New Type of Coding problem Motivated by Database Theory[*][†]

Gyula O.H. Katona Attila Sali
Alfréd Rényi Institute of Mathematics, HAS
Budapest P.O.B. 127 H-1364 HUNGARY
ohkatona@rnyi.hu, sali@renyi.hu

Abstract

The present paper is intended to survey the interaction between Relational Database Theory and Coding Theory. In particular it is shown how an extremal problem for relational databases gives rise to a new type of coding problem. The former concerns minimal representation of branching dependencies that can be considered as a data mining type question. The extremal configurations involve *d-distance sets* in the space of disjoint pairs of k-element subsets of an n-element set X. Let X be an n–element finite set, $0 < k < n/2$ an integer. Suppose that $\{A_1, B_1\}$ and $\{A_2, B_2\}$ are pairs of disjoint k-element subsets of X (that is, $|A_1| = |B_1| = |A_2| = |B_2| = k, A_1 \cap B_1 = \emptyset, A_2 \cap B_2 = \emptyset$). Define the distance of these pairs by $d(\{A_1, B_1\}, \{A_2, B_2\}) = \min\{|A_1 - A_2| + |B_1 - B_2|, |A_1 - B_2| + |B_1 - A_2|\}$.

1 Introduction

A relational database system of the scheme $R(A_1, A_2, \ldots, A_n)$ will be considered as a matrix, where the columns correspond to the *attributes* A_i's (for example name, date of birth, place of birth etc.), while the rows are the n-tuples of the relation r. That is, a row contains the data of a given *individual*. For the sake of convenience, it is assumed that the rows of the matrix are pairwise distinct. Let Ω denote the set of attributes (the set of

[*]AMS Subject classification Primary 68P15 68R05 Secondary 05D05. Keywords: Relational database, functional and branching dependencies, code, d-distance set.

[†]The work was supported by the Hungarian National Foundation for Scientific Research grant numbers T029255, T034702 and by the Research Grant of the Hungarian Academy of Sciences.

the columns of the matrix). Let $A \subseteq \Omega$ and $b \in \Omega$. We say that b *(functionally) depends* on A (see [1, 7]) if the data in the columns of A determine the data of b, that is there exist no two rows which agree in A but differ in b. We denote this by $A \longrightarrow b$. A set function on the subsets of Ω can be defined with the help of functional dependency.

Definition 1.1 *Let M be the matrix of a relational database. The function $\mathcal{C}_M\colon 2^\Omega \longrightarrow 2^\Omega$ is defined by*

$$\mathcal{C}_M(A) = \{b\colon b \in \Omega,\ A \longrightarrow b\}$$

for any $A \subseteq \Omega$. We shall write $\mathcal{C}$ instead of $\mathcal{C}_M$ if it does not cause confusion.

The function defined above has the following three properties.

Proposition 1.2

$$\begin{array}{ll} 1) & A \subseteq \mathcal{C}(A), \\ 2) & A \subseteq B \implies \mathcal{C}(A) \subseteq \mathcal{C}(B), \\ 3) & \mathcal{C}(\mathcal{C}(A)) = \mathcal{C}(A). \end{array}$$

Set functions satisfying properties 1)-3) are called *closure operations*. Armstrong proved that the above correspondence could be reversed.

Theorem 1.3 ([1]) *For any given closure $\mathcal{C}$ there exists a matrix M such that*

$$\mathcal{C}_M = \mathcal{C}.$$

It is evident that a matrix with a small number of rows cannot yield a complicated closure. Furthermore, as closures and database matrices are equivalent by Armstrong's theorem, the following number is a measure of complexity of closures.

Definition 1.4 *Let $\mathcal{C}$ be a closure on Ω. Then let*

$$s(\mathcal{C}) = \min_{M:\mathcal{C}_M=\mathcal{C}} \{\textit{number of rows in } M\}.$$

The data mining aspect of $s(\mathcal{C})$ is that if we want to explore dependencies of a database, then the number of tuples (records) can rule out those dependency systems whose minimum representation requires a larger number of rows. It is very hard to determine $s(\mathcal{C})$ for an arbitrary closure $\mathcal{C}$. However, there are nice combinatorial results for certain closures.

