From database theory to secret sharing

by

Péter M. Gergely

Submitted to

Central European University, Budapest

Department of Mathematics and its Applications

In partial fulfillment of the requirements for the degree of Doctor of Philosophy in Mathematics and its Applications

Supervisor: Gyula O. H. Katona

Budapest, Hungary

2011

Abstract

A database can be considered as a matrix where a sufficiently large set of columns (a *key*) will uniquely determine the rows. Given the number of columns and the family of minimal keys an interesting task is to find the minimal number of rows with which a matrix can be constructed that fulfills these requirements.

Secret sharing is based on a similar situation. There is a treasure box with many keyholes with each participant holding a key. The box can only be opened if at least a given number of key holders pool their resources together. This can also be formulated by a matrix where the participants are represented by the columns and the treasures are the indices of the rows. Here one can also ask what the minimum number of rows is with which this situation can be reached.

In both problems insufficient resources should not determine the row/secret. However, in real life applications we need stronger security. We cannot allow the possibilities to be narrowed down to the point where brute force is enough to find the correct one. Also we may require that the keys of the participants must be changed from time to time, in other words there must be many sets of partial secrets which properly function.

The main task of the present thesis is to investigate mathematical problems lying in-between. There is a very wide class of problems which are special cases of the general problem obtained as a common generalization of problems in database theory and secret sharing.

Contents

1 Introduction

A database can be considered as an $n \times m$ matrix M. The columns represent kinds of data, first name, last name, date of birth, etc., the entries in one row are the data of a given individual. (Suppose that the rows are different.) Let Ω denote the set of columns where $|\Omega| = m$. We say that a set $K \subset \Omega$ of columns is a *key*, if the data in the columns belonging to K uniquely determine the row. In other words, there are no two rows which are equal in the columns belonging to K but different in some other column. K is a *minimal key* if it is a key, but none of its proper subsets is a key. Of course the family of minimal keys can be quite complicated, but here it will be supposed that they all have the same size. Let $1 \leq \ell$ be an integer and suppose that the family of minimal keys is equal to $\binom{\Omega}{\ell}$ (= family of all ℓ-element subsets). In other words, every ℓ-element set of columns is a key, but no $\ell - 1$-element set of columns is a key. Choosing ℓ columns and fixing the entries in these columns they determine at most one such row, but choosing $\ell - 1$ columns there are at least two different rows which are equal in these columns. For instance, if $\ell = 1$ then this definition means that the entries in any one column are all different, but the empty set is not a key, in other words, the number of rows is at least two.

Let $s(m, \ell)$ denote the minimum number of rows n in an $n \times m$ matrix M in which all ℓ-element sets of columns are the minimal keys. This question

was asked in [4], where also the lower bound

$$\binom{s(m,\ell)}{2} \geq \binom{m}{\ell-1} \tag{1}$$

was found. It was shown that the bounds obtained from (1) for $\ell = 1$ and 2 are sharp. For $\ell = 3$ (1) gives $s(m,3) \geq m$. It was conjectured in [4] that $s(m,3) = m$ and a hint was given for its solution. The conjecture was partially solved in [3] and it was finally proved in [1] for all $m \geq 7$. There are many interesting related results surveyed in [5]. (Papers which appeared later: [8], [12], [14], [15], [18], [19], [20].)

For general ℓ (1) leads to $s(m,\ell) \geq cm^{\frac{\ell-1}{2}}$. Paper [3] proved that this lower bound is asymptotically sharp, that is $c'm^{\frac{\ell-1}{2}}$ is an upper bound with some $c = c(\ell) < c' = c'(\ell)$.

An important branch of *cryptology*, the theory of *secret sharing* is based on a very similar situation. There is a treasure box with m keyholes. Each of the m persons possesses a hard key. Any ℓ of them can open the box, but $\ell - 1$ of them cannot. In other words each of them has a partial secret, his own key. ℓ of these partial secrets are sufficient to determine the main secret, the information needed to open the treasure box, but no $\ell - 1$ of these partial secrets determine the main secret. This can also be formulated by an $n \times m$ matrix M where the persons are represented by the columns and the ith main secret is the index of the ith row. The partial secret p of the person is an entry in his column. In other words he knows that the secret/row can only be one of the rows where p is the entry in his column in

the matrix. Suppose, again that the rows of the matrix are different. Then the assumption is that fixing the entries of the matrix in any ℓ columns there is exactly one row containing these prescribed entries in the given columns, while if only $\ell - 1$ of these entries are fixed in $\ell - 1$ columns, then one can find two distinct rows in which the entries in the given columns are equal. Here one can also ask what the minimum number of rows (main secrets) is with which this situation can be reached. Of course having all ℓ-element subsets as minimal keys is a very specific case, in general we could determine any family of qualified subsets to serve as the *access structure* of our secret sharing scheme.

