On the size of concept lattices

Yasuaki Kobayashi[1]

Abstract: In this paper, we give an upper bound of the size of concept lattices, where the size of a concept lattice is measured by the number of concepts plus the number of arcs in its line diagram. We show that the size of the concept lattice of a formal context (X, Y, R) is $2^{\min(|X|,|Y|)}(|X| + |Y|)^{O(1)}$, which is essentially tight up to a polynomial factor. To establish this bound, we analyze a combinatorial structure of concept lattices through the theory of minimal separators and potential maximal cliques. More specifically, we give a characterization of arcs in the line diagram of concept lattices by means of potential maximal cliques in an associated co-bipartite graph, which might be of interest in its own right.

Keywords: Concept Lattice, Formal Concept Analysis, Minimal Separator, Potential Maximal Clique

1. Introduction

Concept lattices, also known as *Galois lattices*, are widely used in several areas such as association rule mining [8], [13], frequent itemset generation [12], information retrieval [9], software engineering [11], social network analysis [10], and web document management [4]. In such applications, a crucial and time consuming step is building the concept lattice from a given data. Given this, many algorithms for building concept lattices are proposed in the literature [5], [7]. However, to the best of the author's knowledge, no non-trivial upper bound on the size of concept lattices is known, where the size of a concept lattice is measured by the sum of the number of concepts and the number of arcs in the Hasse diagram of the concept lattice.

In this paper, we give, for the first time, a non-trivial upper bound on the size of the concept lattice of a formal context (X, Y, R).

Theorem 1. *Let (X, Y, R) be a formal context. The size of the concept lattice is $2^{\min(|X|,|Y|)}(|X| + |Y|)^{O(1)}$. Moreover, this upper bound is tight up to a polynomial factor in $|X| + |Y|$.*

To show this bound, we exploit the result of Berry and Sigayret [1]. They gave a method of analyzing concept lattices by using the theory of *minimal separators*. We borrow some tools to analyze concept lattices given by [1] and we extend their tools by exploiting the theory of *potential maximal cliques*. More specifically, we give a characterization of the covering relations between two concepts in concept lattices.

2. Preliminaries

In this section, we give some notations and terminologies used throughout our paper.

Let G be an undirected graph. We denote by $V(G)$ the set of vertices of G and by $E(G)$ the set of edges of G. Let $X \subseteq V(G)$.

We write $N(X)$ to denote the set of neighbors of X, that is, $N(X) = \{u \in V(G) \setminus X : \{u, v\} \in E(G)$ for some $v \in X\}$. The subgraph of G induced on X is denoted by $G[X]$.

2.1 Concept Lattices

Let X and Y be non-empty sets and let R be a binary relation, that is, $R \subseteq X \times Y$. We usually call X the set of *objects* and Y the set of *attributes*. We say that an object $x \in X$ has an attribute $y \in Y$ if $(x, y) \in R$. The ordered triplet (X, Y, R) is called a *formal context*. Let $K = (X, Y, R)$ be a formal context. Consider two operators over K. Operators $^{\uparrow} : 2^X \to 2^Y, ^{\downarrow} : 2^Y \to 2^X$ are defined as: for $A \subseteq X$ and for $B \subseteq Y$,

$$A^{\uparrow} = \{y \in Y : (x, y) \text{ for every } x \in A\}$$
$$B^{\downarrow} = \{x \in X : (x, y) \text{ for every } y \in B\},$$

that is, $A^{\uparrow}$ is the set of all attributes that every object in A has and $B^{\downarrow}$ is the set of all objects that have all attribute in B.

Let $A \subseteq X$ and $B \subseteq Y$. We say that the pair (A, B) is a *formal concept* (or simply *concept*) of K if $A^{\uparrow} = B$ and $B^{\downarrow} = A$. Note that two pairs $(\emptyset, Y)$ and $(X, \emptyset)$ are concepts. We say that a concept is *trivial* if it is one of the above two concepts. A concept is *non-trivial* if it is not a trivial concept. When $C = (A, B)$ is a concept of K, A is called the *extent* of C and B is called the *intent* of C.

