the first partition block, from the r-rooted LABFS decomposition of $\mathbb{K} = (O, A, E)$, that $P \cup B$ meets. To match traditional FCA notation, we let $\mathfrak{B}_i(\mathbb{K})$ denote the set of all concepts in concept level i.

The next sequence of statements establishes that the entire extent and intent of every concept of $\mathbb{K} = (O, A, E)$ must be included in three consecutive levels of the partition $\{L_0, L_1, ..., L_n\}$ of $O \cup A$.

Lemma 3. *Let $\mathbb{K} = (O, A, E)$ be an r-rooted formal context and let X be a nonempty subset of $O \cup A$. For $X \subseteq L_i$ and $i \geq 0$, it follows $X' \subseteq L_{i-1} \cup L_{i+1}$.*

Proof: Let $o \in X'$. First we suppose $o \in O$. Since X is nonempty, there exists some $x \in X$, and so $o \in X'$ implies $(o, x) \in E$. By Lemma 2, $E = \bigcup_{i \in \mathbb{N} \cup \{0\}} S_i$, so $x \in X \subseteq L_i$ implies that $(o, x) \in S_{i-1} \cup S_i \cup S_{i+1}$. In any case, $o \in L_{i-1} \cup L_{i+1}$. The same argument applies for $o \in A$.

Proposition 4. *Given an r-rooted formal context $\mathbb{K} = (O, A, E)$,*

$$\mathfrak{B}(\mathbb{K}) = \bigsqcup_{i \in \mathbb{N} \cup \{0\}} \mathfrak{B}_i(\mathbb{K}).$$

Also, if $(P, B) \in \mathfrak{B}_i(\mathbb{K})$, then

$$P \subseteq L_{i+1}, \quad B \subseteq L_i \cup L_{i+2} \quad if \quad i \quad is \quad even,$$

$$B \subseteq L_{i+1}, \quad P \subseteq L_i \cup L_{i+2} \quad if \quad i \quad is \quad odd.$$

Proof: First note that every concept of a graph-connected formal context must have a nonempty extent or a nonempty intent, since we assume that every formal context has nonempty object and attribute sets. Thus every concept (P, B) must satisfy $(P \cup B) \cap L_i \neq \emptyset$ for some i, and there must be a least such value i for which this is true, by the well-ordering of $\mathbb{N} \cup \{0\}$. This value i determines the r-concept level that (P, B) lies in.

For the second statement we argue the even case, leaving the similar argument in the odd case to the reader. Suppose $(P, B) \in \mathfrak{B}_{2s}$. Then $B \cap L_{2s} \neq \emptyset$, while $P \cap L_j = B \cap L_j = \emptyset$ for all $j < 2s$. We claim that $B' \subseteq L_{2s+1}$. By Lemma 3, $B' \subseteq L_{2s-1} \cup L_{2s+1}$. But $B' = P$, and if $P \cap L_{2s-1} \neq \emptyset$, then this contradicts $(P, B) \in \mathfrak{B}_{2s}(\mathbb{K})$, so we conclude that $B' \subseteq L_{2s+1}$. Again, by Lemma 3 $B = B''$ must be a subset of $L_{2s} \cup L_{2s+1}$.

The following statement is an immediate consequence of the definitions, and is recorded for later reference:

Lemma 5. *Let $\mathbb{K} = (O, A, E)$ be a formal context and suppose $H \subseteq E$. If $B \subseteq A$, then $B^H \subseteq B^E$. If $P \subseteq O$, then $P^H \subseteq P^E$.*

Proposition 6. *Let $\mathbb{K} = (O, A, E)$ be a graph-connected formal context, and let a root $r \in A$ be given. Then $\mathfrak{B}_i(\mathbb{K}) \subseteq \mathfrak{B}(S_i)$.*

Proof: Suppose $(P, B) \in \mathfrak{B}_i$. We consider the case where i is even, and leave the similar odd case to the reader. By Proposition 4, $P \subseteq L_{i+1}$ and $B \subseteq L_i \cup L_{i+2}$. Since $(P, B) \in \mathfrak{B}_i \subseteq \mathfrak{B}(\mathbb{K})$, it follows that $P^E = B$ and $B^E = P$. Thus, by Lemma 5,

$$P^{S_i} \subseteq P^E = B \quad and \quad B^{S_i} \subseteq B^E = P.$$

To show the inclusion $B \subseteq P^{S_i}$, we suppose $b \in B = P^E$. Then for every $p \in P$, pEb. Since $P \subseteq L_{i+1}$ and $B \subseteq L_i \cup L_{i+2}$, we conclude that for every $p \in P$,

$$pEb \;\; and \;\; (p, b) \in L_{i+1} \times (L_i \cup L_{i+2}),$$

so for every $p \in P$, pS_ib, that is, $b \in P^{S_i}$. Thus $B \subseteq P^{S_i}$.

