BOLYAI SOCIETY
MATHEMATICAL STUDIES

Combinatorics,
Paul Erdős is Eighty (Volume 1)
Keszthely (Hungary), 1993, pp. 91–105.

On the Number of Closure Operations

G. BUROSCH, J. DEMETROVICS*, G. O. H. KATONA*,
D. J. KLEITMAN and A. A. SAPOZHENKO

[5] introduces a natural partial order for the closure operations in a fixed finite n-element set. It is shown that this partially set is ranked. In the present paper estimates are given on the number of closure operations of fixed rank. On the other hand, asymptotical results are given on the number of closure operations determined uniquely by the family of subsets whose closure is the underlying set (keys).

1. Introduction

Codd [7] and Armstrong [2] introduced the system of functional dependencies as a model of a database. We, however, prefer another equivalent variant, the closure operation (see e.g. [5]). Let X be the (finite) set of *attributes*, that is, the set of types of data. The elements of X are words like "name", "date of birth", "age", *etc.* Some of the data determine some other data uniquely. For instance, the date of birth determines the age. Let $A \subseteq X, a \in X, a \notin A$. We say that A *determines* a and write $A \to a$ iff the set of data in A determines the data in a, more precisely, there are no two individuals having the same data in A and different in a. The function $\mathcal{L} : 2^X \to 2^X$ is defined by

$$\mathcal{L}(A) = \{a : A \to a\}.$$

---

*The work of these authors is supported by the Hungarian National Foundation for Scientific Research, grant number 2575

This function obviously possesses the following properties:

$$A \subseteq \mathcal{L}(A),$$

$$A \subseteq B \quad \text{implies} \quad \mathcal{L}(A) \subseteq \mathcal{L}(B),$$

$$\mathcal{L}(\mathcal{L}(A)) = \mathcal{L}(A).$$

Such a function is known as a *closure operation* or briefly a *closure.* Therefore a closure is a possible model of a database.

$$\mathcal{L}(\emptyset) = \emptyset$$

is a rather natural assumption for closures formed from databases. In the present paper we will use the name closure for the functions satisfying this additional condition.

The most natural question concerning a poset is to determine its size. The present authors investigated this question in [6]. We proved that the number $\alpha(n)$ of closures on an n-element set satisfies the following inequalities:

$$2^{\binom{n}{\lfloor \frac{n}{2} \rfloor}} \leq \alpha(n) \leq 2^{2\sqrt{2}\binom{n}{\lfloor \frac{n}{2} \rfloor}(1+o(1))}.$$

This has been substantially improved by Alekseev [1] showing that

$$\alpha(n) = 2^{\binom{n}{\lfloor \frac{n}{2} \rfloor}(1+o(1))}$$

as we conjectured.

In [5] we introduced a partially ordered set $\mathbf{P}$ of the closures as a model of changing databases:

$$\mathcal{L}_1 \leq \mathcal{L}_2 \quad \text{iff} \quad \mathcal{L}_1(A) \supseteq \mathcal{L}_2(A) \quad \text{holds for all} \quad A \subseteq X.$$

The integer valued function $r(x), x \in \mathbf{P}$ is a *rank function* iff $0 \leq r(x)$ for all $x \in \mathbf{P}$, $r(x) = 0$ for some $x \in \mathbf{P}$ and $x < y, x, y \in \mathbf{P}$, $\nexists\, z : x < z < y$ imply $r(y) = r(x) + 1$. It has been proved in [5] that the partially ordered set $\mathbf{P}$ of the closures has a rank function and the maximum rank is $2^n - 2$. The set of elements with rank k is called the k^{th} *level.* In [5] we investigated the minimum (maximum) number of immediate neighbours (degree in the Hasse-diagram) at a given level.

In Section 2 we give the order of magnitude of the number of elements in the k^{th} and $(2^n - 2 - k)^{\text{th}}$ level, resp., for fixed k, and for $n \to \infty$. It is interesting that the poset is much thinner at the top. the size of the k^{th} level is exponential in k while the size of the $(2^n - 2 - k)^{\text{th}}$ level is a polynomial of k.

The sets A satisfying $\mathcal{L}(A) = X$ are called the *keys* of the closure $\mathcal{L}$. This concept has obviously an important role for the databases. The data of a key determine the individual uniquely. The really important ones are the *minimal keys:* they are keys containing no other key as a proper subset. The family of all minimal keys is denoted by $\mathcal{K} = \mathcal{K}(\mathcal{L})$. It is easy to see that $A, B \in \mathcal{K}, A \neq B$ imply $A \not\subset B$. The families satisfying this condition are called *Sperner families.* $\mathcal{K}$ is a *saturated Sperner family* iff it is Sperner but $\mathcal{K} \cup \{A\}$ is non-Sperner for all $A \in 2^X - \mathcal{K}$. It is proved in [5] that for any non-empty Sperner family $\mathcal{S}$ there is a closure $\mathcal{L}$ such that $\mathcal{K}(\mathcal{L}) = \mathcal{S}$. If $\mathcal{K}$ is saturated then $\mathcal{L}$ is unique. The converse is not true. An example of a non-saturated Sperner family is given in [5] which determines the closure uniquely. Section 3 investigates the asymptotic behaviour of the number of these Sperner families, determining the closure $\mathcal{L}$ uniquely.