Definition 1.5 *Let $\mathcal{C}_n^k$ denote the following closure on Ω:*

$$\mathcal{C}_n^k(X) = \begin{cases} X & \text{if } |X| < k \\ \Omega & \text{otherwise.} \end{cases}$$

The following lemma gives a general lower bound for $s(\mathcal{C}_n^k)$.

Lemma 1.6 ([10])

$$\binom{s(\mathcal{C}_n^k)}{2} \geq \binom{n}{k-1}.$$

The exact value of $s(\mathcal{C}_n^k)$ is determined for certain values of k.

Theorem 1.7 ([10]) *The following equalities hold:*

$$\begin{array}{ll} a)\ s(\mathcal{C}_n^1) = 2, & c)\ s(\mathcal{C}_n^{n-1}) = n, \\ b)\ s(\mathcal{C}_n^2) = \left\lceil \frac{1+\sqrt{1+8n}}{2} \right\rceil, & d)\ s(\mathcal{C}_n^n) = n+1. \end{array}$$

We give the proof of Case b) as an example.

Proof of Case b) of Theorem 1.7
Let $s = s(\mathcal{C}_n^2)$. Lemma 1.6 gives $\binom{s}{2} \geq n$. Note that the number of the right hand side of equality in Case b) is the smallest s satisfying the previous inequality. If s is such, then we construct a matrix M with s rows such that $\mathcal{C}_M = \mathcal{C}_n^2$ as follows:

$$M = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 1 & 1 & 1 & \dots & 1 \\ 0 & 2 & 2 & \dots & 2 & 0 & 0 & 0 & \dots & 2 \\ 3 & 0 & 3 & \dots & 3 & 0 & 3 & 3 & \dots & 3 \\ 4 & 4 & 0 & \dots & 4 & 4 & 0 & 4 & \dots & 4 \\ 5 & 5 & 5 & \dots & 5 & 5 & 5 & 0 & \dots & 5 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ s & s & s & \dots & 0 & s & s & s & \dots & s \end{pmatrix}.$$

There is a pair of zeros in every column of M such that for different columns the zeros are in different pairs of rows, which implies that every one-element subset of Ω is closed. This can be done by the choice of s. On the other hand, no two rows agree in more than one column, so if $A \subseteq \Omega$ with $|A| > 1$, then $\mathcal{C}_M(A) = \Omega$. ∎

The exact value of $s(\mathcal{C}_n^k)$ is not known for $k > 3$. However, if k is fixed, then its asymptotic behavior is known.

Theorem 1.8 ([8]) *If k is fixed and $n > n_0(k)$, then*

$$c_1(k)n^{\frac{k-1}{2}} \leq s(\mathcal{C}_n^k) \leq c_2(k)n^{\frac{k-1}{2}}.$$

The lower bound in Theorem 1.8 follows from Lemma 1.6. The upper bound is proven by a construction involving polynomials over a finite field. Füredi proved some bounds for the "other end" of the range of k.

Theorem 1.9 ([17]) *If k is fixed and $n > n_0(k)$, then*

$$c_3(k)n^{\frac{2k+1}{3}} \leq s(\mathcal{C}_n^{n-k}) \leq c_4(k)n^k.$$

As an example of interaction between database type problems and other fields of combinatorics, let us now consider the case $k = 3$. From Lemma 1.6 we obtain that

$$\binom{s(\mathcal{C}_n^3)}{2} \geq \binom{n}{2},$$

hence $s = s(\mathcal{C}_n^3) \geq n$. Equality holds if we can construct an $n \times n$ matrix M such that:

1) for any distinct $a, b, c \in \Omega$ there are two rows equal in columns a and b, but different in c.
2) for any distinct $a, b, c \in \Omega$ there are no two rows equal in all of them.

Consider the dual problem. A column naturally determines a partition of the set Y of rows, by the equalities of its entries. We say that a partition *covers* the pair (α, β) $(\alpha, \beta \in Y,\ \alpha \neq \beta)$ iff α and β are in the same class of the partition. We can state the previous two properties as follows.

Find n partitions of Y ($|Y| = n$) such that:
1') for any two partitions there exists a pair (α, β) covered by both,
2') no pair (α, β) is covered by three different partitions.