The two problems look mathematically identical until this point. In the case of the second problem, however, there are stronger assumptions concerning the security. When the $\ell - 1$ people with the $\ell - 1$ (hard)keys cannot open the box, they cannot do much more. However, in the case of practical (electronic) situations it is supposed that the "mechanical structure" of the treasure box is also known by everybody. Mathematically, the participants all know the matrix M. If the partial secrets (entries) of the $\ell - 1$ persons determine two rows then they know that one of the two main secrets open the treasure box. They simply try both! Therefore one supposes here that there are many rows which are equal in the columns belonging to a given set of $\ell - 1$ columns.

But there is another difference between the assumptions in database theory and secret sharing. In the case of databases an $\ell - 1$-element set of

columns is not a key, if there are different rows for one given set of prescribed entries in these columns. In terms of treasure box and of hard keys it means that no $\ell - 1$ (hard) keys can open the box. But strong security requires that the keys of the persons must be changed from time to time, in other words there must be many sets of partial secrets which properly function. In the extreme case all choices function, that is choosing the partial secrets (entries in the corresponding columns) in any way and taking $\ell - 1$ columns there are many distinct rows which have the prescribed entries in these $\ell - 1$ rows. This problem was solved in the paper of Shamir ([21]). (More precisely, Shamir's problem was formulated in a probabilistic way: $\ell - 1$ persons do not receive any information on the main secret what ever choice of partial secrets are given.)

Let us summarize the differences between the two problems in the matrix form. There is no difference concerning the choice of ℓ columns. Prescribing the entries in these ℓ columns in any way there is at most one row containing these entries in the given columns. But the two problems are very different concerning the choice of $\ell - 1$ columns. In the case of databases there must be one choice of entries such that there are at least two rows containing these entries in the given columns. In the case of (absolutely secure) secret sharing, for every choice of entries for the $\ell - 1$ columns there must be many rows equal to these entries.

The main task of the present thesis is to investigate mathematical problems lying in-between. There is a very wide class of problems which are

special cases of the general problem obtained as a common generalization of problems in database theory and secret sharing.

In Section 2 we describe Shamir's original idea and provide alternative solutions for some weaker secret sharing schemes based on his design.

In Section 3 we will investigate the minimum number of rows in a matrix in which the followings hold. Any $\ell = 2$ columns form a key, but no single column does in a stronger sense: not only that there is a pair of equal entries in each column, but there are at least k equal entries. Also we will look at graphs serving as access structures which, as we will see, can be regarded as a more general case of this one.

These definitions do not depend on the actual values (entries) in the matrix, only on their equality in the columns, we can replace the column by the partition of the set of n rows defined by equality on its entries (throughout this thesis the two terminologies will be used alternatively). The conditions on these partitions are nothing else but conditions on the sizes of the ℓ-wise and $\ell - 1$-wise intersections. So, in the case of $\ell = 2$ the condition is that each partition has a class with size at least k, but the pairwise intersections of these classes do not have two or more elements. This is a problem of Steiner systems, which have been solved years ago. We have found a short asymptotic solution in [9].

In Section 4 we present some relevant constructions concerning matrix representations of closure operations done by Demetrovics, Füredi and Katona ([3]) in order to lay the groundwork for our main result.

Finally in Section 5 the analogous problem for $\ell = 3$ is considered. In the form of partitions our conditions are the following ones: 1. for any two partitions there is a c-element subset, which is covered by both of them, each 2-element subset of $[n]$ is covered by at most two different partitions. In [10] we found asymptotic bounds on the minimum size of the underlying set $[n]$ where such partitions can be constructed.

2 Shamir's (k, n) Threshold Scheme

2.1 Definitions

In a secret sharing scheme with access structure $A \subset 2^n$ (where n is the number of participants) the goal is to divide a secret S into n pieces $S_1, \ldots, S_n$ in such a way that:

(i) knowing a set B of S_i pieces makes S computable if $\exists C : C \in A, C \subseteq B$; (ii) if no such C exists then knowledge of a set B of S_i pieces leaves S completely undetermined (in the sense that all of its possible values are equally likely).

If A consists of precisely of all k-element subsets of 2^n then this scheme is called a (k, n) threshold scheme.

Shamir's scheme described in [21] is based on polynomial interpolation: given k points in the 2-dimensional plane $(x_1, y_1), \ldots, (x_k, y_k)$ with distinct x_i's, there is one and only one polynomial $q(x)$ of degree $k-1$ such that $q(x_i) = y_i$ for all i. To divide the secret S into pieces S_i, we pick a random $k-1$ degree polynomial $q(x) = a_0 + a_1 * x + \cdots + a_{k-1} * x^{k-1}$ in which $a_0 = S$, and evaluate: $S_1 = q(1), \ldots, S_i = q(i), \ldots, S_n = q(n)$. A piece received by the ith participant will be the ordered pair (i, S_i). Knowing any k of these pieces is sufficient to find the coefficients of $q(x)$ by interpolation and then evaluate $S = q(0)$. On the other hand knowledge of just $k-1$ of these values is not enough to calculate S. To make this claim more precise, we use

modular arithmetic instead of real arithmetic. The set of integers modulo a prime number p forms a field in which interpolation is possible. Given an integer valued secret S, we pick a prime p which is bigger than both S and n. The coefficients $a_1, \ldots, a_{k-1}$, in $q(x)$ are randomly chosen from a uniform distribution over the integers in $[0, p)$, and the values $D_1, \ldots, D_n$ are computed modulo p.