Similarly, $B^E \subseteq B^{S_i}$, and since $B = P^{S_i}$ and $P = B^{S_i}$, we conclude that $(P, B) \in \mathfrak{B}(S_i)$.

Thus, the i^{th} r-concept level of an r-rooted formal context is included in the set of formal concepts of S_i, its i^{th} LABFS subrelation. Unfortunately, the reverse inclusion does not hold, but the following theorem shows that the set of non-trivial concepts in $\mathfrak{B}(S_i)$ that intersect L_i consists of formal concepts of $\mathbb{K}$ that are in the r-concept level i along with those ordered pairs (P, B) that generate (in $\mathbb{K}$) formal concepts in the r-concept level $i-1$.

Theorem 7. *Let $\mathbb{K} = (O, A, E)$ be an r-rooted formal context. If $(P, B) \in \mathfrak{B}(S_i)$, with $P \neq \emptyset$, $B \neq \emptyset$ and $L_i \cap (P \cup B) \neq \emptyset$, then*

1. *If $(P \cup B) \cap L_{i+2} \neq \emptyset$, then $(P, B) \in \mathfrak{B}_i(\mathbb{K})$.*
2. *If $(P \cup B) \cap L_{i+2} = \emptyset$, then*
   There exists $(Q, C) \in \mathfrak{B}_{i-1}(\mathbb{K})$ such that $P \subseteq Q$ and $B \subseteq C$, or
   $(P, B) \in \mathfrak{B}_i(\mathbb{K})$.

Proof: Suppose $(P, B) \in \mathfrak{B}(S_i)$, so that $P^{S_i} = B$ and $B^{S_i} = P$.

For 1., suppose $(P \cup B) \cap L_{i+2} \neq \emptyset$. We will assume that i is even, that is, that $B \cap L_{i+2} \neq \emptyset$, and leave the odd i case to the reader. By Lemma 5,

$$B = P^{S_i} \subseteq P^E \quad and \quad P = B^{S_i} \subseteq B^E$$

so we need only examine the reverse inclusions.

Let $b \in P^E$. Since $P \neq \emptyset$, there exists some $q \in P$ such that qEb. Also, $P \subseteq L_{i+1}$ (by construction of S_i and the assumption that i is even), so it follows from Lemma 2 that $b \in L_i \cup L_{i+2}$. Now we have that for all $p \in P$, pEb and $p \in L_{i+1}$ and $b \in L_i \cup L_{i+2}$. This is equivalent to saying that for all $p \in P$, pEb and $(p, b) \in L_{i+1} \times (L_i \cup L_{i+2})$, that is, for all $p \in P$, pS_ib. Thus $b \in P^{S_i} = B$, so we conclude that $P^E \subseteq B$, and we have shown that $B = P^E$.

Let $p \in B^E$. Since $B \cap L_i \neq \emptyset$, it follows that there is some $b_1 \in B \cap L_i$ such that pEb_1, and since $B \cap L_{i+2} \neq \emptyset$, it follows that there is some $b_2 \in B \cap L_{i+2}$ such that pEb_2. Now pEb_1 implies that $p \in L_{i-1} \cup L_{i+1}$ and pEb_2 implies that $p \in L_{i+1} \cup L_{i+3}$, so p is trapped:

$$p \in (L_{i-1} \sqcup L_{i+1}) \cap (L_{i+1} \sqcup L_{i+3}) = L_{i+1}.$$

Thus $B^E \subseteq L_{i+1}$. Now $p \in B^E$ and $p \in L_{i+1}$ imply that, for all $b \in B$, pEb and $p \in L_{i+1}$, which is equivalent to saying that, for all $b \in B$, pEb and $b \in L_i \cup L_{i+2}$ and $p \in L_{i+1}$. The latter expression is equivalent to $(p, b) \in L_{i+1} \times (L_i \cup L_{i+2})$, so, for all $b \in B$, pS_ib. Thus $p \in B^{S_i} = P$, and we conclude that $P = B^E$. This completes the proof of the first case.

To prove the second case, suppose $(P \cup B) \cap L_{i+2} = \emptyset$. We assume i is even and argue this case; thus we assume that $B \cap L_{i+2} = \emptyset$. The proof for odd i is similar, and is omitted.