2. On the number of closures of fixed rank

Let $\mathcal{L}$ be a closure. The *closed sets* C are defined by $\mathcal{L}(C) = C$. It is known that the family $\mathcal{Z} = \mathcal{Z}(\mathcal{L})$ of closed sets possesses the following properties:

$$\emptyset, X \in \mathcal{Z}, \tag{2.1}$$

$$A, B \in \mathcal{Z} \quad \text{implies} \quad A \cap B \in \mathcal{Z}. \tag{2.2}$$

The families satisfying (2.1) and (2.2) are called *intersection semi-lattices.* It is shown in [8] that the function $f : \mathcal{L} \to \mathcal{Z}(\mathcal{L})$ is a bijection between the set of closures and the set of intersection semi-lattices. Thus the number of closures is equal to the number of intersection semi-lattices. It is proved in [5] that the partially ordered set $\mathbf{P}$ defined in the introduction has a rank function and the rank of a closure $\mathcal{L} \in \mathbf{P}$ is equal to the *number of closed sets with respect to $\mathcal{L}$ minus 2.* We have to subtract 2 because the smallest closure must have rank 0 but it has two closed sets ($\emptyset$ and X). So the number $\alpha(n, k)$ of closures on n elements with rank k is equal to the number of intersection semi-lattices on n elements and consisting of $k + 2$ members. E.g. $\alpha(n, 0) = 1$ and the only closure of rank 0 has two closed sets: $\emptyset$ and X. If $k = 1$, then the number of closed sets is 3, they are $\emptyset, X$ and any set A different from them. Therefore $\alpha(n, 1) = 2^n - 2$.

Let $\mathcal{Z} = \{Z_1, \ldots, Z_m\}$ be a family of subsets. The non-empty ones of the sets

$$Z_1^{\varepsilon_1} \cap Z_2^{\varepsilon_2} \cap \ldots Z_m^{\varepsilon_m}$$

are called *atoms,* where ε is either 0 or 1 and $Z^1 = Z, Z^0 = \overline{Z}$. It is easy to see that the atoms form a partition of X.

Lemma 1. *If $\mathcal{Z}$ is an intersection semi-lattice of $k+2$ members then it has at most $k+1$ atoms.*

Proof. We use induction on k. If $k = 0$ then the statement is trivial. Suppose that $k > 0$ and that the statement is true for smaller values of k. Take a member $A \in \mathcal{Z}$ such that $A \neq X$ and there is no $B \in \mathcal{Z}, B \supset A, B \neq X$. We may suppose that $\mathcal{Z} = \{Z_1, \ldots, Z_{k+2}\}$ where $A = Z_{k+2}$. Consider a fixed sequence $(\varepsilon_1, \ldots, \varepsilon_{k+1})$. Suppose that $\varepsilon_i = 1$ for some i $(1 \leq i \leq k+1)$. Then $Z_i^{\varepsilon_i} \cap A \in \mathcal{Z}$ and consequently $Z_i^{\varepsilon_i} \cap A = \mathcal{Z}_j$ hold for some j $(1 \leq j \leq k+1)$. This implies

$$\mathcal{Z}_j \cap Z_i^{\varepsilon_i} \cap A = \mathcal{Z}_j \cap Z_i^{\varepsilon_i},$$

$$\overline{\mathcal{Z}}_j \cap Z_i^{\varepsilon_i} \cap \overline{A} = \overline{\mathcal{Z}}_j \cap Z_i^{\varepsilon_i},$$

and either

$$\bigcap_{r=1}^{k+1} Z_r^{\varepsilon_r} \cap A = \bigcap_{r=1}^{k+1} Z_r^{\varepsilon_r} \tag{2.3}$$

or

$$\bigcap_{r=1}^{k+1} Z_r^{\varepsilon_r} \cap \overline{A} = \bigcap_{r=1}^{k+1} Z_r^{\varepsilon_r}. \tag{2.4}$$

It is obvious that the atoms of $\mathcal{Z}$ form a refinement of the partition defined by $\mathcal{Z}-\{A\}$. However (2.3) and (2.4) show that the atoms of $\mathcal{Z}-\{A\}$ defined by at least one $\varepsilon_i = 1$ are also atoms of $\mathcal{Z}$. The only atom of $\mathcal{Z}-\{A\}$ which could be cut into two parts is

$$\overline{Z}_1 \cap \overline{Z}_2 \cap \ldots \cap \overline{Z}_{k+1}.$$

Therefore the number of atoms could be increased by at most one. The number of atoms of $\mathcal{Z} - \{A\}$ is at most k, since it is an intersection semi-lattice and hence the number of atoms of $\mathcal{Z}$ is at most $k+1$. ■