However, the number of pairs of partitions is also $\binom{n}{2}$ and different pairs of partitions cannot cover the same pair of elements by 2'). Thus, we may conclude that 1') and 2') (consequently 1) and 2)) are equivalent to:

(i) for any two partitions there is exactly one pair of elements, which is covered by both,
(ii) each pair of elements is covered by exactly two different partitions.

Definition 1.10 *A collection of partitions satisfying (i) and (ii) is called an* orthogonal double cover*.*

The existence of orthogonal double covers in various forms were investigated and a new branch of design theory grew out of this database motivation, see [2, 5, 6, 18, 19, 20, 21, 22, 26], among others.

2 Branching Dependencies

Functional dependencies have turned out to be very useful. Many existing data base managing systems are based on this concept. Let us consider the following example. Suppose that $\Omega = \{A_1, A_2, A_3, A_4\}$ and $A_1 \longrightarrow A_2$ and $A_3 \longrightarrow A_4$ hold. If we store the whole matrix in the memory of a computer, then it requires $4N_1N_3$ registers in the worst case, where N_1 (N_3) denotes the number of possible different values of A_1 (A_3). Indeed, A_1 and A_3 can take values independently, but they determine A_2 and A_4, respectively. Thus, the number different rows is at most N_1N_3. However, using the given functional dependencies, we can save a lot of memory. Indeed, it is enough to store the matrix consisting of the columns A_1 and A_3 ($2N_1N_3$ registers) together with two little matrices each having two columns. One contains values of A_1 and A_2 in the first and second columns, respectively. The first column contains all possible values of A_1, while the second one contains the values determined by the dependency $A_1 \longrightarrow A_2$. The other small matrix is built up from A_3 and A_4 in the same way. The number of stored values is at most $2N_1N_3 + 2(N_2 + N_4)$, which is usually significantly smaller than $4N_1N_3$.

In [12] a more general (weaker) dependency was introduced. We describe it first in a very particular case, then we show the usefulness of the concept. Let $A \subseteq \Omega$ and $b \in \Omega$, we say that b $(1,2)$-depends on A if the values in A determine the values in b in a "two-valued" way. That is, there exist no three rows same in A but having three different values in b. We denote it by $A \stackrel{(1,2)}{\longrightarrow} b$. Similarly, $A \stackrel{(1,q)}{\longrightarrow} b$ if there exist no $q+1$ rows each having the same values in columns of A, but containing $q+1$ different values in the column b.

Let us suppose that the database consists of the trips of an international transport truck, more precisely, the names of the countries the truck enters. For the sake of simplicity, let us suppose, that the truck goes through exactly four countries in each trip, (counting the start and endpoints, too) and does not enter a country twice during one trip. Suppose furthermore, that there are 30 possible countries and one country has at most five neighbors. Let A_1, A_2, A_3, A_4 denote the first, second, third and fourth country as attributes. It is easy to see that $A_1 \stackrel{(1,5)}{\longrightarrow} A_2$, $\{A_1, A_2\} \stackrel{(1,4)}{\longrightarrow} A_3$ and $\{A_2, A_3\} \stackrel{(1,4)}{\longrightarrow} A_4$. Now, we cannot decrease the size of the stored matrix, as in the case of functional ((1,1)-) dependency, but we can decrease the range of the elements of the matrix. The range of each element of the original matrix consists of 30 values, names of countries or some codes of them (5 bits each, at least). Let us store a little table ($30 \times 5 \times 5 = 750$ bits) that contains a numbering of the neighbors of each country, which assigns to them the numbers 0,1,2,3,4 in some order. Now we can replace attribute A_2 by these numbers (A_2^*), because the value of A_1 gives the starting country and the value of A_2^* determines the second country with the help of the little table. The same holds for the attribute A_3, but we can decrease the number of possible values even further, if we give a table of numbering the possible third countries for each A_1, A_2 pair. In this case, the attribute A_3^* can take only 4 different values. The same holds for A_4, too. That is, while each element of the original matrix could be encoded by 5 bits, now for the cost of two little auxiliary tables we could decrease the length of the elements in the second column to 3 bits, and that of the elements in the third and fourth columns to 2 bits.