Let us now assume that $k-1$ of these n pieces are obtained by an opponent. For each possible value S' in $[0, p)$ he can construct exactly one polynomial $q'(x)$ of degree $k-1$ such that $q'(0) = S'$ and $q'(i) = S_i$ for the $k-1$ given pieces. By construction, these p possible polynomials are equally likely, and thus there is absolutely nothing the opponent can deduce about the real value of S.

Some of the useful properties of this (k, n) threshold scheme are:

(1) The size of each piece does not exceed the size of the original data.

(2) When k is kept fixed, S_i pieces can be dynamically added or deleted without affecting the other S_i pieces.

(3) By giving a participant multiple S_i pieces we can create a hierarchical scheme in which we can assign a weight to each participant according to their importance. For example if the secret is the ability to sign checks in a company then we can give the company's president three values of $q(x)$, each vice-president two values of $q(x)$ and each executive one value of $q(x)$. This way a $(3, n)$ threshold scheme enables checks to be signed either by any three executives, or by any two executives one of whom is a vice-president,

or by the president alone.

2.2 Databases and weak schemes

Shamir's scheme is strong in the sense that after obtaining $k-1$ pieces the adversary still cannot reduce the number of possible secrets from the original. However in the main part of this paper we will consider 'weaker' schemes in the sense that we do not require that all secrets remain valid after the opponent manages to get hold of $k-1$ pieces only that there should be at least a pre-determined number d of them that are still possible.

Let us translate the scheme into a database of which a simple model is a matrix. Each row in the matrix corresponds to a secret, each column corresponds to a participant. Shamir's scheme would translate into a (p^k, n) matrix with each row representing one of the p^k polynomials possible over the field. It can be easily proved that the construction is minimal in the number of rows necessary for the given access structure. Our main goal is to minimize the number of rows needed in the case of other, weaker schemes. However finding a solution with i rows won't necessarily mean that we can easily obtain a solution for any $j > i$ as well because simply adding a row is not always possible without damaging the structure of the database. This new row could introduce completely new valid choices of values and it is not at all guaranteed that these choices are already covered by our database. Therefore finding the minimum number of rows won't always be enough, sometimes we'll need multiple constructions to show that the database can

be realized with the given parameters.

Let us see a specific example. Given Shamir's polynomial secret sharing structure, we'd like to modify it so that after learning $k-1$ values the adversary still has $\frac{p}{c}$ secrets to choose from, with $c \geq 2, c \in \mathbb{Z}^+$. Following the original proof it is easy to see that this database requires at least $(\frac{p}{c})^k$ rows. We will show 3 solutions for this problem, all with different number of rows.

Solution 1: We start with all of the p^k rows from the original structure and we delete each of those where the value given to a specific participant is not divisible by c. This will leave us with $\frac{p^k}{c}$ rows. Suppose that an adversary manages to learn $k-1$ shares. If the share of this designated person is among them, then all possible choices for the secret are still valid. If not, then knowing this additional information means that only $\frac{p}{c}$ secrets remain possible. So this construction achieves the desired result with $\frac{p^k}{c}$ rows and has a further property: after learning $k-2$ shares the adversary can't discard any choice yet, only after learning $k-1$ shares can he get additional information.

Solution 2: We use the original design but for $\frac{p}{c}$ instead of p. This gives us $(\frac{p}{c})^k$ rows but only $\frac{p}{c}$ possible secrets. Therefore we duplicate the database c times, each with different values and secrets. Now we have $c(\frac{p}{c})^k$ rows and p secrets. With this construction the adversary will know immediately after obtaining only 1 share which $\frac{p}{c}$ secrets remain possible but he won't get any further information out of the next $k-2$ shares. In some sense this is the

exact opposite of the previous solution.

Solution 3: Again we start with the original design modified for $\frac{p}{c}$ instead of p secrets. We partition the rows into $\frac{p}{c}$ sets, according to the secret corresponding to them. We then further partition each set into c roughly equal subset independent from the others. For each subset we change the secret corresponding to it, so that no two new secrets will be equal. It doesn't matter if the subsets aren't of equal size or the way we partition them. With this construction we will have $c\frac{p}{c} = p$ secrets in our database. Because we didn't modify the underlying structure, nor the data received by the participants it follows from the original proof that after obtaining $k-1$ shares the adversary will still have $\frac{p}{c}$ rows to choose from. These rows come from different sets, so the secret corresponding to any 2 of them cannot be the same. This construction uses the absolute minimum number of rows possible, but we cannot say anything about the behavior of the structure between the first and the $k-1$th information obtained by the adversary.