Consider a partial order $\leq$ on the set of concepts in a formal context K: for contexts (A, B) and (A', B'), $(A, B) \leq (A', B')$ if and only if $A \subseteq A'$. Equivalently, $(A, B) \leq (A', B')$ if and only if $B \supseteq B'$. It is well-known that this partial order relation $\leq$ on the set of concepts forms a complete lattice, called a *concept lattice* [14]. In this lattice, we say that a concept $C = (A, B)$ is a *predecessor* of a concept $C' = (A', B')$ if $A' \subseteq A$ (and hence $B \subseteq B'$). In particular, C is an *immediate predecessor* of C' if C is a predecessor of C' and there is no concept $C'' = (A'', B'')$ satisfying $A' \subset A'' \subset A$ (and hence $B \subset B'' \subset B'$).

2.2 Minimal Separators

Berry and Sigayret [1] gave several tools for analyzing concept

[1] Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606–8501, Japan

lattices. The key to their results is exploiting results of minimal separators.

Let S be a vertex set of an undirected graph G. We say that S is an a, b-*separator* of G if there is no path between a and b in $G[V(G)\backslash S]$. Furthermore, S is said to be a *minimal a, b-separator* of G if no proper subset of S is an a, b-separator of G. A *minimal separator* of G is a vertex set that is a minimal a, b-separator of G for some $a, b \in V(G)$.

It is well-known that minimal separators can be characterized by using the notion of full components. Let S be a vertex set and let C be a component of $G[V(G)\backslash S]$. Then we call C a *component associated to* S. We say that C is a *full* component associated to S if $N(C) = S$.

Lemma 1 (forklore). *A separator S is a minimal separator of G if and only if there are at least two full components associated to S.*

2.3 Potential Maximal Cliques

Dirac [3] gave a characterization of the class of chordal graphs by using minimal separators. Minimal separators were also used for computing the treewidth and a minimum fill-in of several classes of graphs [6]. This argument was extended by Bouchitté and Todinca [2]. The crucial idea of their results is using the notion of *potential maximal cliques*. A set $\Omega \subset V(G)$ is called a *potential maximal clique* if there is a minimal triangulation of G in which Ω is a maximal clique. Bouchitté and Todinca gave the following characterization for potential maximal cliques.

Lemma 2 ([2]). *Let Ω be a vertex set of H and let Q be the set of component associated to Ω. Then Ω is a potential maximal clique of H if and only if*

(1) there is no full component in Q and

(2) Ω is a clique in the graph obtained from H by adding an edge between every pair of vertices in $N(C)$ for each component $C \in Q$.

Moreover, for each $C \in Q$, $N(C)$ is a minimal separator of G.

3. Main Result

This section is dedicated to give an upper bound on the size of concept lattices. Here, we consider the size of a lattice as the sum of the number of concepts and the number of arcs in the Hasse diagram of the lattice.

Let $K = (X, Y, R)$ be a formal context. We assume that for every $x \in X$, there is $y \in Y$ with $(x, y) \in R$, and for every $y \in Y$, there is $x \in X$ with $(x, y) \in X$. Let $\mathcal{L}$ be the concept lattice of the formal context K. Since any two concepts have different extent and different intent, we obviously have the following upper bound on the number of concepts.

Lemma 3. *The number of concepts in $\mathcal{L}$ is at most $2^{\min(|X|,|Y|)}$.*

This upper bound is essentially tight. To see this, we need a characterization due to [1]. Let $G_{\mathcal{L}}$ be a co-bipartite graph with vertex set $X \cup Y$, where both $G_{\mathcal{L}}[X]$ and $G_{\mathcal{L}}[Y]$ are cliques, and there is an edge between $x \in X$ and $y \in Y$ if $(x, y) \notin R$. Berry and Sigayret [1] showed that there is a one-to-one correspondence between the set of non-trivial concepts in $\mathcal{L}$ and the set of minimal separators in $G_{\mathcal{L}}$.

Lemma 4 ([1]). *Let $C = (A, B)$. C is a non-trivial concept in $\mathcal{L}$ if*

and only if $V(G_{\mathcal{L}}) \backslash (A \cup B)$ is a minimal separator of $G_{\mathcal{L}}$. Moreover both A and B are full components associated to this minimal separators.