First we show $B = P^E$. Since $(P, B) \in \mathfrak{B}(S_i)$, it follows that $P^{S_i} = B$ and $B^{S_i} = P$, so $B = P^{S_i} \subseteq P^E$.

We now prove $P^E \subseteq B$. Let $x \in P^E$. As $P \subseteq L_{i+1}$ and $P \neq \emptyset$, there exists some $k \in P$ such that kEx, and thus $x \in L_i \cup L_{i+2}$. But if $x \in L_{i+2}$, then $P \subseteq L_{i+1}$ implies $x \in P^{S_i} \cap L_{i+2}$, that is, $x \in B \cap L_{i+2}$, which contradicts the assumption that $B \cap L_{i+2} = \emptyset$. Thus $x \in L_i$, so $x \in P^{S_i} = B$. This shows that $B = P^E$.

Either $B^E = P^{EE} = P$ or $B^E = P^{EE} \supsetneq P$. In the former case, $(P, B) \in \mathfrak{B}_i$. In the latter case, we will prove $(P^{EE}, P^E) \in \mathfrak{B}_{i-1}$.

Suppose $P \subsetneq P^{EE}$. We claim $B^E = P^{EE} \subseteq L_{i-1} \cup L_{i+1}$ and $P^{EE} \cap L_{i-1} \neq \emptyset$. Since $B \subseteq L_i$, Lemma 3 implies that $B^E \subseteq L_{i-1} \cup L_{i+1}$. If $B^E \subseteq L_{i+1}$, then $B \subseteq L_i$ implies $B^{S_i} = B^E$, that is, $P = B^{S_i} = B^E = P^{EE}$, a contradiction. Thus $B^E \cap L_{i-1} \neq \emptyset$.

We conclude that $(P^{EE}, P^E) \in \mathfrak{B}_{i-1}$, so we set $Q = P^{EE}$ and $C = P^E$ to complete the proof.

Recall the standard attribute set embedding $\mu : A \to \mathfrak{B}(\mathbb{K})$ given by

$$\mu(a) = (\{a\}', \{a\}'') ,$$

and the corresponding object embedding γ. To ease the description of the concept lattice of the subrelation generated from the root r and the elements in all the levels up to some level L_i, we extend the usual notation, by defining $\mu[B] = (B', B'')$, for $B \subseteq A$, and similarly, $\gamma[P] = (P'', P')$, for $P \subseteq O$.

Definition: Fix $\mathbb{K} = (O, A, E)$. For any subset $B \subseteq A$, define

$$\downarrow_{\mathfrak{B}(\mathbb{K})} \mu[B] = \bigcup_{b \in B} \{C \in \mathfrak{B}(\mathbb{K}) \mid b \in intent(C)\}$$

and for $P \subseteq O$

$$\uparrow_{\mathfrak{B}(\mathbb{K})} \gamma[P] = \bigcup_{p \in P} \{C \in \mathfrak{B}(\mathbb{K}) \mid p \in extent(C)\}.$$

In plain language, these subsets of $\mathfrak{B}(\mathbb{K})$ are the order ideal of $\mathfrak{B}(\mathbb{K})$ determined by $\mu[B]$ and the order filter of $\mathfrak{B}(\mathbb{K})$ determined by $\gamma[P]$, where γ and μ are the object- and attribute-embedding maps of [Ganter and Wille (99)]. Our purpose in introducing these subsets of $\mathfrak{B}(\mathbb{K})$ is to describe the connection between the unions

$$\bigcup_{j=0,1,\ldots,i} \mathfrak{B}_j(\mathbb{K}) ,$$

which are clearly nested as i ranges from 0 to its maximum value, and the concepts of unions of the subrelations S_j. Note that if L_0, L_1, ..., L_{k-1}, L_k is the full list of levels of an r-rooted formal context $\mathbb{K}$, then

$$\mathfrak{B}(\mathbb{K}) = \bigcup_{j=1,\ldots,k-1} \mathfrak{B}_j(\mathbb{K}) \ .$$