We call two families $\mathcal{Z}_1$ and $\mathcal{Z}_2$ *weakly isomorphic* (in notation $\mathcal{Z}_1 \approx \mathcal{Z}_2$) iff there are bijections f and g between the sets of atoms and members, resp., such that $f(A) \subseteq g(Z)$ iff $A \subseteq Z$ (A is an atom of $\mathcal{Z}_1$, $f(A)$ is an atom of $\mathcal{Z}_2$, $Z \in \mathcal{Z}_1, g(Z) \in \mathcal{Z}_2$). A family $\mathcal{A}$ is called *atomic* iff all of its atoms have size 1. $\mathcal{Z}_1$ is a *contraction* of $\mathcal{Z}_2$ iff $\mathcal{Z}_1 \approx \mathcal{Z}_2$ and $\mathcal{Z}_1$ is atomic. It is easy to see that the weak isomorphism is an equivalence relation and that there is an atomic family in each equivalence class. On the other hand, it is clear that one class contains either only intersection semi-lattices or none of its elements has this property. Denote by $\mathbf{C}_n(\mathcal{Z})$ the class of families on n elements, weakly isomorphic to $\mathcal{Z}$.

Denote by $\mathbf{I}(n,k)$ the set of intersection semi-lattices on n elements and consisting of $k+2$ members. Partition this set collecting all elements, having the same (= strongly isomorphic) contraction, into one class:

$$\mathbf{I}(n,k) = \bigcup_{\mathcal{A} \in \mathbf{R}} \mathbf{C}_n(\mathcal{A}) \tag{2.5}$$

where $\mathbf{R}$=$\mathbf{R}(n,k)$ is a set of atomic families representing each equivalence class of $k+2$-member intersection semi-lattices on n elements, exactly once.

Let $\mathcal{A} \in \mathbf{R}$ and use the notation $\{a_1, a_2, \ldots, a_r\}$ for the groundset of $\mathcal{A}$. Denote by $s(\mathcal{A})$ the number of permutations $i_1, \ldots, i_r$ such that the permutation $a_j \to a_{i_j}$ brings $\mathcal{A}$ into itself. (That is, $s(\mathcal{A})$ is the order of the group of automorphism of $\mathcal{A}$.)

Lemma 2. *Suppose that $\mathcal{A} \in \mathbf{R}$ is a family on r elements. Then*

$$|\mathbf{C}_n(\mathcal{A})| \sim \frac{r^n}{s(\mathcal{A})}$$

holds (r is fixed, $n \to \infty$).

Proof. Let $\{a_1, a_2, \ldots, a_r\}$ be the groundset of $\mathcal{A}$. Suppose that $\mathcal{Z} \in \mathbf{C}_n(\mathcal{A})$ and that its groundset is $X = \{x_1, x_2, \ldots, x_n\}$. $\mathcal{A}$ and $\mathcal{Z}$ and are weakly isomorphic, therefore there are functions f and g mapping their atoms and members, respectively, preserving inclusion. Define the function $h(x)$ $(x \in X)$ by $h(x) = i$ iff $x \in f(a_i)$. In this way we defined a function h for each pair of (weakly isomorphic) $\mathcal{A}$ and $\mathcal{Z}$. $\mathcal{A}$ can be chosen in $s(\mathcal{A})$ ways, $\mathcal{Z}$ can be chosen in $|\mathbf{C}_n(\mathcal{A})|$ ways, independently. That is, the number of such functions is $s(\mathcal{A})|\mathbf{C}_n(\mathcal{A})|$. The atoms are non-empty, so the functions take on all possible values from 1 to r. On the other hand, as it is easy to see, all such functions can be obtained from some pair $\mathcal{A}, \mathcal{Z}$. The number of functions defined on an n-element set, taking on integer values from 1 to r and having all values at least once can be obtained by the inclusion-exclusion formula:

$$r^n + \sum_{i=1}^{r} \binom{r}{i} (r-i)^n (-1)^i = s(\mathcal{A})|\mathbf{C}_n(\mathcal{A})|.$$

The statement of the lemma easily follows. ■

Let us introduce the following notation:

$$\eta(k) = \sum_{\mathcal{A} \in \mathbf{S}_k} \frac{1}{s(\mathcal{A})},$$

where $\mathbf{S}_k$ is the set of non-isomorphic atomic intersection semi-lattices with $k+2$ members and exactly $k+1$ atoms.

Theorem 1. *The number $\alpha(n,k)$ of closures on n elements and having rank k satisfies*

$$\alpha(n,k) \sim \eta(k)(k+1)^n$$

where k is fixed, $n \to \infty$.