It is easy to see, that the same idea can be applied in each case when we store the paths of a graph, whose maximal degree is much less than the number of its vertices or when we want to store the sequence of states of a process, where the number of all possible states is much larger, than the number of possible successor states of a state or in any case when there hold many $(1, q)$-dependencies, where q is small.

The general concept that was studied in [12, 13, 14] is the (p, q)-dependency ($1 \leq p \leq q$ integers)

Definition 2.1 *Let M be an $m \times n$ matrix, with column set Ω. Let $A \subseteq \Omega$ and $b \in \Omega$. We say that b (p, q)-*depends *on A if there are no $q + 1$ rows of M such that they contain at most p different values in each column of A, but $q + 1$ different values in b.*

The functional dependency discussed in the previous section is a special case, namely it is the $(1, 1)$-dependency. For a given matrix M we define a function from the family of subsets of Ω into itself as follows.

Definition 2.2 *Let M be the given matrix. Let us suppose, that $1 \leq p \leq q$. Then the mapping $\boldsymbol{J}_{Mpq}: 2^\Omega \to 2^\Omega$ is defined by*

$$\boldsymbol{J}_{Mpq}(A) = \left\{ b \colon A \xrightarrow{(p,q)} b \right\}.$$

$\boldsymbol{J}_{Mpq}$ in general is not a closure. Its properties are studied in [12] and [27]. However $s_{pq}(\mathcal{C}_n^k)$ can be defined as the minimum number of rows of a matrix in which $\boldsymbol{J}_{Mpq}$ is exactly $\mathcal{C}_n^k$. The exact value of $s_{pq}(\mathcal{C}_n^k)$ is known in a few cases only (see [13]).

The following theorem is interesting for its proof.

Theorem 2.3

$$s_{12}(\mathcal{C}_n^2) = \min\left\{ s \textit{ integer: } \binom{s}{3} \geq 2\,n \right\},$$

provided $n > 452$.

We prove the upper bound in Theorem 2.3 via construction. In fact, we consider the number of rows m to be given, and construct $n = \lfloor \binom{m}{3}/2 \rfloor$ columns so that the $(1,2)$-dependency in that matrix will be exactly $\mathcal{C}_n^2$. The construction is based on the following theorem, which leads to coding theory type generalizations.

Theorem 2.4 *Let* $|X| = n$ *and* $2 \leq k$. *The family of all* k*-subsets of* X *can be partitioned into* $\lfloor \binom{n}{k}/2 \rfloor$ *unordered pairs, so that paired* k*-subsets are disjoint and if* A_1, B_1 *and* A_2, B_2 *are two such pairs with* $|A_1 \cap A_2| \geq \lceil \frac{k+1}{2} \rceil$, *then* $|B_1 \cap B_2| < \lceil \frac{k+1}{2} \rceil$, *provided* $n > n_0(k)$.

Let us suppose, that m is an integer that satisfies $\binom{m}{3} \geq 2\,n$. A matrix with m rows and n columns will be constructed that $(1,2)$-represents $\mathcal{C}_n^2$. Let us denote the set of rows by X. Apply Theorem 2.4 with $q = 3$ and $k = 2$ to obtain disjoint pairs of 3-subsets of X. There are $\lfloor \binom{m}{3}/2 \rfloor$, that is, at least n such pairs. Choose n of them. We construct a column from such a pair, as follows. Put 1's in the rows indexed by the first 3-set, 2's in the rows indexed by the second one, and all different entries, that are at least 3, in the other positions.

If a and b are two distinct columns, then there are no 3 rows that agree in both a and b, because we used all distinct 3-subsets of rows, hence $\{a, b\} \stackrel{(1,2)}{\longrightarrow} \Omega$. On the other hand, if a is constructed from the pair of 3-subsets A_1, A_2 and b is constructed from B_1, B_2, then either $|A_1 \cap B_1| < 2$ or $|A_2 \cap B_2| < 2$, so there are 3 rows which contain all identical entries in column a, but all distinct ones in column b, hence $a \stackrel{(1,2)}{\not\longrightarrow} b$. ∎

Theorem 2.4 is proved using the following Hamiltonian type theorem.