Corollary 8. *Let $\mathbb{K} = (O, A, E)$ be an r-rooted formal context, such that the intent of $1_{\mathfrak{B}(\mathbb{K})}$ is empty and the extent of $0_{\mathfrak{B}(\mathbb{K})}$ is empty. Let $i \in \mathbb{Z}$ satisfy $i \geq 2$. Define the formal context $\mathbb{K}_i = (O_i, A_i, U_i)$ by setting*

$$U_i = \bigcup_{j=0,1,\ldots,i-2} S_j,$$

and letting O_i and A_i be the domain and codomain of U_i, respectively. Let $\mathcal{C}_i$ be the set of concepts (P, B) of $\mathbb{K}_i$ such that $P \neq \emptyset$ and $B \neq \emptyset$ and $P \cup B \nsubseteq L_{i-1} \cup L_i$. Then $\mathcal{C}_i$ is equal to

$$[(\bigcup_{j<i-2,\ j\ odd} \uparrow_{\mathfrak{B}(\mathbb{K})} \gamma[L_j]) \setminus 1_{\mathfrak{B}(\mathbb{K})}] \cup [\bigcup_{j\leq i-2,\ j\ even} \downarrow_{\mathfrak{B}(\mathbb{K})} \mu[L_j]) \setminus 0_{\mathfrak{B}(\mathbb{K})}] \ .$$

The proof of the Corollary is left to the reader. The most important note is that the concepts in $\mathcal{C}_i$ are all in $\mathfrak{B}(\mathbb{K})$ by Theorem 7, because of the condition $P \cup B \nsubseteq L_{i-1} \cup L_i$ and the definition of U_i.

Of course, a corresponding statement can be made for odd values of i. The key point to take from Corollary 8 is that for increasing values of i, the unions that express the concepts in $\mathcal{C}_i$ are nested, so that, as i grows – that is, as the LABFS decomposition of $\mathbb{K}$ unfolds the full relation – the union expresses which are the corresponding concepts in the full concept lattice. The appearance of the levels L_j in the unions expressing $\mathcal{C}_i$ indicates the tight connection between LABFS levels and the concepts they generate.

4 Unfolding a Concept Lattice Using LABFS

In practice, the effect of Corollary 8 is that we can choose a root, say, some attribute $r \in A$, and unfold the lattice diagram from the corresponding attribute concept $\mu(r)$. Suppose the rooted LABFS decomposition has been determined, and we have limited our attention to the subcontext generated by the union of the subrelations S_j, as appears in Corollary 8. The concepts of $\mathfrak{B}(\mathbb{K})$ that can be computed from this subrelation are, first, those non-trivial concepts below $\mu(r)$, the concept determined by the root r. Next we compute those concepts which include objects in L_1 in their extent but exclude r in their intent; this includes the attribute concepts $(\{\alpha\}', \{\alpha\}'')$ with $\alpha \in L_2$, except for those such that $\alpha \rightarrow r$. Next come the concepts which include attributes in L_2 (except for the attribute concepts which have already appeared), and so on, up to L_i, save for those two-part concepts (described in Theorem 7) including only elements from L_{i-1} and L_i which must be verified in the full context $\mathbb{K} = (O, A, E)$. In this section, we first present an example illustrating the unfolding of the concept lattice from a root concept, and follow that with a discussion of the utility of unfolding in the context of viewing association rules via a concept lattice.

4.1 An Example of Unfolding

Now consider the formal context $\mathbb{F}$ in Figure 2 below, with 14 objects and 13 attributes, already arranged to display the attr-1−rooted LABFS decomposition. This synthetic data is both sparse and connected. The context is followed by the lattice $\mathfrak{B}(\mathbb{F})$, with 34 concepts and 69 edges.

14	13												
	attr-1	attr-2	attr-3	attr-4	attr-5	attr-6	attr-7	attr-8	attr-9	attr-10	attr-11	attr-12	attr-13
obj-1	X	X											
obj-2		X	X										
obj-3			X	X									
obj-4				X	X								
obj-5					X	X							
obj-6					X		X						
obj-7						X	X						
obj-8							X	X					
obj-9								X	X				
obj-10									X	X			
obj-11										X	X	X	
obj-12										X	X		X
obj-13										X		X	X
obj-14											X	X	X

Figure 2. A graph-connected formal context $\mathbb{F}$, in LABFS-specific tabular form.