3 Asymptotic packing of k-sets

3.1 Introduction

Let $[n] = \{1, 2, \ldots, n\}$ be an n-element set, $k \geq 3$ an integer. The family $\mathcal{P}(n,k) \subset \binom{[n]}{k}$ is called a *packing* if $F, G \in \mathcal{P}(n,k)$ $(F \neq G)$ implies $|F \cap G| < 2$. A packing is called *exact* if there is an element $H \in \mathcal{P}(n,k)$ for every pair $a, b \in [n]$ $(a \neq b)$ such that $\{a,b\} \subset H$. These exact covers are also called *Steiner systems* and denoted by $\mathcal{S}(n,k)$. There are some obvious divisibility conditions for n and k which are needed for the existence of an exact cover $\mathcal{S}(n,k)$. Richard Wilson's classical theorem [22] states that these necessary conditions are also sufficient.

It is easy to see by double counting that

$$|\mathcal{P}(n,k)| \leq \frac{\binom{n}{k}}{\binom{k}{2}} \tag{3.1}$$

holds for every packing with equality for an exact packing. Wilson's theorem gives an infinite sequence of n's (k is fixed) for which an exact cover $\mathcal{P}(n,k)$ exists. Our modest goal, as described in [9], is to give a cover $\mathcal{P}(n,k)$ for every n so that the inequality (3.1) is an asymptotic equality (k is fixed). However our construction is very simple. More effort is needed to prove its asymptotic behavior than to describe the construction itself.

A packing $\mathcal{P}(n,k)$ defines a graph $G = ([n], E)$ where E consists of the pairs $\{a,b\}$ $(a \neq b)$ which are subsets of the members of $\mathcal{P}(n,k)$. The graph will be denoted by $G(\mathcal{P}(n,k))$.

The Turán graph $T(p,k)$ has pk vertices, $V = V_1 \cup \ldots \cup V_k$, $V_i \cap V_j = \emptyset (i \neq j)$ and $|V_1| = \ldots = |V_k| = p$. The vertices a and b are joined by an edge if and only if $a \in V_i, b \in V_j$ and $i \neq j$.

Proposition 3.1 *If p is prime number, $k \leq p$ then there is a packing $\mathcal{P}(pk,k)$ satisfying $G(\mathcal{P}(pk,k)) = T(p,k)$.*

Theorem 3.2 *Let $k \geq 3$ be a fixed integer. There is a packing $\mathcal{P}(n,k)$ for every n such that*

$$\lim_{n\to\infty} \frac{|\mathcal{P}(n,k)|}{\binom{n}{k}} = \binom{k}{2}. \tag{3.2}$$

The construction will be based on Proposition 3.1.

As it turns out Erdős and Hanani proved this result already in 1963 [6], however in their proof they only gave the exact packing of a Turán graph where the size of the classes was very general. In contrast my construction only requires the sizes to be primes. This case can be proven much more easily, although we have to use deeper number theory tools. The creating of the asymptotically good construction is the same as in [6], the important observation is that the asymptotic quality remains true even so.

3.2 Proofs

Proof of Proposition 3.1. Let the vertices of the Turán graph be the ordered pairs (i,j) where $1 \leq i \leq k, 1 \leq j \leq p$, the classes of the vertices

are $V_i = \{(i,j) :\ 1 \le j \le p\}$ $(1 \le i \le k)$. Let c and d be integers satisfying $1 \le c, d \le p$ and define the set

$$F(c,d) = \{(i, ci + d) :\ 1 \le i \le k\}$$

where $ci + d$ is considered $\mod p$. It is obvious that $F(c,d)$ has exactly one element in each V_i, therefore $|F(c,d)| = k$. Only such pairs can be their subsets which are in distinct classes that is the edges of the Turán graph.

Now it will be shown that every such edge $\{(i_1, j_1), (i_2, j_2)\}$ $(i_1 \neq i_2)$ is a subset of exactly one of the sets $F(c,d)$. This inclusion holds if and only if the following equations hold:

$$ci_1 + d \equiv j_1 \pmod{p}$$

$$ci_2 + d \equiv j_2 \pmod{p}.$$

This equation system in c and d has a unique solution when $i_1 \neq i_2$ since its matrix is non-singular.

This proves that the family

$$\mathcal{P}(pk, k) = \{F(c,d) :\ 1 \le c, d \le p\}$$

is a packing satisfying the conditions of the proposition. □

Let $\mathcal{R}(p,k)$ denote the packing obtained in Proposition 3.1. Theorem 3.2 will be proved by combining copies of $\mathcal{R}(p,k)$'s.

Proof of Theorem 3.2.

Construction. Define $p_1(n,k)$ as the largest prime number such that $p_1(n,k)k \le n$. If $i \ge 1$ then let $p_{i+1}(n,k)$ be the largest prime number

satisfying $p_{i+1}(n,k)k \leq p_i(n,k)$. If there is no such prime number, that is, $k^2 > p_i(n,k)$ (as for using Lemma 3.1 $p_i \geq k$ is required) then let us stop and denote the last i by u. (Of course u also depends on n and k, from now on these arguments are also omitted from $p_i(n,k)$ to avoid too long formulas.)

