Obtaining and maintaining polynomial-sized concept lattices

Anne Berry[1] and **Alain Sigayret**[1]

Abstract. We use our definition of an underlying co-bipartite graph which encodes a given binary relation to propose a new approach to defining a sub-relation or incrementally maintaining a relation which will define only a polynomial number of concepts.

1 Introduction

Concept lattices, also known as Galois lattices, have been studied for a long time, for example in the context of Social Sciences (see [1]), but recently, Wille and Ganter's work have introduced new perspectives and applications, and the use of concept lattices is rapidly emerging in many areas related to Artificial Intelligence and Data Mining, such as Data Base Management, Machine Learning, and Frequent Set Generation (see e.g. [24], [23], [22], [18], [14]).

The main drawback of concept lattices is that they may be of exponential size. This makes it impossible, in practise, to compute and span the entire structure they describe. It is thus of primeval importance to be able to navigate the lattice efficiently, or to be able to define a polynomial sized sub-lattice which contains the right information.

Though it is experimentally known that the greater the number of "ones" of the binary relation is, the more concepts the corresponding lattice will tend to have, there is no formal known characterization of the family of binary relations which will define a polynomial number of concepts. Thus in order to make a lattice smaller and more feasible, we can always remove "ones" from the relation, but this will not, in general, yield a sub-lattice, and there are even cases in which this may increase the number of concepts which the relation defines.

In this paper, we propose a formal approach which will help define binary relations with a polynomial number of concepts.

Our tool, surprisingly enough, is an emerging area of the theory of undirected graphs: the theory of minimal separation. A separator is a group of vertices, the removal of which disconnects the graph, just as the removal of an internal node in a tree will disconnect the tree into several sub-trees. Minimal separation, introduced by Dirac in 1961 to characterize chordal graphs, a family of graphs which is a direct generalization of trees, has recently given rise to a consistent body of research, with many new results for a variety of graph classes, and even in the general case for arbitrary graphs.

In a recent contribution in the Data Mining area (see [7]), we introduce a new encoding for a given binary relation, by using a graph constructed on the complement of the relation. We show that there is a one to one correspondence between the concepts defined by the relation and the minimal separators defined by this underlying graph.

[1] LIMOS UMR CNRS 6158, Ensemble Scientifique des Cézeaux, Université Blaise Pascal, BP 10125, 63 173 Aubière, cedex, France. E-mail: berry@isima.fr, sigayret@isima.fr

This fundamental remark enables us to apply the powerful recent tools defined on minimal separation to concept lattices: given any relation, we are able to compute a poly-sized sub-relation resulting in a sub-lattice of the original lattice. This is done by repeatedly choosing a minimal separator of the underlying graph and forcing it into a clique, a process which computes a partial embedding of the graph into a chordal graph. We are also able to decompose the lattice, and propose a new way of generating the cover of an element, as well as several algorithmic processes to generate the set of concepts, based on minimal separator enumeration algorithms.

In this paper, we focus on the possibility of reducing a lattice to a poly-sized sub-lattice or maintaining a relation which we are sure to define a poly-sized concept lattice, by using a different and larger class of graphs: the class of weakly chordal graphs, a large superclass of chordal graphs.

Our contribution is intended as a position paper, presenting a prospective research direction, interesting both for the field of Formal Concept Analysis, and for the field of Graph Theory; indeed, any theoretical or experimental breakthrough in one of these two areas would lead to results in the other.

2 Preliminaries

In order to make this paper self-contained, we will need to give some preliminaries, especially since we need graph theoretic concepts which are not generally used in KDD problematics.

2.1 Lattices

Given a finite set $\mathcal{P}$ of "properties" or "attributes" and a finite set $\mathcal{O}$ of "objects" or "tuples", we will consider a binary relation R as a proper subset of the Cartesian product $\mathcal{P} \times \mathcal{O}$; we will refer to the triple $(\mathcal{P}, \mathcal{O}, R)$ as a **context**, and denote $\mathcal{L}(R)$ the corresponding concept lattice.