Proof. (2.5) implies

$$\alpha(n,k) = \sum_{\mathcal{A} \in \mathbf{R}} |\mathbf{C}_n(\mathcal{A})|.$$

The number of terms depends only on k, therefore their asymptotic behaviour can be considered separately. The asymptotic behaviour is given in Lemma 2. Only the terms with $r = k+1$ count, by Lemma 1, the rest is $o((k+1)^n)$. Collecting these terms we obtain the statement of the theorem. ■

Now we will investigate $\alpha(n, 2^n - 2 - k)$ for fixed k and large n. If $k = 0$, we must have all sets in the family of closed sets, $\alpha(n, 2^n - 2) = 1$. Let $k = 1$. Only one set is missing from 2^X. Since the family of closed sets is an intersecting semi-lattice, the missing set must have $n-1$ elements. This missing set can be chosen in n different ways, $\alpha(n, 2^n - 3) = n$.

For general k, let $\mathcal{Z}$ be a family of closed sets which is an intersection semi-lattice consisting of $2^n - k$ members. As the case $k = 1$ indicates, it is better to consider the family of the complements of the complementing family:

$$\mathcal{C} = \{\overline{A} : A \in 2^X - \mathcal{Z}\}$$

It is easy to see that $\mathcal{C}$ possesses the following properties:

$$\emptyset, X \notin \mathcal{C}, \tag{2.6}$$

$$|\mathcal{C}| = k, \tag{2.7}$$

$$A \in \mathcal{C},\ A = B \cup C \text{ implies either } B \in \mathcal{C} \text{ or } C \in \mathcal{C}. \tag{2.8}$$

Now we prove some more properties about its *inner atoms,* that is, the atoms contained in at least one member of $\mathcal{C}$. The atom contained in no member (if it is non-empty) is called the *outer atom.*

Lemma 3. *If $\mathcal{C}$ satisfies (2.6)–(2.8) then the sizes of its inner atoms are* 1 *and their number is at most k.*

Proof. Suppose that one of the inner atoms, say D has at least two elements. There is a set A such that $D \subset A \in \mathcal{C}$. It has a partition $A = B \cup C$ dividing D into two non-empty parts. Then, by (2.8), either B or C is in $\mathcal{C}$, contradicting the assumption that D is an atom. This contradiction proves that the inner atoms are of one element.

We prove the second statement of the lemma by induction on k. It is trivial for $k = 0$, suppose that it is true for $k - 1 \geq 0$ and prove it for k. Suppose that $\mathcal{C}$ satisfies (2.6)–(2.8). Choose a maximal member A of $\mathcal{C}$. Then $\mathcal{C} - \{A\}$ satisfies (2.6)–(2.8) with $k - 1$. By the induction hypothesis, the number of inner atoms of $\mathcal{C} - \{A\}$ is at most $k - 1$. As each of these atoms has only one element, A cannot cut them. The only new inner atom formed by adding A to the family is the atom which is the intersection of A and the complements of all other members of the family $\mathcal{C}$. The number of inner atoms is at most k. ■

Let $\mathbf{J}(n, k)$ denote the set of families $\mathcal{C}$ satisfying (2.6)–(2.8) on n elements. We will classify them according to their contractions. Let $\mathbf{U}$=$\mathbf{U}(n, k)$ be a set of possible contractions, representing each isomorphism class exactly once. Then

$$\mathbf{J}(n, k) = \bigcup_{\mathcal{A} \in \mathbf{U}} \mathbf{D}_n(\mathcal{A}) \tag{2.9}$$

where $\mathbf{D}_n(\mathcal{A})$ denotes the set of families on n elements, weakly isomorphic to $\mathcal{A}$ and having all inner atoms of size 1. (That means, that only the *outer* atom is blown up.)

Lemma 4. *Suppose that $\mathcal{A} \in \mathbf{U}$ is a family on $r + 1$ elements. Then*

$$|\mathbf{D}_n(\mathcal{A})| \sim \frac{n^r}{s(\mathcal{A})}$$

holds (r is fixed, $n \to \infty$).

Proof. Let $\{a_0, a_1, \ldots, a_r\}$ be the groundset of $\mathcal{A}$. Suppose that $\mathcal{C} \in \mathbf{D}_n(\mathcal{A})$ and that its groundset is $X = \{x_1, \ldots, x_n\}$. We may assume $r < n$, thus the members of $\mathbf{D}_n(\mathcal{A})$ have an outer atom, consequently so does $\mathcal{A}$, as well. Suppose that it is a_0. $\mathcal{A}$ and $\mathcal{C}$ are weakly isomorphic therefore there are functions f and g mapping their atoms and members, resp., preserving inclusion. Define the function $h(x)$ $(x \in X)$ by $h(x) = i$ iff $x \in f(a_i)$. In this way we defined a function h for each pair $\mathcal{A}, \mathcal{C}$. $\mathcal{A}$ can be chosen in $s(\mathcal{A})$ ways, $\mathcal{C}$ can be chosen in $|\mathbf{D}_n(\mathcal{A})|$ ways, independently. That is, the number of such functions is $s(\mathcal{A})|\mathbf{D}_n(\mathcal{A})|$. The function h takes on the values $0, 1, \ldots, r$, and there is exactly one $x \in X$ such that $h(x) = i$ for $i = 1, 2, \ldots, r$. On the other hand, it is easy to see that all such functions can be obtained for some pair $\mathcal{A}, \mathcal{C}$. The number of such functions is obviously