Let $\mathcal{T}_u(n,k) = \mathcal{R}(p_u,k)$. Since $G(\mathcal{R}(p_{u-1},k))$ is a Turán graph with classes of size $p_{u-1} \geq p_u k$, one can place an isomorphic copy of $\mathcal{T}_u(n,k)$ in each of these classes in such a way that the so obtained family of k-element subsets on $[p_{u-1}k]$ is a packing. Denote this family by $\mathcal{T}_{u-1}(n,k)$. Suppose that $\mathcal{T}_i(n,k)$ is a packing of k-element subsets on $[p_i k]$. The graph $G(\mathcal{R}(p_{i-1},k))$ is a Turán graph with classes of size $p_{i-1} \geq p_i k$. Place isomorphic copies of $\mathcal{T}_i(n,k)$ in each of these classes. The so obtained family of k-element subsets, $\mathcal{T}_{i-1}(n,k)$ is a packing on $[p_{i-1}k]$. The construction of $\mathcal{T}_1(n,k)$ uses $\mathcal{R}(p_1,k)$, there is sufficient room for this, since $p_1 k \leq n$. To indicate the slight difference between the underlying sets, the notation $\mathcal{T}(n,k)$ will be used when $\mathcal{T}_1(n,k)$ is considered on $[n]$ (rather than $[p_1 k]$).

Investigating the asymptotic behavior. Let $h_i(n,k)$ be the number of pairs which are not covered by the members of $\mathcal{T}_i(n,k)$ that is the number of edges of the complement of $G(\mathcal{T}_i(n,k))$ (with the underlying set $[p_i k]$). On the other hand, $h(n,k)$ is the same thing as $h_1(n,k)$, but on the underlying set $[n]$.

Lemma 3.3

$$h(n,k) = o\left(\binom{n}{2}\right).$$

Proof. $h_u(n,k) = k\binom{p_u}{2}$ is trivial. Suppose that $h_i(n,k)$ is known, try to express $h_{i-1}(n,k)$. The non-covered pairs in one class of $\mathcal{T}_{i-1}(n,k)$ are either in the here placed $\mathcal{T}_i(n,k)$ or one of their end-points are in the $p_{i-1} - p_i k$ element set not covered by the elements of of the copy of $\mathcal{T}_i(n,k)$. The number of edges of the previous kind is $h_i(n,k)$ while the number of edges of the latter kind is at most $p_{i-1}(p_{i-1} - p_i k)$. This is true for every class, therefore

$$h_{i-1}(n,k) \leq k\left(h_i(n,k) + p_{i-1}(p_{i-1} - p_i k)\right). \tag{3.3}$$

Here it is crucial that we took the largest prime number p_i satisfying $p_i k \leq p_{i-1}$ and the fact that this is near to $\frac{p_i}{k}$, that is, the prime numbers are densely situated on the number line. This is why the following statement will be used to obtain a good upper bound on (3.3). It is an easy consequence of the prime number theorem. (See e.g. [2].)

Theorem 3.4 *Let $0 < \varepsilon$ be a real number. If $r > r(\varepsilon)$ then there is a prime number between $r(1-\varepsilon)$ and r.*

The following almost special case of this theorem is also needed, an elementary proof of which can be found in [7].

Theorem 3.5 (Chebyshev) *If $2 \leq N$ is a natural number then there is a prime number between N and $2N$.*

Since p_1 is the largest prime number with $p_1 \leq \frac{n}{k}$, the inequality $\frac{n}{2k} \leq p_1$ holds by Theorem 3.5. Similarly, $\frac{p_i}{2k} \leq p_{i+1}$ is also true. Hence we have $\frac{n}{(2k)^u} \leq p_u < k^2$ and u, as a function of n, tends to infinity with n.

By definition we have

$$k^a p_b \leq p_{b-a} (a < b). \tag{3.4}$$

This implies $k^{u-v} p_u \leq p_v$. Let $v = \lfloor \frac{u}{2} \rfloor$ what also tends to the infinity with n. Choose $n(\varepsilon)$ so that $r(\varepsilon) \leq p_v$ (where $v = v(n(\varepsilon), k)$) and

$$\frac{1}{k^{v-1}} \leq \varepsilon \tag{3.5}$$

both hold. Then $r(\varepsilon) \leq p_i$ holds for $i \leq v$. Applying Theorem 3.4 with $r = \frac{p_{i-1}}{k} \geq p_i$ we obtain that $(1-\varepsilon)\frac{p_{i-1}}{k} \leq p_i$ holds for $i \leq v$. This is equivalent to $p_{i-1} - p_i k \leq \varepsilon p_{i-1}$. Inequality (3.3) implies

$$h_{i-1}(n,k) \leq k \left(h_i(n,k) + \varepsilon p_{i-1}^2 \right) \text{ if } i \leq v. \tag{3.6}$$

Induction on j proves the generalization of (3.6):

$$h_{i-j}(n,k) \leq k^j h_i(n,k) + \varepsilon \sum_{\ell=1}^{j-1} k^{j+1-\ell} p_{i-\ell}^2 + \varepsilon k p_{i-j}^2 \ (1 \leq j < i). \tag{3.7}$$