Example 2.1 *Let* $(\mathcal{P}, \mathcal{O}, R)$ *be a context, with* $\mathcal{P}= \{a, b, c, d, e, f\}$, $\mathcal{O}= \{1, 2, 3, 4, 5, 6\}$, *and* R:

	a	*b*	*c*	*d*	*e*	*f*
1		x	x	x	x	
2	x	x	x			
3	x	x				x
4				x	x	x
5			x	x		
6	x					

The associated concept lattice $\mathcal{L}(R)$ *is shown in Figure 1.*

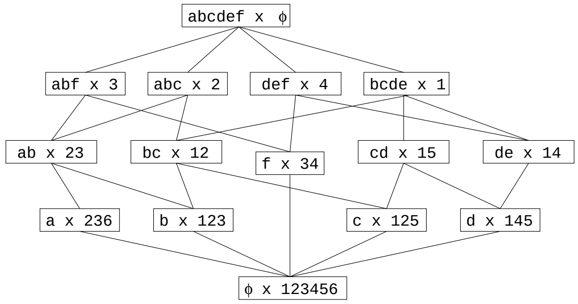

Figure 1. Concept lattice $\mathcal{L}(R)$ of relation R from Example 2.1.

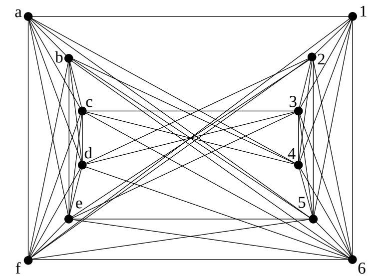

Figure 2. Underlying graph G_R of relation R from Example 2.1.

2.2 Graphs

The graphs used in this paper are finite and undirected. A graph is denoted $G = (V, E)$; V is the vertex set, $|V| = n$, and $E \subset (V \times V)$ is the edge set, $|E| = m$.

For $X \subset V$, $G(X)$ denotes the subgraph induced by X in G (only the edges which have both endpoints in X are retained).

The **neighborhood** of vertex x (the set of vertices y such that xy is an edge of E) is denoted by $N(x)$. If xy is an edge of E, we say that x and y **see** each other. For $C \subset V$, $N(C) = \cup_{x \in C}(N(x)) - C$. A vertex is said to be **universal** if it sees all the other vertices of the graph.

A **clique** is a set X of vertices such that $\forall x, y \in X, x \neq y, xy \in E$.

A graph is said to be **chordal** or **triangulated** if it contains no chordless induced cycle of length strictly greater than three. **Minimal triangulation**, also called minimal chordal completion, is the process of embedding a graph into a chordal graph by the addition of an inclusion-minimal set of edges.

The basic notion we use in this work is that of minimal separator.

A **separator** S of a connected graph G is a subset of vertices such that subgraph $G(V - S)$ is disconnected. S is called an **xy-separator** if x and y lie in different connected components of $G(V - S)$; S is called a **minimal xy-separator** if S is an xy-separator and no proper subset of S separates x from y. Finally, S is called a **minimal separator** if there is some pair $\{x, y\}$ of vertices such that S is a minimal xy-separator.

A minimal separator S of a graph G is characterized by the fact that there are at least two distinct connected components A and B of $G(V - S)$ such that $N(A) = N(B) = S$; A and B are called *full components*. S is then a minimal ab-separator for any pair $\{a, b\}$ of vertices where $a \in A$ and $b \in B$.

S is called a **clique separator** if it is a separator and a clique; we will say that we **saturate** a non-clique separator S if we add all missing edges necessary to make S into a clique.

With a context $(\mathcal{P}, \mathcal{O}, R)$, we associate an underlying co-bipartite graph $G_R = (V, E)$ with vertex set $V = \mathcal{P} \cup \mathcal{O}$, where $\mathcal{P}$ and $\mathcal{O}$ are cliques, and there is an xy edge in E, $x \in \mathcal{P}, y \in \mathcal{O}$ iff (x, y) is **not** in R.

We show that if S is a minimal separator of G_R, then $G(V - S)$ has exactly two connected components, A and B, with $A \times B$ a concept of $\mathcal{L}(R)$, and vice-versa.

Example 2.2 *The underlying graph G_R of relation R from Example 2.1 is shown in Figure 2. $S = \{a, d, e, f, 3, 4, 5, 6\}$ is a minimal separator of G_R, as illustrated by Figure 3. S separates $A = \{b, c\}$ from $B = \{1, 2\}$, and $bc \times 12$ is a concept of R and an element of $\mathcal{L}(R)$.*

Note that $N(b) = \{a, c, d, e, f, 4, 5, 6\}$, $N(c) = \{a, b, d, e, f, 3, 4, 6\}$, so $N(A) = (N(b) \cup N(c)) - A = \{a, d, e, f, 3, 4, 5, 6\} = S$, which shows that S is indeed a minimal separator of G_R.