Therefore the number of concepts in $\mathcal{L}$ is exactly equal to the number of minimal separators of $G_{\mathcal{L}}$ plus two (trivial concepts). Conversely let G be a co-bipartite graph with bipartition (X, Y) of the vertex set. We say that G is *well-formed* if $N(\{x\}) \cap Y \neq \emptyset$ for every $x \in X$ and $N(\{y\}) \cap X \neq \emptyset$ for every $y \in Y$. Let $R_G \subseteq X \times Y$ be a binary relation such that $(x, y) \in R_G$ for every pair $\{x, y\} \notin E(G)$ with $x \in X$ and $y \in Y$. Lemma 4 shows in fact that $S \subseteq V(G)$ is a minimal separator of G if and only if $(X \backslash S, Y \backslash S)$ is a concept in the formal context (X, Y, R_G). Thus we immediately have the following corollary.

Corollary 1. *Let G be a well-formed co-bipartite graph with bipartition (X, Y). Then the number of minimal separators in G is at most $2^{\min(|X|,|Y|)}$.*

To see a lower bound, let us consider a co-bipartite graph $G = (X \cup Y, E)$ with $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$ such that $E = \{\{x_i, x_j\} : 1 \leq i < j \leq n\} \cup \{\{y_i, y_j\} : 1 \leq i < j \leq n\} \cup \{\{x_i, y_i\} : 1 \leq i \leq n\}$. In other words, G forms two cliques of the same size and has a perfect matching between cliques. This co-bipartite graph is indeed well-formed. To separate x_1 and y_n in G at least one of x_i and y_i must be chosen as a separator for each $1 < i < n$. On the other hand, for any x_1, y_n-separator S with $\{x_i, y_i\} \subseteq S$, both $S \backslash \{x_i\}$ and $S \backslash \{y_i\}$ are also x_1, y_n-separators. Thus every minimal x_1, y_n-separator contains exactly one of x_i and y_i for each $1 < i < n$. This implies the number of minimal separators is at least 2^{n-2}, which is tight up to a constant multiplicative factor for the bound in Lemma 3.

In the rest of this section, we give an upper bound on the number of pairs of concepts C and C' such that C is an immediate successor of C'. To do so, we need to show several lemmas.

Lemma 5. *Let $C = (A, B)$, $C' = (A', B')$ be concepts in $\mathcal{L}$ such that $A' \subset A$ (and hence $B \subset B'$). Then C is an immediate predecessor of C' if and only if $(A \backslash A') \cup (B' \backslash B)$ is a clique in $G_{\mathcal{L}}$.*

Proof. Let $W = A \backslash A'$ and let $Z = B' \backslash B$. As C is a predecessor of C', both W and Z are non-empty.

To prove the forward direction, suppose that C is an immediate predecessor of C'. Moreover suppose, for contradiction, that there are $w \in W$ and $z \in Z$ that are not adjacent to each other. Consider the complement bipartite graph H of $G_{\mathcal{L}}[W \cup Z]$. Since w and z are adjacent to each other in H, there is a maximal biclique C in H that contains both w and z. Define $A'' = A' \cup (V(C) \cap X)$ and $B'' = B \cup (V(C) \cap Y)$. Observe that, in G, every vertex in $x \in X \backslash A''$ is adjacent to some vertex in B'' since otherwise $C \cup \{x\}$ is a biclique of H with $\{w, z\} \subseteq V(C)$, which contradicts the maximality of C. Therefore B'' is a full component associated to $V(G_{\mathcal{L}}) \backslash (A'' \cup B'')$. A symmetric argument shows that A'' is also a full component. By Lemma 1, $V(G_{\mathcal{L}}) \backslash (A'' \cup B'')$ is a minimal separator of $G_{\mathcal{L}}$. This implies that (A'', B'') is a non-trivial concept of $\mathcal{L}$ with $A' \subset A'' \subset A$ and $B \subset B'' \subset B'$. This is contradicting to the fact that C is an immediate predecessor of C'.