$$n(n-1)\ldots(n-r+1) = s(\mathcal{A})|\mathbf{D}_n(\mathcal{A})|.$$

The statement of the lemma easily follows. ∎

Let us introduce the following notation:

$$\theta(k) = \sum_{\mathcal{A} \in \mathbf{T}_k} \frac{1}{s(\mathcal{A})},$$

where $\mathbf{T}_k$ is the set of non-isomorphic atomic families satisfying (2.6)–(2.8) on a $(k+1)$-element set.

Theorem 2. *The number $\alpha(n, 2^n - 2 - k)$ of closures on n elements and having rank $2^n - 2 - k$ satisfies*

$$\alpha(n, 2^n - 2 - k) \sim \theta(k) n^k$$

where k is fixed, $n \to \infty$.

Proof. (2.9) implies

$$\alpha(n, 2^n - 2 - k) = \sum_{\mathcal{A} \in \mathbf{U}} |\mathbf{D}_n(\mathcal{A})|.$$

The number of terms depend only on k, therefore their asymptotic behaviour can be considered separately. The asymptotic behaviour is given in Lemma 4. Only the terms with $r = k$ count, by Lemma 3, the rest is $o(n^k)$. Collecting these terms the statement of the theorem is obtained. ∎

There is an apparent similarity between $\eta(k)$ and $\theta(k)$. The families in $\mathbf{S}_k$ and $\mathbf{T}_k$ become very similar if $\emptyset$ and the groundset are omitted from the families belonging to the latter one: they have k members different from $\emptyset$ and the groundset, they are atomic with $k+1$ atoms. However (2.2) and (2.8) are really different as the following examples show for $k = 3$: $\{\{1,2\}, \{1,3\}, \{2\}, \{3\}\}, \{\{1\}, \{1,2\}, \{1,2,3\}\}$.

Problem 1. *Give good estimates on $\eta(k)$ and $\theta(k)$.*

Theorems 1 and 2 show that the sizes of the lowest levels grow exponentially while starting from the top the sizes grow only polynomially. Although this is more or less true for the first and last cn levels where c is a small positive constant, this observation makes it probable, that the lowest part of the partially ordered set $\mathbf{P}$ is much fatter than the upper part, for a longer interval.

Problem 2. *Determine the level of largest size in* $\mathbf{P}$.

3. The number of closures uniquely determined by their keys

The system $\mathcal{K} = \mathcal{K}(\mathcal{L})$ of minimal keys of a closure $\mathcal{L}$ is determined uniquely by the closure $\mathcal{L}$. It is a non-empty Sperner family. Conversely, for any non-empty Sperner family $\mathcal{S}$ there is a closure $\mathcal{L}$ satisfying $\mathcal{S} = \mathcal{K}(\mathcal{L})$. This is, however, not unique. Denote by $\beta(n), \gamma(n)$ and $\delta(n)$ the number of Sperner families, Sperner families determining the closure uniquely and saturated Sperner families, resp. The inequality

$$\delta(n) \leq \gamma(n) \leq \beta(n) \leq \alpha(n)$$

is obvious. Korshunov [11] gave very good estimates on $\beta(n)$. However, we need here only the simpler estimate of Kleitman [10]:

$$2^{\binom{n}{\lfloor \frac{n}{2} \rfloor}} \leq \beta(n) \leq 2^{\binom{n}{\lfloor \frac{n}{2} \rfloor}(1 + \frac{c \ln n}{\sqrt{n}})}.$$

In this section $\gamma(n)$ is investigated. First we prove that it is not much smaller than $\beta(n)$.

Theorem 3.

$$2^{\binom{n}{\lfloor \frac{n}{2} \rfloor}(1 + o(1))} \leq \gamma(n).$$

In the proof, we use bipartite graphs. Let $G = (A, B; E)$ be a bipartite graph, denote the minimum degree of the vertices in B and the maximum degree of the vertices in A by d_B and D_A, resp. A set $C \subseteq A$ is called a *covering* iff each vertex of B is connected to some vertex in C. Its minimum size is denoted by τ_A. A *fractional covering* is a non-negative real valued function determined on A such that the sum of the values associated with the vertices connected to any given vertex in B is at least 1. The minimum of the total sum of the values of a fractional covering is denoted by τ_A^*. It is easy to see that any covering defines a fractional covering, if 1 and 0 are associated with the elements of C and $\overline{C}$, resp. Hence $\tau_A^* \leq \tau_A$ follows. Lovász [12], Sapozhenko [13] and Stein [14] independently proved that

$$\tau_A \leq (1 + \ln D_A)\tau_A^*. \qquad \textbf{(3.1)}$$

The function associating $\frac{1}{d_B}$ with each vertex is a fractional covering. This implies

$$\tau_A^* \leq \frac{|A|}{d_B}. \qquad (3.2)$$

(3.1) and (3.2) prove the following lemma.