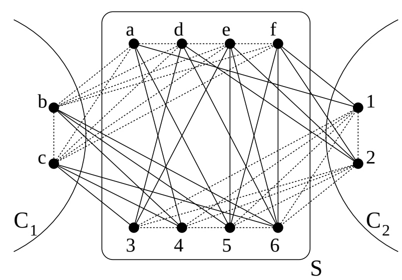

Figure 3. Separator $S = \{a, d, e, f, 3, 4, 5, 6\}$ of G_R.

This enables us to use existing algorithms for generating the minimal separators of a graph (see [21], [20],[15], [5]) to efficiently generate the concepts, matching the best complexities of [16] and [9].

Moreover, if we add in G_R the edges necessary to make S into a clique, defining a new relation R', which is obtained from R by deleting the corresponding crosses, then concept lattice $\mathcal{L}(R')$ is the sub-lattice obtained from $\mathcal{L}(R)$ by removing all the elements which are not comparable to $A \times B$.

We thus propose a generic process which will automatically produce a sub-lattice:

Given a binary relation R, compute the corresponding underlying graph G_R; find a minimal separator S of G_R; saturate S by adding any missing edges; remove from R the corresponding crosses.

The new lattice obtained is strictly smaller than the previous one; the process can be repeated, and it will end in a polynomial number of steps when a maximal chain of the lattice is obtained. The corresponding underlying graph will be a chordal graph, a class known to have very few minimal separators (less than the number of vertices).

Example 2.3 *Let us saturate separator S=$\{a, d, e, f, 3, 4, 5, 6\}$ of G_R from Example 2.2, representing concept $bc \times 12$. Edges $a3$, $a6$, $d4$, $d5$, $e4$, $f3$ and $f4$ will be added to G_R and the corresponding crosses removed from the relation, defining a new relation R':*

	a	*b*	*c*	*d*	*e*	*f*
1		*x*	*x*	*x*	*x*	
2	*x*	*x*	*x*			
3		*x*				
4						
5			*x*			
6						

Figure 4 gives the very restricted sublattice $\mathcal{L}(R')$ obtained. Saturating S has caused concepts $a \times 236$, $ab \times 23$, $abf \times 3$, $d \times 145$, $cd \times 15$, $f \times 34$, $def \times 4$ and $de \times 14$ to disappear from the lattice.

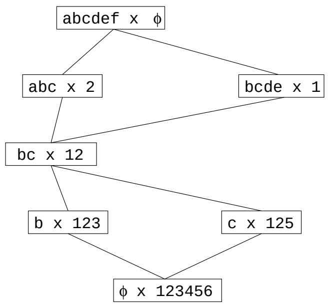

Figure 4. Concept lattice $\mathcal{L}(R')$.

There exist a variety of ways of finding a minimal separator in a graph in linear time; one can, for example, easily compute the minimal separators included in the neighborhood of a vertex x which is not universal, by computing the connected components of the graph $G(V - (N(x) \cup \{x\}))$; for each connected component C thus defined, $N(C)$ is a minimal separator of G.

Moreover, when the graph G is co-bipartite, a clique minimal separator can be used to decompose the graph into two proper subgraphs G_1 and G_2 such that any minimal separator of G is a minimal separator of either G_1 or G_2. Thus, after a first minimal separator is saturated, the next minimal separator can be searched for on a strictly smaller graph (with fewer vertices and fewer edges).

3 Computing a sub-relation by weakly chordal graph embedding

The process described in the previous section ensures that, given any binary relation, we can obtain a sub-relation defining a poly-sized sub-lattice, a theoretical breakthrough.

When applied experimentally, however, this process turns out to be too brutal, as every step will define a new articulation point in the lattice, and causes too many concepts to disappear. We are thus led to propose a more subtle way of modifying the lattice, which yields a poly-sized lattice which is larger and retains more information.

Repeatedly saturating a minimal separator in an arbitrary graph will yield a minimal triangulation of the input graph, which is an embedding into a chordal graph by addition of an inclusion-minimal set of edges (see [17], [2]).

One of the interesting properties of chordal graphs which we use is that the resulting graph has a polynomial number of minimal separators, in fact very few (less than the number of vertices).

It seems logical, in order to be less reductive, to try embedding the graph into another larger class of graphs. Recent research has shown that there are other graph classes with only a polynomial number of minimal separators; the most promising one is the class of weakly chordal graphs (also called weakly triangulated graphs), which was introduced by Hayward in 1985 (see [11]) as a generalization of chordal graphs, and which define a large super-class of chordal graphs.