Using (3.4) in (3.7)

$$h_{i-j}(n,k) \leq k^j h_i(n,k) + \varepsilon \sum_{\ell=1}^{j-1} \frac{1}{k^{j-1-\ell}} p_{i-j}^2 + \varepsilon k p_{i-j}^2 = k^j h_i(n,k) + \varepsilon p_{i-j}^2 k \sum_{\ell=1}^{j} \frac{1}{k^{j-\ell}}$$

$$\leq k^j h_i(n,k) + \varepsilon p_{i-j}^2 \frac{k^2}{k-1} \ (i \leq v)$$

is obtained. We will actually use this inequality for $i = v$ and $j = v - 1$:

$$h_1(n,k) \leq k^{v-1} h_v(n,k) + \varepsilon p_1^2 \frac{k^2}{k-1}. \tag{3.8}$$

Obvious analogues of (3.3) and (3.6) are

$$h(n,k) \le k(h_1(n,k) + n(n - p_1 k))$$

and

$$h(n,k) \le k(h_1(n,k) + \varepsilon n^2).$$

Combine the latter one with (3.8) to obtain

$$h(n,k) \le k^v h_v(n,k) + \varepsilon p_1^2 \frac{k^3}{k-1} + \varepsilon k n^2 \le$$

$$k^v h_v(n,k) + \varepsilon \left(\frac{n}{k}\right)^2 \frac{k^3}{k-1} + \varepsilon k n^2 = k^v h_v(n,k) + \varepsilon n^2 \frac{k^2}{k-1}. \tag{3.9}$$

Recall that $\mathcal{T}(n,k)$ is a family of k-element subsets of $[n]$. The total number of pairs of elements (potential edges) is $\binom{n}{2}$. The ratio of the uncovered pairs and this number of all pairs for $\mathcal{T}(n,k)$ can be upperbounded using (3.9).

$$\frac{h(n,k)}{\binom{n}{2}} \le \frac{k^v h_v(n,k)}{\binom{n}{2}} + \frac{\varepsilon n^2 \frac{k^2}{k-1}}{\binom{n}{2}}. \tag{3.10}$$

Let us give now an upper bound on the second term of the right hand side applying $n \ge p_1$ and (3.4):

$$\frac{k^v h_v(n,k)}{\binom{n}{2}} \le \frac{k^v h_v(n,k)}{\binom{k^{v-1}p_v k}{2}} = k \cdot \frac{h_v(n,k)}{\binom{p_v k}{2}} \cdot \frac{kp_v - 1}{k^v p_v - 1}.$$

The second factor is at most 1, the third one is less than ε by (3.5). That is, the first term of the right hand side of (3.10) is less than εk if $n \ge n(\varepsilon)$. The second term of the right hand side of (3.9) is at most $3\varepsilon \frac{k^2}{k-1}$ if $n \ge 3$. This proves that (3.10) is at most $\varepsilon \left(k + 3\frac{k^2}{k-1}\right)$ for $n(\varepsilon) \le n$. $\square_{\mathrm{L}}$

Lemma 3.3 implies that the number of edges contained in the members of $\mathcal{T}(n,k)$ is

$$|G(\mathcal{T}(n,k))| = \binom{n}{2} - o\left(\binom{n}{2}\right).$$

Every member contains exactly $\binom{k}{2}$ pairs, therefore

$$|\mathcal{T}(n,k)| = \frac{\binom{n}{2} - o(\binom{n}{2})}{\binom{k}{2}}$$

proves the theorem. $\square_{\mathrm{T}}$

3.3 Graphs as access structures

Recently there has also been a lot of interest in graphs as access structures. The idea is the following:

1. Every point represents a participant.

2. If two points are joined with an edge then the shares given to the participants corresponding to the points are enough to reconstruct the secret.

3. The shares corresponding to an independent set of points give no clue about the secret whatsoever.

It is easy to see that this approach corresponds to an access structure where the size of any minimal key is exactly 2 and conversely, any such access structure can be represented by a simple graph. There are already lots of results concerning these kinds of access structures, I would like to present a new one:

Theorem 3.6 *If a graph $G = (V, E)$ contains n totally independent edges, that is $2n$ points $a_1, \ldots, a_n, b_1, \ldots, b_n$ such that $\forall i (a_i, b_i) \in E$ and $\forall i \neq$*

$j(a_i, a_j), (b_i, b_j), (a_i, b_j) \notin E$, then the access structure corresponding to G cannot be realized with less then p^{2^n} rows, where p is the number of possible secrets.