Lemma 5. *Let $G = (A, B; E)$ be a bipartite graph and assume that the degree of any vertex in B is at least d_B (≥ 1). Then there is a subset $C \subseteq A$ such that each vertex of B is connected to some vertex of C and*

$$|C| \leq (1 + \ln D_A)\frac{|A|}{d_B}.$$

Lemma 6. *Let G be a bipartite graph like in Lemma 5. Then there is a subset $C \subset A$ such that each vertex of B is connected to at least two vertices of C and*

$$|C| \leq 2(1 + \ln D_A)\frac{|A|}{d_B - 1}.$$

Proof. Choose C_1 according to Lemma 5. Let B' the set of vertices in B having exactly one neighbour in C_1. Define $A' = A - C$, and let G' be the graph induced by $A' \cup B'$. It is easy to see that the degree of any vertex from B' (in G') is at least $d_B - 1$. Apply Lemma 5 for G'. There exists a C_2 such that each vertex of B' is connected to some vertex of C_2 and

$$|C_2| \leq (1 + \ln D_{A'})\frac{|A'|}{d_B - 1} \leq (1 + \ln D_A)\frac{|A|}{d_B - 1}. \qquad (3.3)$$

It is easy to see that $C = C_1 \cup C_2$ satisfies the requirements of the lemma. The inequality is a consequence of the inequality for $|C_1|$ and (3.3). ■

Lemma 7. *There is a family $\mathcal{F}$ of $\lfloor\frac{n}{2}\rfloor$-element subsets of the n-element X $(n > 1)$ such that*

$$8\frac{\ln n}{n}\binom{n}{\lfloor\frac{n}{2}\rfloor} \geq |\mathcal{F}| \tag{3.4}$$

and

$$\text{for any } \left(\left\lfloor\frac{n}{2}\right\rfloor - 1\right)\text{-element subset } A \text{ of } X \text{ there are two distinct members of } \mathcal{F} \text{ containing } A. \tag{3.5}$$

Proof. We use Lemma 6. Let A and B be the family of all $\lfloor\frac{n}{2}\rfloor$ and $\lfloor\frac{n}{2}\rfloor - 1$-element subsets of X, resp. Then

$$d_B = \lceil\frac{n}{2}\rceil + 1 \text{ and } D_A = \lfloor\frac{n}{2}\rfloor$$

hold. An easy calculation shows the inequality in the lemma. ■

Remark. If $l \leq k \leq n$ are positive integers, then the *Turán number* $T(n,k,l)$ is the minimum number of l-element subsets of an n-element set such that every k-element subset contains at least one of them. There are many papers on these Turán numbers, see e.g. the survey of Brouwer and Voorhoeve [4]. Corollary 2 of [9] gives

$$\frac{1}{\lceil\frac{n}{2}\rceil + 1}\binom{n}{\lfloor\frac{n}{2}\rfloor} \leq T(n, \lceil\frac{n}{2}\rceil + 1, \lceil\frac{n}{2}\rceil).$$

Our Lemma 7 gives an upper estimate which is weaker up to a factor $\ln n$, only. This upper estimate appeared in [3].

Proof of Theorem 3. If $\mathcal{S}$ is a Sperner family on X, define $\mathcal{S}^-$ to be the family of subsets of X containing no member of $\mathcal{S}$ and maximal for this property (that is, A is in $\mathcal{S}^-$ iff there is no member B of $\mathcal{S}$ such that $B \subseteq A$, but for any proper superset C of A one can find such a B). Theorem 4 of [5] states that the non-empty Sperner family $\mathcal{S}$ determines the closure $\mathcal{L}$ (by $\mathcal{S} = \mathcal{K}(\mathcal{L})$) uniquely iff

$$B \subset A \in \mathcal{S}^- \text{ implies that } B \text{ is an intersection of some members of } \mathcal{S}^-. \tag{3.6}$$

It is easy to see that if $\mathcal{S}^-$ is equal to $\mathcal{F}$ of Lemma 7 then $\mathcal{S}$ determines a closure uniquely, by (3.5) and (3.6). Furthermore, $\mathcal{F} \subseteq \mathcal{S}^-$ has the same implication. The number of such families $\mathcal{S}^-$ is at least

$$2^{\binom{n}{\lfloor\frac{n}{2}\rfloor}(1 - 8\frac{\ln n}{n})},$$

by (3.5). However, $\mathcal{S}_1^- \neq \mathcal{S}_2^-$ implies $\mathcal{S}_1 \neq \mathcal{S}_2$. Hence we have the desired number of families $\mathcal{S}$ determining the closure uniquely. ■