Theorem 2.5 *Let* $G_0 = (V, E_0)$ *and* $G_1 = (V, E_1)$ *be simple graphs on the same vertex set* $|V| = N$, *such that* $E_0 \cap E_1 = \emptyset$. *The 4-tuple* (x, y, z, v) *is called an* alternating cycle *if* (x, y) *and* (z, v) *are in* E_0 *and* (y, z) *and* (x, v) *are in* E_1. *Let* r *be the minimum degree of* G_0 *and let* s *be the maximum degree of* G_1. *Suppose, that*

$$2\,r - 8\,s^2 - s - 1 > N,$$

then there is a Hamiltonian cycle in G_0 such that if (a,b) and (c,d) are both edges of the cycle, then (a,b,c,d) is not an alternating cycle.

The pairs of disjoint k-subsets are obtained from neighboring vertices of a Hamiltonian cycle of type above. G_0 and G_1 are as follows. The vertex set V consists of the k-subsets of X, $|V| = \binom{n}{k} = N$. Two k-subsets are adjacent in G_0 if their intersection is empty, while two k-subsets are adjacent in G_1 if they intersect in at least $\lceil \frac{k+1}{2} \rceil$ elements.

3 Coding type questions

Enomoto and Katona [16] realized that Theorem 2.4 really speaks about a certain kind of distance. Define the *distance* of the pairs $\{A_1, B_1\}$ and $\{A_2, B_2\}$ by

$$d(\{A_1, B_1\}, \{A_2, B_2\}) = \min\{|A_1 - A_2| + |B_1 - B_2|, |A_1 - B_2| + |B_1 - A_2|\}.$$

Theorem 3.1 *Let $|X| = n$. The family of all k-element subsets of X can be partitioned into disjoint pairs (except possibly one if $\binom{n}{k}$ is odd), so that $d(\{A_1, B_1\}, \{A_2, B_2\}) \geq k$ holds for any two such pairs $\{A_1, B_1\}$ and $\{A_2, B_2\}$, provided $n > n_0(k)$.*

It is obvious that $|A_1 \cap A_2| \geq \lceil \frac{k+1}{2} \rceil$ and $|B_1 \cap B_2| \geq \lceil \frac{k+1}{2} \rceil$ imply $d((A_1, B_1), (A_2, B_2)) \leq 2\lfloor \frac{k-1}{2} \rfloor \leq k-1$ for sets satisfying $A_1 \cap B_1 = A_2 \cap B_2 = \emptyset$, therefore the following theorem is really a sharpening of Theorem 2.4 .

The proof of Theorem 3.1 follows the line of that of Theorem 2.4, the difference is that a strengthening of Theorem 2.5 is needed, which involves weighted Hamiltonian cycles.

One can show ([16]) that d is really a distance in the "space" of all disjoint pairs of k-element subsets of X. Theorem 3.1 answers really a coding type question, how many elements can be chosen from this space with large pairwise distances. We say that a set $\mathcal{C}$ of unordered disjoint pairs of k-subsets of an n-set is an (n, k, d)–code if the distance of any two elements is at least d.

Let $C(n, k, d)$ be the maximum size of an (n, k, d)-code. $C'(n, k, d)$ denotes the same under the additional condition that a

k–element subset may occur only once in the pairs $\{A, B\} \in \mathcal{C}$ as A or B.

With this notation Theorem 3.1 states that $C'(n, k, k) = \left\lfloor \frac{1}{2}\binom{n}{k} \right\rfloor$. Observe that Theorem 3.1 implies $C'(n, k, d) = \left\lfloor \frac{1}{2}\binom{n}{k} \right\rfloor$ for all $d \leq k$. On the other hand, $C'(n,k,d) = C(n,k,d)$ holds when $k < d$, since large distances do not allow the repetition of a k-element set. Therefore the remaining unsolved cases of $C'(n,k,d)$ belong to $C(n,k,d)$.

Brightwell and Katona [4] gave a general upper and lower bound for the numbers $C(n,k,d)$.

Theorem 3.2 *Let $d \le 2k \le n$ be integers. Then*

$$C(n,k,d) \le \frac{1}{2} \frac{n(n-1)\cdots(n-2k+d)}{k(k-1)\cdots\lceil\frac{d+1}{2}\rceil \cdot k(k-1)\cdots\lfloor\frac{d+1}{2}\rfloor}$$

holds.