Proof. Consider the points $a_1, \ldots, a_n$. These are independent, so for any valid choice of values $c_1, \ldots, c_n$, there exist at least p rows where $\forall i$ c_i appears in the column of a_i. The same holds for $a_1, \ldots, a_{n-1}, b_n$, with values $c_1, \ldots, c_{n-1}, d_n$. $(a_n, b_n) \in E$, therefore each (c_n, d_n) pair can appear at most in 1 row. We can give a lower estimate for the number of rows that have $c_1, \ldots, c_{n-1}$ in the columns $a_1, \ldots, a_{n-1}$: Any c_n appears at least p times in a_n. In these rows the value in b_n must be different everywhere, so there are at least p different valid values in b_n. These too must appear at least p times in the column, independently from the others. Therefore there are at least p^2 rows with the same values in $a_1, \ldots, a_{n-1}$. Of course this is true regardless of our choice in $c_1, \ldots, c_{n-1}$. We can apply the same argument for $a_1, \ldots, a_{n-2}, b_{n-1}$ as well. Now we can give a lower estimate for the number of rows that have $c_1, \ldots, c_{n-2}$ in the columns $a_1, \ldots, a_{n-2}$: Any c_{n-1} appears at least p^2 times in a_{n-1}. In these rows the value in b_{n-1} must be different everywhere, so there are at least p^2 different valid values in b_{n-1}. These too must appear at least p^2 times in the column, independently from the others. Therefore there are at least $(p^2)^2 = p^{2^2}$ rows with the same values in $a_1, \ldots, a_{n-2}$. This argument can be used recursively until we get the following result: the number of rows having the same value in a_1 is at least $p^{2^{n-1}}$. This is of course true for b_1 as well and the values appearing in

b_1 in those rows where c_1 is given must be different as $(a_1, b_1) \in E$. So b_1 contains at least $p^{2^{n-1}}$ different values which (independently from each other) appear at least $p^{2^{n-1}}$ times. This proves the theorem. $\square$

4 Closure operations

4.1 Definitions

Let us now return to our other interpretation when the database is represented by a matrix. Given a subset $A \subseteq X$ suppose that the values of the columns belonging to A are known. As there may exist more than one row containing these values in the columns specified by A the row is not necessarily determined. However all these rows might share the same data in a column $b \notin A$. We say that b belongs to the closure $L(A)$ of A if this is true for b for any choice of values in the columns belonging to A. Formally defined:

Let M be a matrix of m rows and n columns, X denoting the set of columns. If $A \subseteq X, a \in X$ and M contains no two rows equal in A but different in a then we say that A *implies* a. The *closure* of A is

$$L_M(A) = \{a : a \in X, A \text{ implies } a\} \tag{4.1}$$

.

The following rules can be easily seen to be valid for $L_M = L$:

$$A \subseteq L(A), \tag{4.2}$$

$$A \subseteq B \Rightarrow L(A) \subseteq L(B), \tag{4.3}$$

$$L(L(A)) = L(A). \tag{4.4}$$

A function $L : 2^X \to 2^X$ is called a *closure operation* if it satisfies (4.2)-(4.4).

For the other direction, let L be an arbitrary closure operation on X which is an n-element set. Then there exists an $m \times n$ matrix M such that $L_M = L$ ([3]). We say that M *represents* L. The definition of our main target is the following:

$$s(L) = \min\{m : M \text{ is an } m \times n \text{ matrix, } L_M = L\}. \tag{4.5}$$

Now we define the class of keys, an important subset determined by the closure operation. K is said to be a key in L if $L(K) = X$. $\mathcal{K} = \mathcal{K}(L)$ denotes the family of *minimal keys* (K is a minimal key if it is a key but no proper subset of K is a key). It can be easily seen that $K_1, K_2 \in \mathcal{K}, K_1 \neq K_2$ imply $K_1 \not\subset K_2$. Those families of subsets that satisfy this condition are called *Sperner-families*, therefore $\mathcal{K}$ is a Sperner-family. We say that a matrix M *represents* a given Sperner-family $\mathcal{K}$ if $\mathcal{K} = \mathcal{K}(L_M)$ is true. The maximal non-keys are called *antikeys*. Their family is defined by

$$\mathcal{K}^{-1} = \{A : \nexists B \in \mathcal{K}, B \subseteq A, \text{ and } \forall C \subseteq X, A \subset C : \ \exists D \in \mathcal{K}, D \subseteq C\}.$$

Lemma 4.1 *[3] M represents the Sperner-family $\mathcal{K}$ iff for any $A \in \mathcal{K}^{-1}$ M has two different rows having the same entries in the columns in A and any two rows equal in $K \in \mathcal{K}$ are equal everywhere.*

Proof. If M represents $\mathcal{K}$, then $\mathcal{K} = \mathcal{K}(L_M)$ is true. $K \in \mathcal{K}$ implies $L_M(K) = X$, the second condition obviously holds. Similarly, $A \in \mathcal{K}^{-1}$ implies $L_M(A) \neq X$ and from this we obtain the first condition.

Conversely, if both conditions are true for M and $\mathcal{K}$, then (i) $L_M(A) \neq X$ holds for any $A \in \mathcal{K}^{-1}$ and (ii) $L_M(K) = X$ holds for any $K \in \mathcal{K}$.

(ii) and (4.3) imply that $L_M(C) = X$ if $C \supseteq K$ for some $K \in \mathcal{K}$. If we suppose that C is not a superset of a member of $\mathcal{K}$, then by definition there exists an $A \in \mathcal{K}^{-1}$ such that $C \subseteq A$. From (i) and (4.3) we get $L_M(C) \neq X$. Therefore $L_M(C) = X$ holds exactly for the supersets of members $\mathcal{K} : \mathcal{K} = \mathcal{K}(L_M)$. □

The following definition is an analogue of (4.5):

$$s(\mathcal{K}) = \min\{m : M \text{ is an } m \times n \text{ matrix representing } \mathcal{K}\}. \tag{4.6}$$

where $\mathcal{K}$ is a Sperner-family on an n-element set.