Theorem 4.

$$\lim_{n\to\infty} \frac{\gamma(n)}{\beta(n)} = 0.$$

Proof. It is essentially based on the definitions and results of [11]. Denote by $\mathcal{B}_k^n$ the family of all k-element subsets of an n-element set. Let $\mathcal{A} \subseteq \mathcal{B}_k^n$. The vertices of the graph $G(\mathcal{A})$ are the members of $\mathcal{A}$ and two vertices are connected iff their intersection is of size $k-1$. A subfamily of $\mathcal{A}$ corresponding to the vertices of a connected component of $G(\mathcal{A})$ is called a component of $\mathcal{A}$. Let $\mathbf{M}(n)$ denote the set of all Sperner families on n elements. Two cases are distinguished in the proof.

Case 1. $n = 2q$. Put

$$\mathbf{L}(n) = \{\mathcal{A} \in \mathbf{M}(n) : \mathcal{A} \text{ satisfies conditions } (i) - (iii)\}$$

where

(*i*) $\mathcal{A} \subset \mathcal{B}_{q-1}^n \cup \mathcal{B}_q^n \cup \mathcal{B}_{q+1}^n,$

(*ii*) $\mathcal{A} \cap \mathcal{B}_{q-1}^n$ and $\mathcal{A} \cap \mathcal{B}_{q+1}^n$ are non-empty,

(*iii*) both $\mathcal{A} \cap \mathcal{B}_{q-1}^n$ and $\mathcal{A} \cap \mathcal{B}_{q+1}^n$ have only components of cardinality 1 or 2.

It is easy to see that $\mathbf{M}^0(n) \subset \mathbf{L}(n)$, where $\mathbf{M}^0(n)$ is defined on page 83 in [11]. Since $\mathbf{M}^0(n) = \mathbf{M}^4(n)$ (defined on page 70 in [11]), thus Theorem 1, Lemma 14.5, (16.1) and (16.2) prove that $|\mathbf{M}^0(n)|$ and $|\mathbf{M}(n)|$ are asymptotically equal. Hence the same holds for $|\mathbf{L}(n)|$, that is,

$$|\mathbf{L}(n)| \sim |\mathbf{M}(n)|.$$

We will now verify that $\mathcal{S}^- \in \mathbf{L}(n)$ implies that there are $B \subset A \in \mathcal{S}^-$ not satisfying (3.6), supposing $n = 2q > 2$. Choose A from $\mathcal{S}^- \cap \mathcal{B}_{q+1}^n$. Two cases are distinguished: A is a member of a component of size either 1 or 2. In the first case, a q-element subset B cannot be a subset of any other member of $\mathcal{S}^-$, therefore it cannot be an intersection. In the second case, let $A = A_1$ and A_2 be the two elements of the component. If $B \subset A_1$ but $B \neq A_1 \cap A_2$ then B is not contained in any other member of $\mathcal{S}^-$, therefore it cannot be an intersection. Such a B exists since $q + 1 > 1$.

This proves that there are asymptotically as many families $\mathcal{S}^-$ not satisfying (3.6) as the number of Sperner families. However, $\mathcal{S}_1^- \neq \mathcal{S}_2^-$ implies $\mathcal{S}_1 \neq \mathcal{S}_2$, thus the same statement is true for the number of families

$\mathcal{S}$ not satisfying (3.6). That is, by Theorem 4 of [5], almost all Sperner families are such that they do not determine the closure uniquely. The proof in this case is complete.

Case 2. $n = 2q + 1$. The proof in this case is basically the same as in the previous case, but the structure is more tedious. Put

$$\mathbf{L}_1(n) = \mathbf{L}(n)$$

and

$$\mathbf{L}_2(n) = \{\mathcal{A} \in \mathbf{M}(n) : \mathcal{A} \text{ satisfies conditions } (iv) - (vi)\}$$

where

(*iv*) $$\mathcal{A} \subset \mathcal{B}_q^n \cup \mathcal{B}_{q+1}^n \cup \mathcal{B}_{q+2}^n,$$

(*v*) $$\mathcal{A} \cap \mathcal{B}_q^n \text{ and } \mathcal{A} \cap \mathcal{B}_{q+2}^n \text{ are non} - \text{empty},$$

(*vi*) both $\mathcal{A} \cap \mathcal{B}_q^n$ and $\mathcal{A} \cap \mathcal{B}_{q+2}^n$ have only components of cardinality 1 or 2.