Definition 3.1 *A graph G is said to be **weakly chordal** or **weakly triangulated** if neither G nor the complement of G contains an induced chordless cycle of length strictly greater than 4.*

This generalizes chordal graphs which are defined as graphs with no induced chordless cycle of length strictly greater than 3; the requirement added for the complement ensures that the class remains a perfect graph class, which is important because many hard problems become polynomial on such classes.

The class of weakly chordal graphs has been well-studied, and has given rise to many results, with an interesting incremental composition scheme (see [12]), several efficient recognition algorithms (see [13], [6]) and many properties which are similar to those of chordal graphs (see [6], [4])).

Moreover, graphs of this class are shown in [6] to have no more than $O(m)$ minimal separators, that is not more than it has edges.

It is thus interesting to experiment embedding our underlying graph into a weakly chordal graph. In order to do this, we use the following algorithm:

We first embed the graph into a chordal graph, then remove the added edges one by one until the graph becomes critical, by which we mean that any other edge removal will yield a graph which fails to be weakly triangulated.

This process has been shown in [19] to yield a minimal triangulation using chordal graphs instead of weakly chordal ones as critical graphs. The classes of chordal and weakly chordal graphs exhibit such stricking similarities that we conjecture that there is a strong theoretical correlation between the minimal separators of such a weak chordal completion and the minimal separators of the input graph, as is the case with classical chordal completion.

Consequently, we have tested on binary relations the embedding into a weakly chordal graph, and obtained encouraging results, as we not only obtain a polynomial lattice showing structural similarities with the lattice defined by the original relation, but we also often actually obtain a sub-lattice, which is without articulation point and thus retains more information than when saturating a minimal separator.

Example 3.2 *Initial relation R from Example 2.1 defines an underlying graph G_R which is not weakly chordal. Adding edge $f4$ will embed the graph into a weakly chordal graph. A new relation R'', without cross $f4$, is obtained:*

	a	*b*	*c*	*d*	*e*	*f*
1		*x*	*x*	*x*	*x*	
2	*x*	*x*	*x*			
3	*x*	*x*				*x*
4				*x*	*x*	
5			*x*	*x*		
6	*x*					

The corresponding lattice $\mathcal{L}(R'')$ is shown in Figure 5. Note how it is richer and more interesting than $\mathcal{L}(R')$ obtained by saturating a minimal separator and illustrated in Figure 4.

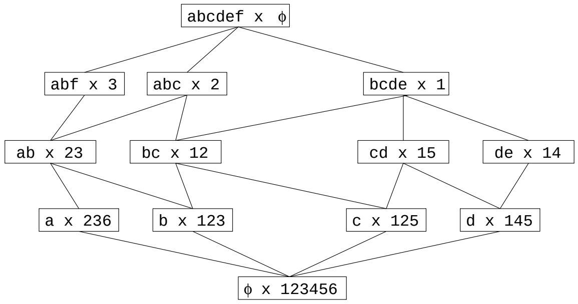

Figure 5. Concept lattice $\mathcal{L}(R'')$ of relation R'' from Example 3.2.

Another interesting aspect of weakly chordal graphs is that they are endowed with characterizing edge-addition construction schemes (see [12]), which makes it easy to maintain a modified binary relation with a weakly chordal underlying graph.

4 Conclusion

In this paper, we have characterized a new family of binary relations which define concept lattices of polynomial size, namely those associated with co-bipartite weakly chordal graphs. We feel that this is important, as little is known on the size of the lattice defined by a relation.

We have shown this to be useful to extract a poly-sized lattice from an exponential lattice by working on a sub-relation, but this can also help maintain relations with a small associated concept lattice in applications such as organizing object hierarchies or choosing samples for machine learning, since recognition of weakly chordal graphs is polynomial. The question remains open as to whether a binary relation with an underlying weakly chordal co-bipartite graph can constitute a correct sampling for a large arbitrary relation; and if this is not the case, it would be interesting to find out why, as this would mean that there is a semantic aspect behind the structure of the underlying graph.

As mentioned in our introduction, relations which define an exponential number of concepts tend to be dense (to have many ”ones”); the corresponding underlying graph would then have $O(n)$ edges (instead of $O(n^2)$), so a corresponding weakly chordal underlying subgraph would retain only $O(n)$ concepts. This may turn out to be insufficient, so it would be interesting to investigate a further generalization of weakly chordal graphs into a superclass which has a greater yet tractable number of minimal separators.