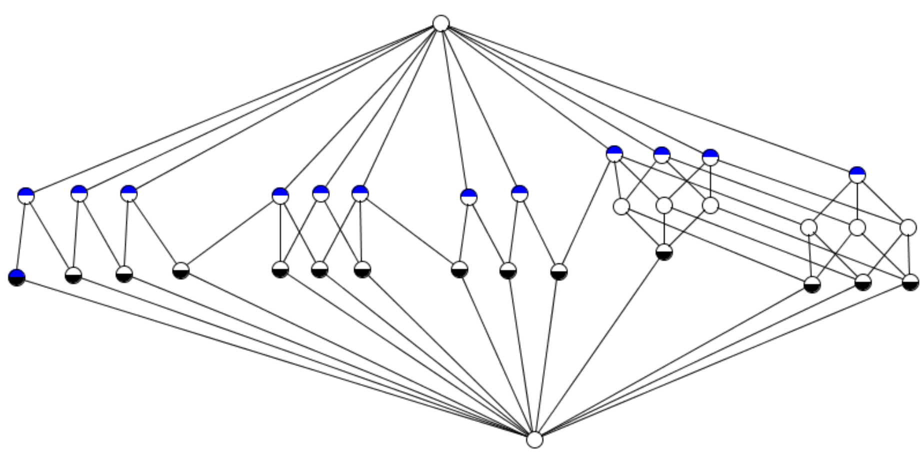

Figure 3. The lattice $\mathfrak{B}(\mathbb{F})$, drawn by Concept Explorer.

The concept lattice in Figure 3 is drawn using Concept Explorer [ConExp]. After minimal manipulation of the diagram, we see that its structure is that of a 4-fence lattice, glued at a coatom to a Boolean algebra with 3 atoms, glued at a coatom to a 3-fence lattice with one additional leg, glued at a coatom to a Boolean algebra with 4 atoms (a j-fence lattice is a j-fence with a top and bottom added).

We have implemented a program, Decompose, which converts an input of a formal context and a root to an output of the LABFS decomposition of the relation, also allowing the user to output the subcontext generated by levels L_i through L_j, for $i < j$. The concept lattice of the output subcontext can then be viewed in Concept Explorer [ConExp]. The next figure shows the output of Decompose on the input of the context in Figure 2, along with its determination of the context generated by subrelation $S_0 \cup S_1 \cup . . . \cup S_9$, and finally the conversion of that subcontext to the input format of Concept Explorer.

```
FenceBA-3FenceBA-4.txt
| 27 | 13 | 14 | 32 | 0.175824
Root | # of Vert | # of Edges | # of Levels
attr-1 27 32 20
L0 1 1 1.0 (0,0)       L10 2 3 0.75 (11,12)
L1 1 1 1.0 (1,1)       L11 2 1 0.5 (13,14)
L2 1 1 1.0 (2,2)       L12 1 1 1.0 (15,15)
L3 1 1 1.0 (3,3)       L13 1 1 1.0 (16,16)
L4 1 1 1.0 (4,4)       L14 1 1 1.0 (17,17)
L5 1 1 1.0 (5,5)       L15 1 1 1.0 (18,18)
L6 1 1 1.0 (6,6)       L16 1 3 1.0 (19,19)
L7 1 1 1.0 (7,7)       L17 3 6 0.66 (20,22)
L8 1 2 1.0 (8,8)       L18 3 3 1.0 (23,25)
L9 2 2 0.5 (9,10)      L19 1 0 0 (26 26)
```

Figure 4. The LABFS overview of $\mathbb{F}$*, from Decompose.*

Decompose displays the statistics associated with each new level, and between new levels and the previous level. The display shows the number of new vertices in level L_i, the number of edges from level L_i to level L_{i+1}, and the density of edges relative to the maximum possible $|L_i \times L_{i+1}|$. This feature provides an overview of the LABFS results, relative to root r (in the example shown, r =attr-1), so that sparser sections of the data can be distinguished from less sparse sections. In Figure 4, Decompose indicates the portions of the binary relation that form a fence: the consecutive levels reading "Li 1 1 1.0 (j, j)" indicate levels of the LABFS that have one new vertex, one edge to level L_{i+1}, density 1.0, and where the new vertices range from j to j in the index list.

After the LABFS overview, Decompose allows the user to output the subcontext generated by levels L_i to L_j. For example below, by requesting levels L_0 through L_{11}, the subrelation $S_0 \cup S_1 \cup . . . \cup S_9$ is saved to a file, and similarly by requesting levels L_{11} through L_{19}, the subrelation $S_{11} \cup S_{12} \cup \ldots \cup S_{19}$ is saved to a file. If requested, Decompose also converts a given context to the input format of Concept Explorer (ConExp), so its lattice can be viewed.

A	B	C	D	E	F	G	H
	attr-1	attr-2	attr-3	attr-4	attr-5	attr-6	attr-7
obj-1	X	X					
obj-2		X	X				
obj-3			X	X			
obj-4				X	X		
obj-5					X	X	
obj-6					X		X
obj-7						X	X
obj-8							X

Figure 5a. The subcontext $\mathbb{F}_{0-11}$ of $\mathbb{F}$, viewed in ConExp.