Conversely, suppose that $W \cup Z$ is a clique in $G_{\mathcal{L}}$. If C is not an immediate predecessor of C', there must be a concept

$C'' = (A'', B'')$ with $A' \subset A'' \subset A$ and $B \subset B'' \subset B'$. Since C'' is non-trivial, by Lemma 4, $V(G_{\mathcal{L}}) \setminus (A'' \cup B'')$ is a minimal separator of $G_{\mathcal{L}}$. This implies that there are no edges between $A'' \setminus A'$ and $B'' \setminus B$. This is contradicting to the fact that $W \cup Z$ is a clique in $G_{\mathcal{L}}$. □

Lemma 6. *Let $C = (A, B), C' = (A', B')$ be concepts of a concept lattice $\mathcal{L}$. Then C is an immediate predecessor of C' if and only if $V(G_{\mathcal{L}}) \setminus (A' \cup B)$ is a potential maximal clique of $G_{\mathcal{L}}$.*

Proof. Suppose C is an immediate predecessor of C'. Let $\Omega = V(G_{\mathcal{L}}) \setminus (A' \cup B)$. We will show that Ω is a potential maximal clique of $G_{\mathcal{L}}$. Let C be a component associated to Ω. Observe that there are no edges between A' and B'. This follows from Lemma 4 when C' is a non-trivial concept and from the fact that either A' or B' is empty when C' is a trivial one. As $B \subset B'$, there are no edges between A' and B as well. Therefore either $C = A'$ or $C = B$. If C is a trivial concept, B must be empty, and if C' is a trivial concept, A' must be empty. Thus C is an extent or an intent of a non-trivial concept, and hence, by Lemma 4, $N(C)$ is a minimal separator of $G_{\mathcal{L}}$ and C is a full component associated to it. When $C = A'$, $N(A')$ is a minimal separator that separates A' and B'. Since Ω contains $B' \setminus B$, $N(C)$ is strictly contained in Ω. Therefore C is not a full component associated to Ω. The same consequence also holds when $C = B$. Hence Ω has no full component associated to it.

Let $G'_{\mathcal{L}}$ be a graph obtained from $G_{\mathcal{L}}$ by adding an edge between every pair of vertices in $N(A')$ and that in $N(B)$. In order to prove that Ω is a potential maximal clique, it suffices to show that Ω is a clique in $G'_{\mathcal{L}}$. Suppose first that C is a trivial concept. Then we have $A = X$ and $B = \emptyset$. Let us note that, in this case, $\Omega = (A \setminus A') \cup Y$. We have the fact that Ω is a clique in $G'_{\mathcal{L}}$ from Lemma 5. Applying a symmetric argument to the case where C' is a trivial concept proves Ω is a clique in $G'_{\mathcal{L}}$. Suppose next that either C or C' is a non-trivial concept. By Lemma 4, $V(G_{\mathcal{L}}) \setminus$ is a minimal separator of $G_{\mathcal{L}}$, and B is a full component associated to it. Thus we have that $X \setminus A \subseteq N(B)$. Similarly, we have that $Y \setminus B' \subseteq N(A')$. As $G_{\mathcal{L}}$ is co-bipartite, we also have $X \setminus A' \subseteq N(A')$ and $Y \setminus B \subseteq N(B)$. This gives the fact that for every pair $(x, y) \in (X \setminus A') \times (Y \setminus B)$, if at least one of $x \in X \setminus A$ or $y \in Y \setminus B'$ holds, then x and y are adjacent to each other in $G'_{\mathcal{L}}$. Moreover if both $x \in A \setminus A'$ and $y \in B' \setminus B$, by Lemma 5, x and y are adjacent to each other in $G_{\mathcal{L}}$. Therefore Ω is a clique in $G'_{\mathcal{L}}$.

To prove the other direction suppose $\Omega = V(G_{\mathcal{L}}) \setminus (A' \cup B)$ is a potential maximal clique of $G_{\mathcal{L}}$. We will show that $C = (B^{\downarrow}, B)$ and $C' = (A', A'^{\uparrow})$ are concepts of $\mathcal{L}$, and moreover C is an immediate predecessor of C' in $\mathcal{L}$.