Our approach offers a generic tool for analyzing the structure of relations, as well as for estimating the size of the concept lattice defined by a given relation, as hundreds of graph classes are known and have been extensively studied, and this body of knowledge can now be put to use in the field of concept lattices.

Acknowledgment

We wish to thank the referees for many helpful suggestions which helped us to improve this paper.

REFERENCES

[1] M. Barbut and B. Monjardet. Ordre et classifi cation. *Classiques Hachette*, 1970.

[2] A. Berry. Désarticulation d'un graphe. *PhD Dissertation, Université Montpellier II*, 1998.

[3] A. Berry and J.-P. Bordat. Orthotreillis et séparabilité dans un graphe non-orienté. *Mathématiques, Informatique et Sciences Humaines*, 146:5–17, 1999.

[4] A. Berry and J.-P. Bordat. Triangulated and Weakly Triangulated Graphs: Simpliciality in Vertices and Edges. *Communication, 6th International Conference on Graph Theory (ICGT'00), 2000.*

[5] A. Berry, J.-P. Bordat and O. Cogis. Generating all the minimal separators of a graph. *International Journal of Foundations of Computer Science*, 11:397-404, 2000.

[6] A. Berry, J.-P. Bordat, and P. Heggernes. Recognizing weakly triangulated graphs by edge separability. In *Nordic Jour. Comput.*, vol 7, no 3, pp 164–177, 2000.

[7] A. Berry and A. Sigayret. Representing a concept lattice by a graph. *Workshop on Discrete Mathematics for Data Mining, Proc. 2nd SIAM Workshop on Data Mining, Arlington (VA), April 11–13, 2002.*

[8] G. Birkhoff. Lattice Theory. *American Mathematical Society*, 3rd Edition, 1967.

[9] B. Ganter. Two basic algorithms in concept analysis. *Preprint 831, Technische Hochschule Darmstadt*, 1984.

[10] M.C. Golumbic. Algorithmic Graph Theory and Perfect Graphs. *Academic Press*, New York, 1980.

[11] R. Hayward. Weakly triangulated graphs. *J. Comb. Theory*, 39:200–208, 1985.

[12] R. Hayward. Generating weakly triangulated graphs. *J. Graph Theory*, 21:67–70, 1996.

[13] R. Hayward, J. Spinrad, and R. Sritharan. Weakly chordal graph algorithms via handles. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'00), 2000*.

[14] M. Huchard, H. Dicky and H. Leblanc. Galois lattice as a framework to specify building class hierarchies algorithms. *Theoretical Informatics and Applications*, 34:521–548, 2000.

[15] T. Kloks and D. Kratsch. Listing all minimal separators of a graph. *SIAM Journal on Computing*, 27:605–613, 1998.

[16] L. Nourine and O. Raynaud. A Fast Algorithm for building Lattices. *Information Processing Letters*, 71:199–204, 1999.

[17] A. Parra. Structural and algorithmic aspects of chordal graph embeddings. *PhD Dissertation, Technische Universität Berlin*, 1996.

[18] J.L. Pfaltz and C.M. Taylor. Scientifi c Knowledge Discovery through Iterative Transformation of Concept Lattices. *Workshop on Discrete Mathematics for Data Mining, Proc. 2nd SIAM Workshop on Data Mining, Arlington (VA), April 11–13, 2002.*

[19] D. Rose, R.E. Tarjan, and G. Lueker. Algorithmic aspects of vertex elimination on graphs. *SIAM Journ. Comput.*, 5:146–160, 1976.

[20] H. Shen. Separators are as simple as cutsets. *Asian Computer Science Conference, Purket, Thailand, December 10-13, 1999; Lecture Notes in Computer Science*, 172:347–358, 1999.

[21] H. Sheng and W. Liang. Effi cient enumeration of all minimal separators in a graph. *Theoretical Computer Science*, 180: 169–180, 1997.

[22] P. Valtchef, R. Missaoui, and R. Godin. A Framework for Incremental Generation of Frequent Closed Item Sets. *Workshop on Discrete Mathematics for Data Mining, Proc. 2nd SIAM Workshop on Data Mining, Arlington (VA), April 11–13, 2002.*

[23] A. Yahia, L. Lakhal and J.-B. Bordat. Designing Class Hierarchies of Object Database Schemes. *Proceedings 13e journées Bases de Données avancées (BDA'97)*, 1997.

[24] M. J. Zaki, S. Parthasarathy, M. Ogihara and W. Li. New Algorithms for Fast Discovery of Association Rules. *Proceedings of 3rd Int. Conf. on Database Systems for Advanced Applications, April 1997.*