It is not too hard to check that Theorem 3.2 implies that if $2 \le k \le n/2$ then $C(n,k,k+1) < \lfloor\frac{1}{2}\binom{n}{k}\rfloor$, that is, Theorem 3.1 cannot be stated with $k+1$ instead of k. It was conjectured that the upper bound in Theorem 3.2 is asymptotically correct. This conjecture has recently been settled in affirmative by Bollobás, Katona and Leader in [3].

Theorem 3.3 ([3])

$$\lim_{n\to\infty} \frac{C(n,k,d)}{n^{2k-d+1}} = \frac{1}{2} \frac{1}{k(k-1)\cdots\lceil\frac{d+1}{2}\rceil \cdot k(k-1)\cdots\lfloor\frac{d+1}{2}\rfloor}.$$

However [4] also conjectured that there are infinitely many n for any given k and d with equality in the upper estimate of Theorem 3.2. Bollobás et.al. devised a method to show this for the case $C(n,k,2k-1)$. It is based on the following concept. The distance $\delta(a,b)$ of two integers mod m $(1 \le a,b \le m)$ is defined by

$$\delta(a,b) = \min\{|b-a|, |b-a+m|\}.$$

(Imagine that the integers $1,2,\ldots,m$ are listed around the circle clockwise uniformly. Then $\delta(a,b)$ is the smaller distance around the circle from a to b.) $\delta(a,b) \le \frac{m}{2}$ is trivial. Observe that $b-a \equiv d-c \bmod m$ implies $\delta(a,b) = \delta(c,d)$.

We say that the pair $A = \{a_1,\ldots,a_k\}, B = \{b_1,\ldots,b_k\} \subset \{1,\ldots,m\}$ of disjoint sets is *antagonistic* mod m if

(i) all the $k(k-1)$ integers $\delta(a_i,a_j)(i \ne j)$ and $\delta(b_i,b_j)(i \ne j)$ are different,

(ii) the k^2 integers $\delta(a_i,b_j)(1 \le i,j \le k)$ are all different and

(iii) $\delta(a_i,b_j) \ne \frac{m}{2}(1 \le i,j \le k)$.

If there is a pair of disjoint antagonistic k-element subsets mod m then $2k^2+1 \le m$ must hold by (ii) and (iii).

Open Problem 3.4 *Is there a pair of disjoint, antagonistic k-element sets* mod $2k^2+1$*?*

We have an affirmative answer only in three cases.

Proposition 3.5 *There is a pair of disjoint, antagonistic k-element sets* mod $2k^2+1$ *when $k=1,2,3$.*

Using antagonistic pairs of disjoint k-subsets, one can state bounds for $C(n,k,2k-1)$.

Lemma 3.6 *If there is a pair of disjoint, antagonistic k-element sets* mod m*, then $C(m,k,2k-1)\geq m$.*

Proposition 3.7 *Suppose that there is Steiner family $\mathcal{S}(n,2k^2+1,2)$ and a disjoint, antagonistic pair of k-element subsets* mod $2k^2+1$ *then*

$$C(n,k,2k-1)=\frac{n(n-1)}{2k^2}.$$

Now applying Proposition 3.5, the following theorem was obtained in [3].

Theorem 3.8 *If $n\equiv 1,9$* mod 72 *then $C(n,2,3)=\frac{n(n-1)}{8}$. If $n\equiv 1,19$* mod 342 *then $C(n,3,5)=\frac{n(n-1)}{18}$.*

There are some results when n is relatively small. [4] gives good lower estimates for $C(2k,k,d)$ some cases. The method of the construction is a modification of the method used by Sloane and Graham [23] proving lower bounds for constant weight codes.

References

[1] W.W. Armstrong, Dependency Structures of database Relationships, *Information Processing 74* (North Holland, Amsterdam, 1974) 580-583.

[2] F.E. Bennett and Lisheng Wu, On minimum matrix representation of closure operations, *Discrete Appl. Math.* **26** (1990) 25-40.