Obviously C is a concept when $B = \emptyset$. So is C' when $A' = \emptyset$. Thus we suppose otherwise. As Ω is a potential maximal clique of $G_{\mathcal{L}}$, by Lemma 2, $N(B)$ is a minimal separator. Let us note that $B^{\downarrow}$ is the set of vertices in X, each of which has no neighbor in B. Therefore $N(B) = (X \setminus B^{\downarrow}) \cup (Y \setminus B)$. By Lemma 4, $(B^{\downarrow}, B)$ is a non-trivial concept in $\mathcal{L}$. A symmetric argument proves that $(A', A'^{\uparrow})$ is a non-trivial concept as well.

Now we will prove that C is an immediate predecessor of C'. To do this, by Lemma 5, it is sufficient to show that $(B^{\downarrow} \setminus A') \cup (A'^{\uparrow} \setminus B)$ is a clique in $G_{\mathcal{L}}$. As $G_{\mathcal{L}}$ is co-bipartite,

both $B^{\downarrow} \setminus A'$ and $A'^{\uparrow} \setminus B$ are cliques in $G_{\mathcal{L}}$. Let $x \in B^{\downarrow} \setminus A'$ and $y \in A'^{\uparrow} \setminus B$ be arbitrary. Let $G'_{\mathcal{L}}$ be a graph obtained from $G_{\mathcal{L}}$ by adding an edge between every pair of vertices in $N(A')$ and every pair of vertices in $N(B)$. Since Ω is a potential maximal clique, by Lemma 2, x and y are adjacent in the filled graph $G'_{\mathcal{L}}$. However observe that $x \notin N(B)$. This follows from that fact that $B^{\downarrow}$ is the set of vertices of X that have no neighbors in B. We can show that $y \notin N(A)$ by a symmetric argument. Therefore x and y are adjacent in $G_{\mathcal{L}}$. Hence the lemma follows. □

We are now ready to give our upper bound on the size of concept lattice. The above lemma says that it suffices to consider the number of potential maximal cliques in the associated well-formed co-bipartite graph $G_{\mathcal{L}}$ of $\mathcal{L}$. The following lemma give such an upper bound.

Lemma 7. *Let G be a well-formed co-bipartite graph with bipartition (X, Y). Then the number of potential maximal cliques in G is $2^{\min(|X|,|Y|)}(|X| + |Y|)^{O(1)}$, where n is the number of vertices in G.*

Proof. The lemma immediately follows, together with Corollary 1, from the following claim: if Ω is a potential maximal clique of G, then there are $x \in X \cap \Omega$ and $y \in Y \cap \Omega$ such that $\Omega \setminus \{x, y\}$ is a minimal x, y-separator of G', where G' is a co-bipartite graph obtained from G by removing the edge between x and y. In the following we will prove this claim.

Let $R_G = \{(x, y) \in X \times Y : \{x, y\} \notin E(G)\}$, and let $\mathcal{L}$ be the concept lattice of a formal context (X, Y, R_G). By Lemma 6, $Y \setminus \Omega$ is an intent of a concept C and $X \setminus \Omega$ is an extent of another concept C' in $\mathcal{L}$. Moreover C is an immediate predecessor of C'. Let $C = (A, B)$ and $C' = (A', B')$, where $B = Y \setminus \Omega, A' = X \setminus \Omega, A = B^{\downarrow}$, and $B' = A'^{\uparrow}$. Choose x from $A \setminus A'$ and y from $B' \setminus B$. This can be done since C is a predecessor of C', that is, $A' \subset A$ and $B' \subset B$. Let C_x and C_y be the components of $G'[V(G') \setminus (\Omega \setminus \{x, y\})]$ that contains x and y, respectively. Since there are no edges between A' and B in G', C_x and C_y must be different. Moreover we have that $C_x = A' \cup \{x\}$ and $C_y = B \cup \{y\}$. To prove the aforementioned claim, we show that both C_x and C_y are full components associated to $S = V(G) \setminus (C_x \cup C_y)$ in G'. Let us first consider the cases where $C_x = \{x\}$ or $C_y = \{y\}$. In these cases, say here $C_x = \{x\}$, by Lemma 5, x is adjacent to every vertex in $B' \setminus B$, and as $B' = Y$, we have $S = N(C_x) \setminus \{y\}$, that is, C_x is a full component associated to S in G'. For the other case, suppose C and C' are not trivial concepts. In this case, by Lemma 4, A' and B are full components associated to $S_{C'} = V(G) \setminus (A' \cup B')$ and $S_C = V(G) \setminus (A \cup B)$ in G, respectively. This implies that every vertex in $S_{C'}$ has a neighbor in A' and every vertex in S_C has a neighbor in B. By Lemma 5, x is adjacent to every vertices in $B' \setminus B$. This implies that $S = S_{C'} \cup (B' \setminus (B)) \setminus \{y\}$ has a neighbor in C_x. Thus C_x is a full component associated to S in G'. We can analogously show that C_y is a full component associated to S in G'. Therefore, by Lemma 4, $\Omega \setminus \{S\}$ is an x, y-minimal separator in G' and hence the lemma holds. □