A	B	C	D	E	F	G
	attr-8	attr-9	attr-10	attr-11	attr-12	attr-13
obj-8	X					
obj-9	X	X				
obj-10		X	X			
obj-11			X	X	X	
obj-12			X	X		X
obj-13			X		X	X
obj-14				X	X	X

Figure 5b. The subcontext $\mathbb{F}_{11-19}$ of $\mathbb{F}$, viewed in ConExp.

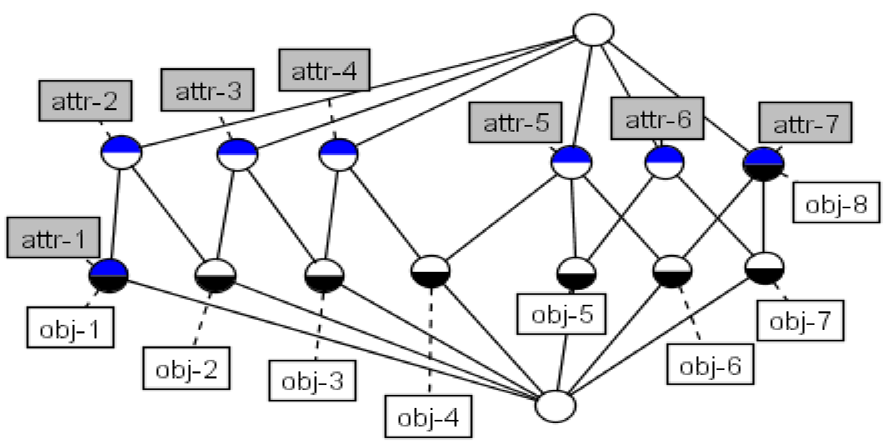

Figure 6a. The lattice of the subcontext $\mathbb{F}_{0-11}$ in Figure 4a, viewed in ConExp.

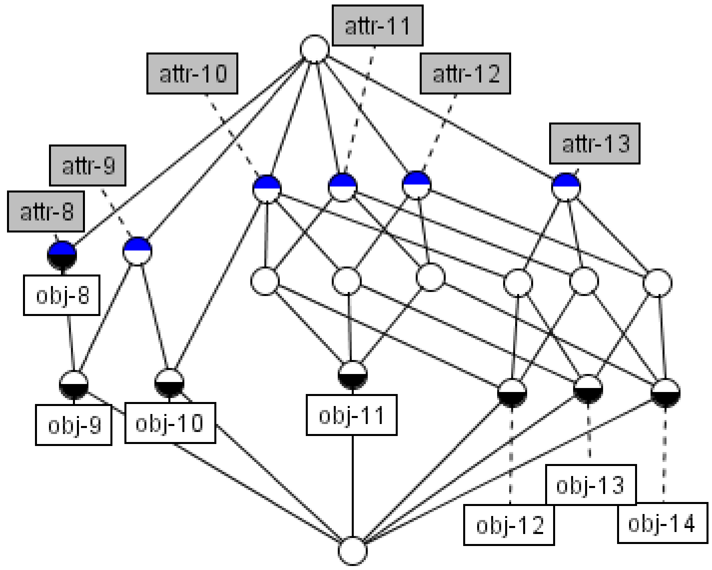

Figure 6b. The lattice of the subcontext $\mathbb{F}_{11-19}$ *in Figure 4b, viewed in ConExp.*

In Figure 6a, note that the concept generated by obj-8 in the subcontext is not the same concept generated in the full context of Figure 2 (it does not include attr-8 as it should), but all other concepts are concepts of the full context. As stated above, only concepts determined exclusively from elements of the last two levels need to be checked for their standing in the full context – all others are "real". A similar comment applies to Figure 6b, so that the only concept from the full concept lattice $\mathfrak{B}(\mathbb{F})$ that does not appear the lattice $\mathfrak{B}(\mathbb{F}_{0-11})$, nor in $\mathfrak{B}(\mathbb{F}_{11-19})$, is the object concept $(\{obj-8\}'', \{obj-8\}')$.

We have not yet implemented a lattice viewer that visually contrasts actual concepts (of the full context) with those that may only be two-part concepts (again, as in Theorem 7). Thus in viewing the lattices of subcontexts we are left to determine which concepts are concepts of the full subrelation. However, flagging the actual concepts should not pose any great challenge, since the results we have presented provide sufficient information to compute only those concepts of induced subcontexts that are in the full concept lattice.