[3] B. Bollobás, G.O.H. Katona and I. Leader, *A coding problem for pairs of subsets*, in preparation.

[4] G. Brightwell and G.O.H. Katona A new type of coding theorem, *Studia Sci. Math. Hungar.* **38** (2001) 139-147.

[5] Yeow Meng Chee, Design-theoretic problems in perfectly $(n-3)$-error-correcting databases, preprint.

[6] M.S. Chung and D.B.West, The p-intersection number of a complete bipartite graph and orthogonal double coverings of a clique, *Combinatorica* **14** (1994), 453-461.

[7] E.F. Codd, A Relational Model of Data for Large Shared Data Banks, *Comm. ACM,* **13** (1970) 377-387.

[8] J. Demetrovics, Z. Füredi and G.O.H. Katona, Minimum matrix representation of closure operations, *Discrete Appl. Math.*, **11** (1985), 115-128.

[9] J. Demetrovics and Gy. Gyepesi, A note on minimum matrix representation of closure operetions, *Combinatorica* **3** (1983) 177-180.

[10] J. Demetrovics, G.O.H. Katona, Extremal combinatorial problems in a relational database, in: *Fundamentals of Computation Theory 81, Proc. 1981 Int. FCT-Conf.*, Szeged, Hungary, 1981, Lecture Notes in Computer Science 117 (Springer, Berlin 1981) pp. 110-119.

[11] J. Demetrovics, G.O.H. Katona, A survey of some combinatorial results concerning functional dependencies in database relations, *Annals of Math. and Artificial Intelligence* **7** (1993) 63-82.

[12] J. Demetrovics, G.O.H. Katona and A.Sali, The characterization of branching dependencies, *Discrete Appl. Math.*, **40** (1992), 139-153.

[13] J. Demetrovics, G.O.H. Katona and A. Sali, Minimal Representations of Branching Dependencies, *Acta Sci. Math. (Szeged)*, **60**, (1995) 213-223.

[14] J. Demetrovics, G.O.H. Katona and A.Sali, Design type problems motivated by database theory *J. Statist. Planning and Inference* **72** (1998) 149-164.

[15] G.A. Dirac, Some theorems on abstract graphs, *Proc. London Math. Soc.*, Ser. 3, **2** (1952), 69-81.

[16] H. Enomoto and G.O.H. Katona, Pairs of disjoint q-element subsets far from each other, *Electronic Journal of Combinatorics* (2001) **8**(No. 2)# R7.

[17] Z. Füredi, Perfect error-correcting databases, *Discrete Appl. Math.* **28** (1990) 171-176.

[18] H.-D.O.F. Gronau and B. Ganter, On two conjectures of Demetrovics, Füredi and Katona concerning partitions, *Discrete Math.* **88** (1991) 149-155.

[19] B. Ganter, H.-D.O.F. Gronau and R.C. Mullin, On orthogonal double covers of K_n, *Ars Combinatoria* **37** (1994), 209-221.

[20] H.-D.O.F. Gronau, R.C. Mullin and P.J. Schellenberg, On orthogonal double covers of K_n and a conjecture of Chung and West, *J. of Combinatorial Designs* **3** (1995), 213-231.

[21] H.-D.O.F. Gronau, R.C. Mullin and P.J. Schellenberg, On orthogonal double covers of $\vec{K_n}$, preprint.

[22] H.-D.O.F. Gronau, R.C. Mullin and A. Rosa, preprint.

[23] R. Graham and N. Sloane, Lower bounds for constant weight codes, *IEEE Trans. on Information Theory* (1980) **26** 37-43.

[24] U. Leck and V. Leck, There is no ODC with all pages isomorphic to $C_4 \cup C_3 \cup C_3 \cup v$, preprint.

[25] L. Lovász, Topological and algebraic methods in graph theory, in: *Graph Theory and Related Topics*, Proc. Conf. Univ. Waterloo, Ontario 1977, 1-14, Academic Press, NY 1979.

[26] A. Rausche, On the existence of special block designs, *Rostock Math. Kolloq.* **35** (1987) 13-20.

[27] A. Sali and A. Sali Sr., Generalized Dependencies in Relational Databases, *Acta Cybernetica* **13** (1998) 431-438.