Observe that complementation of all members of a family is a bijection (denoted by f) between $\mathbf{L}_1(n)$ and $\mathbf{L}_2(n)$. In [11], a certain set $\mathbf{M}_1^{0,1}(n)$ is introduced (page 96). It is a subset of $\mathbf{L}_1(n)$. Theorem 1 and (17.1) of [11] prove that the size of $\mathbf{M}_1^{0,1}(n)$ is $\sim \frac{1}{2}\beta(n)$. Therefore the size of $\mathbf{L}_1(n)$ is at least this much. Since $f(\mathbf{M}_1^{0,1}(n)) \subset f(\mathbf{L}_1(n)) = \mathbf{L}_2(n)$, the same can be stated about the size of $\mathbf{L}_2(n)$. We obtained

$$|\mathbf{L}_1(n) \cup \mathbf{L}_2(n)| = |\mathbf{L}_1(n)| + |\mathbf{L}_2(n)| \tilde{>} \beta(n). \tag{3.7}$$

However,

$$|\mathbf{L}_1(n) \cup \mathbf{L}_2(n)| \leq |\mathbf{M}(n)| = \beta(n).$$

This and (3.7) imply that almost all Sperner families belong to $\mathbf{L}_1(n) \cup \mathbf{L}_2(n)$. One can see, as in case 1) that $\mathcal{S}^- \in \mathbf{L}_1(n) \cup \mathbf{L}_2(n)$ implies that $\mathcal{S}$ does not determine a closure uniquely. The proof can be finished like in the previous case.

Problem 3. *Give better estimates on $\gamma(n)$ and give non-trivial estimates on $\delta(n)$.*

Acknowledgement. We are indebted to Z. Füredi for calling our attention to some papers in the literature.

REFERENCES

[1] Alekseev, On the number of intersection semilattices, *Diskretnaya Mat.* **1**(1984), 129–136 (in Russian).

[2] W. W. Armstrong, Dependency structures of data base relationships, in: *Information Processing,* **74** North Holland, Amsterdam, 1974, 580–583.

[3] A. Asratjan, N. Kuzjurin and A. A. Sapozhenko, A survey of some results on covering problems, *Metody diskretnovo analiza v reshenii kombinatornych zadach = Methods of Discrete Analysis Solving Combinatorial Problems* Novosibirsk, **30**(1977), 46–75 (in Russian).

[4] A. E. Brouwer and M. Voorhoeve, Turán theory and the lotto problem, in: *Packing and covering in combinatorics,* (ed.: A. Schrijver), Mathematical Centre Tracts **106**, Amsterdam, 1979, 99–105.

[5] G. Burosch, J. Demetrovics and G.O.H. Katona, The poset of closures as a model of changing databases, *Order* **4**(1987) 127–142.

[6] G. Burosch, J. Demetrovics, G.O.H. Katona, D.J. Kleitman and A. A. Sapozhenko, On the number of databases and closure operations, *Theoretical Comput. Sci.* **78**(1991) 377–381.

[7] E.F. Codd, A relational model of data for large shared data banks, *Comm. ACM* **13**(1970) 377–387.

[8] J. Demetrovics and G.O.H. Katona, Combinatorial problems of database models, in: *Algebra, Combinatorics and Logic in Computer Science,* (eds.: J. Demetrovics and G.O.H. Katona), Coll. Math. Soc. J. Bolyai, **42**, North-Holland, Amsterdam, 1983, 331–353.

[9] Gy. Katona, T. Nemetz and M. Simonovits, On a graph-problem of Turán, *Mat. Lapok* **15**(1964), 228–238 (in Hungarian).

[10] D. J. Kleitman, On Dedekind's problem: the number of monotone Boolean functions, *Proc. Amer. Math. Soc.* **21**(1969), 677–682.

[11] A.D. Korshunov, On the number of monotone Boolean functions, *Problemy Kibernet.* **38**(1981), 5–108 (in Russian).

[12] L. Lovász, On the ratio of optimal integral and fractional covers, *Discrete Math.* **13**(1975), 383–390.

[13] A. A. Sapozhenko, On the complexity of disjunctive normal forms obtained by greedy algorithm, *Discret. Analiz* **21**(1972), 62–71 (in Russian).

[14] S.K. Stein, Two combinatorial covering theorems, *J. Combinatorial Theory (A)* **16**(1974), 391–397.

Gustav Burosch
O-2500 Rostock,
Goethestr. 21,
Germany

János Demetrovics
Comp. and Automation Inst.,
Hungarian Academy of Sciences,
1111 Budapest, Kende u. 13–17.,
Hungary

e-mail: h935dem@huella.bitnet

Gyula Katona
Mathematical Institute,
Hungarian Academy of Sciences,
Budapest, Pf. 127, 1364 Hungary
`h1164kat@huella.bitnet`

Daniel J. Kleitman
Massachusetts Inst. of Technology,
Cambridge, Mass. 02139,
USA
`e-mail: djk@math.mit.edu`

Alexander A. Sapozhenko
Department of Computer Science,
Moscow State University,
Moscow, Russia