4.2 Rooted LABFS and Association Rules

From the artificial example just considered and Theorem 7, it is apparent that these results regarding the LABFS decomposition may allow (depending on the depth of the relation) an efficient localized computation of the concept lattice. A judicious choice of a root will impact the level of lattice unfolding, and since the complexity of LABFS is linear on the number of edges of the corresponding graph, we foresee that inexpensive searches will help make this choice. The sparser the data, the more useful we expect the LABFS decomposition to be in providing such localization. If in addition the diameter of the associated graph is large then by choosing antipodal roots one expects more lattice unfolding.

On the other hand, the worst-case context $(\{1, 2, 3, ..., n\}, \{1, 2, 3, ..., n\}, \neq)$ is not decomposed at all by LABFS, as any element chosen as root will yield only levels L_0, L_1 and L_2, so the only subrelation is S_0, and it is not a proper subrelation.

The translation of LABFS into the concept lattice allows the lattice to be unfolded from a root attribute. If we determine a subrelation from levels L_0, L_1, L_2, and L_3, then all concepts with the root in their intent will be present, and all these concepts are concepts of the full lattice. Further, any concepts with attributes from L_2 in their intent will have their full extents represented, since we extended as far as L_3, though they may be missing attributes in their intent. It is well-known that any association rule, say $\{\alpha, \beta, \delta\} \to \rho$, has its confidence value $conf_{\mathbb{K}}(\ \{\alpha, \beta, \delta\} \to \rho\ )$ determined by dividing the cardinality of the concept extent $\{\alpha, \beta, \delta\}'$ by the cardinality of the concept extent $\{\alpha, \beta, \delta, \rho\}'$. Suppose we compute $\mathbb{K}_3(\rho)$ from a given context $\mathbb{K}$. If we are focused on what (nearly) implies ρ, there are two advantages offered by working in $\mathbb{K}_3(\rho)$ instead of $\mathbb{K}$: first, only attributes related to ρ through some objects will appear in $\mathfrak{B}(\mathbb{K}_3(\rho))$, e.g. $\alpha, \beta, \delta \in A$ with $\{\alpha, \beta, \delta, \rho\}' \neq \emptyset$; second, the presence of the full ($\mathbb{K}$-)extents for the related concepts $\mu[\{\alpha, \beta, \delta, \rho\}]$ and $\mu[\{\alpha, \beta, \delta\}]$ in $\mathbb{K}_3(\rho)$ ensures that

$$conf_{\mathbb{K}}(\ \{\alpha, \beta, \delta\} \to \rho\ ) = conf_{\mathbb{K}_3(\rho)}(\ \{\alpha, \beta, \delta\} \to \rho\ )\ .$$

A similar observation regarding support can be made as long as we divide by the cardinality $|O|$ of the full object set instead of the cardinality of $O_j(\rho)$, the object set of $\mathbb{K}_j(\rho)$.

Thus, limiting our attention to only levels L_0, L_1, L_2 and L_3 corresponds to a query mode for considering association rules that have the root in their conclusion (note that this is a related, but more theoretically grounded, version of the query process discussed in [Brooke]). Beyond these first three levels, Decompose can indicate how many objects and attributes will be picked up, and the user can decide how deeply to go before truncating to a subcontext. If this is not possible, another root can be chosen for separate examination.

We make one final note regarding the dependence of the depth of the LABFS output on the nature of the input. Even in the presence of very sparse connected data, a complementary attribute will disallow a LABFS decomposition from having any significant depth. Specifically, if a formal context includes complementary attributes α and ζ, and a root $\beta \in A$ is chosen which does not imply α

and does not imply ζ (this will be the case for most epidemiological data, since, e.g., it is often recorded whether a member of a population is Male or Female), then the LABFS decomposition will present at most levels L_0, L_1, ..., L_4. This is because $L_1 = \{\beta\}'$ will include at least one object that satisfies α and at least one object that satisfies ζ, so L_2 must include the attributes α and ζ, which in turn forces L_3 to include all the remaining objects, and then L_4 must include any remaining attributes that did not appear in L_0 or L_2.

5 Conclusion

The determination of the LABFS decomposition(s) associated with a formal context provides an overview of the data – crucial when the data is large, since the lattice will be too large to view (or even possibly to store) – and the results we have presented provide sufficient information to compute only those concepts of LABFS-induced subcontexts that are in the full concept lattice.