4. Remarks

In this paper, we show that the size of a concept lattice of a

formal context (X, Y, R) is $2^{\min(|X|,|Y|)}(|X| + |Y|)^{O(1)}$, where the size of concept lattices is measured by the number of concepts plus the number of arcs in its Hasse diagram. We also show that this bound is tight up to a polynomial factor.

References

[1] A. Berry, A. Sigayret: Representing a concept lattice by a graph. *Discrete Applied Mathematics* 144(1-2), 27–42, 2004.

[2] V. Bouchitté, I. Todinca: Treewidth and Minimum Fill-in: Grouping the Minimal Separators. *SIAM J. Comput.* 31(1), 212–232, 2001.

[3] G.A. Dirac: On rigid circuit graphs. *Abh. Math. Sem. Univ. Hamburg* 25, 71–76, 1961.

[4] T.J. Everts, S.S. Park, B.H. Kang: Using formal concept anaysis with an incremental knowledge acquisition system for web document management. In *Proceedings of ACSC '06*, pp. 247–256, 2006

[5] M. Farach-Colton, Y. Huang: A Linear Delay Algorithm for Building Concept Lattices. In *Proceedings of CPM '08*, pp. 204–216, 2008.

[6] T. Kloks, D. Kratsch, J. Spinrad: On treewidth and minimum fill-in of asteroidal triple-free graphs. *Theor. Comput. Sci.* 175(2), 309–335, 1997.

[7] L. Nourine, O. Raynaud: A fast algorithm for building lattices. *Information Processing Letters* 71, 199–204, 1999.

[8] J.L. Pfaltz, C.M. Taylor: Scientific knowledge discovery through iterative transformation of concept lattices. In *Proceedings of the 2nd SIAM Workshop on Discrete Mathematics and Data Mining*, 2002.

[9] G. Polaillon, M.A. Aufaure, B.L. Grand, M. Soto: FCA for Contextual Semantic Navigation and Information Retrieval in Heterogeneous Information Sysmtems. In: *Proceedings of DEXA '07*, pp. 534–539, 2007.

[10] V. Snášel, Z. Horák, J. Kočíbová, A. Abraham: Analyzing social networks using FCA: complexity aspects. In *Proceedings of WI-IAT '09*, pp. 38–41, 2009.

[11] G. Snelting, F. Tip: Reengineering class hierarchies using concept analysis. In *Proceedings of FSE '98*, pp 99–110, 1998.

[12] P. Valtchef, R. Missaoui, R. Godin: A Framework for incremental generation of frequent closed item sets. In *Proceedings of the 2nd SIAM Workshop on Discrete Mathematics and Data Mining*, 2002.

[13] Y.-Y. Wang, X.-G. Hu: A fast algorithm for mining association rules based on concept lattice. In *Proceedings of ICMLC '04*, pp. 1687–1691, 2004.

[14] R. Wille: Restructing Lattice Theory: An Approach Based on Hierarchies of Concepts. In *Ordered Sets*, pp. 445–470, 1982.