Generalizing our comments about the effect of a complementary pair of attributes on the maximum depth of a LABFS decomposition, note that if there are k attributes such that their neighborhoods correspond to a partition of the set of objects this means that they are mutually independent in the sense that there are no implications among them, and we could present a similar characterization of the depth of LABFS under assumptions that mirror those mentioned for a complementary pair. This suggests their identification and removal, as a preprocessing step before applying LABFS. In any event, LABFS decomposition will be more effective if the root is chosen to maximize the number of levels of decomposition. In graph theoretical terms this corresponds to choosing the root so that the number of levels as close as possible to the diameter of the data bigraph. Related methods that exploit other graph parameters use the notions of cores and cuts and we are currently investigating these approaches. The overall objective is to decompose the data bigraph in an efficient manner that translates into decompositions of the corresponding concept lattice, and the understanding established in this paper is a small step in this larger program. These techniques become imperative when the concept lattice becomes so large that it does not fit in random access memory, so that external memory algorithms are required [Abello and Vitter eds (99)]. Moreover, even if it fits in random access memory a fine graded decomposition allows its visual exploration on a monitor screen, the screen certainly being smaller than memory by several orders of magnitude [Abello and Korn 02].

Acknowledgements

The authors thank the referees for a number of helpful style suggestions and for correcting some errors in the original manuscript. We also acknowledge financial assistance from the National Science Foundation, and in particular Fred Roberts and Mel Janowitz for their support of this research through the DIMACS Computational Epidemiology project.

References

[Agrawal et al. 93] Agrawal, R., Imielinski, T. and A. Swami, A.: "Mining association rules between sets of items in large databases", in ACM SIGMOD Intl. Conf. Management of Data, May 1993.

[Abello and Korn 02] J. Abello, J. Korn: "MGV: A System for Visualizing Massive Multidigraphs", IEEE Transactions on Visualization and Computer Graphics **8** No. 1, January-March 2002.

[Abello and Vitter eds (99)] J. Abello, J. Vitter (eds): "External Memory Algorithms", Vol 50 of the AMS-DIMACS Series in Discrete Mathematics and Theorethical Computer Science, 1999.

[Abello et al. (02)] J. Abello, M. Resende, and S. Sudarsky: "Massive Quasi-Clique Detection", In Proceedings of Latinoamerican Informatics, Springer Verlag LNCS, May 2002.

[Berry and Sigayret 1-02] A. Berry and A. Sigayret: "Representing a concept lattice by a graph", Proceedings of Discrete Maths and Data Mining Workshop, 2nd SIAM Conference on Data Mining (SDM'02), Arlington (VA), April 2002.

[Berry and Sigayret 2-02] A. Berry and A. Sigayret: "Obtaining and maintaining polynomial-sized concept lattices", Proceedings of Workshop FCAKDD (Formal Concept Analysis for Knowledge Discovery in Data bases), ECCAI 02.

[Brooke] S. Brooke, "Multi-Dimensional Location System", `http://www.jasmine.org.uk/~simon/bookshelf/papers/MDLS_RFC.html`

[FCA URL] For a reference list of FCA application papers see: `http://www.mathematik.tu-darmstadt.de/ags/ag1/Literatur/literatur\_en.html`

[Birkhoff (40)] Birkhoff, G.: "Lattice Theory", American Mathematical Society, Providence, R.I., 1st edition, 1940.

[Ganter and Wille (99)] Ganter, B. and Wille, R.: "Formal Concept Analysis: Mathematical Foundations", Springer, NY, 1999. ISBN 3-540-62771-5

[Freese] Freese, R.: "LatDrawWin, a lattice drawing applet", www.math.hawaii.edu/ ˜ ralph/LatDraw.

[Stumme et al. 02] Stumme, G., Taouil, R., Bastide, Y., Pasquier, N. and Lakhal, L.: "Computing iceberg concept lattices with Titanic", Data and Knowledge Engineering (Elsevier), 42 (2002), 189-222.

[Wille 01] Wille, R.: "Why Can Concept Lattices Support Knowledge Discovery in Databases?", Technische Universitat Darmstadt, Fachbereich Mathematik, Preprint Nr. 2158, June 2001. Available at `http://wwwbib.mathematik.tu-darmstadt.de/Math-Net/Preprints/Listen/pp01.html`

[ConExp] Yevtushenko, S., et al: "Concept Explorer", Open source java software available at `http://sourceforge.net/projects/conexp`, Release 1.2, 2003.

[Zaki et al. 98] Zaki, M., and M. Ogihara, M.: "Theoretical Foundations of Associations Rules", in Proceedings of 3rd SIGMOD'98 Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'98), Seattle, Washington, USA, June 1998.