The *k-uniform closure operation* on an n-element groundset X is defined by:

$$L_k^n(A) = \begin{cases} X, \text{ if } |A| \geq k \\ A, \text{ if } |A| < k \end{cases} \tag{4.7}$$

The family of all k-element subsets of X is denoted by $\binom{X}{k}$. Usually there exists several closure operations with the same class $\mathcal{K}$ of minimal keys ($\mathcal{K} = \mathcal{K}(L)$). Our next Lemma states that if $\mathcal{K}$ is the family of all k-element subsets of X, then L is uniquely determined by $\mathcal{K} = \mathcal{K}(L)$.

Lemma 4.2 *[3] Let any closure operation L be defined on an n-element set X. Then*

$$\mathcal{K}(L) = \binom{X}{k} \text{ iff } L = L_k^n.$$

Proof. $\mathcal{K}(L_k^n) = \binom{X}{k}$ is obvious therefore we have to prove the converse statement only. Suppose that $a \in L(A) - A$ for some $A \subseteq X$ such that $|A| < k$. Then one can find a set B satisfying $|B| = k, B \supseteq A \cup \{a\}$. (2) implies $L(B-a) \supseteq B - a$; (4.3) implies $L(B-a) \supseteq L(A) \ni a$. From this $L(B-a) \supseteq B$ follows. We get $L(B-a) = L(L(B-a)) \supseteq L(B) = X$ by (4.4) and (4.3). Consequently, there is a set $B - a$ of cardinality $< k$ with closure X. This contradiction shows that $a \in L(A) - A$ cannot exist if $|A| < k : L(A) = A$. $L(A) = X$ for $|A| \geq k$ is easily obtained from $\mathcal{K}(L) = \binom{X}{k}$ and (4.3). $L = L_k^n$ holds, the lemma is proved. □

4.2 Minimum representation of uniform closure operations

We prove the following statements for sake of completeness.

Lemma 4.3 *[3] If an $m \times n$ matrix M represents $\mathcal{K}$, then*

$$\binom{m}{2} \geq |\mathcal{K}^{-1}|.$$

Proof. If $A \in \mathcal{K}^{-1}$, then by Lemma 4.1 there exist two different rows i, j such that they are equal in A. Take another element B of $\mathcal{K}^{-1}$. For

B two such rows also exist, let them be i' and j'. If these unordered pairs $\{i,j\},\{i',j'\}$ are equal, then these two different rows are equal in $A\cup B$. Therefore $L(A\cup B)\neq X$ and there exists a $C\supseteq A\cup B$ with $C\in\mathcal{K}^{-1}$. By the definition of $\mathcal{K}^{-1}$ this is only possible when $C=A$ and $C=B$, contradicting our original supposition $A\neq B$. Consequently for different members of $\mathcal{K}^{-1}$ we are able to assign different pairs of rows satisfying the above condition, so the number of pairs of rows of M must be $\geq|\mathcal{K}^{-1}|$. □

Lemma 4.4 *[3]*

$$\binom{s(L_k^n)}{2}\geq\binom{n}{k-1}. \tag{4.8}$$

Proof. Let M be an $s(L_k^n)\times n$ matrix representing L_k^n. By Lemma 4.2, M also represents $\mathcal{K}(L_k^n)=\binom{X}{k}$. It is easy to see that $\mathcal{K}^{-1}(L_k^n)=\binom{X}{k-1}$. Then (4.8) follows by Lemma 4.3. □

We will show that (4.8) gives a fairly good lower estimate on $s(L_k^n)$. It is sharp for $k=1,2,n-1$. It seems to be sharp for $k=3$ and $n\geq 7$.

Theorem 4.5 *[3]*

$$s(L_1^n)=2,\qquad s(L_2^n)=\lceil(1+\sqrt{1+8n})/2\rceil,$$
$$s(L_{n-1}^n)=n,\quad s(L_n^n)=n+1.$$

where $\lceil x\rceil$ denotes the smallest integer $\geq x$.

Proof. By Lemma 4.2, $s(L_k^n)=s(\mathcal{K}(L_k^n))=s(\binom{\binom{X}{k}}{})$. We use this last form for the proof.

For $k = 1$, (4.8) gives $s(L_1^n) \geq 2$. The construction of two rows, one filled with 0 everywhere, the other filled with 1 everywhere proves the equality.

For $k = 2$, (4.8) gives

$$\binom{s(L_2^n)}{2} \geq n \tag{4.9}$$

.

Suppose now that $s(L_2^n)$ satisfies (4.9), we will construct an $s(L_2^n) \times n$ matrix M that represents $\binom{X}{2}$: any column of M will contain exactly two zeros, with different columns containing different pairs of zeros. All of the other entries of the ith row will be equal to i ($1 \leq i \leq s(L_2^n)$). Using Lemma 4.1 it is obvious that M represents $\binom{X}{2}$. The least integer satisfying (4.9) can be expressed in the form in the theorem.

For $k = n - 1$, (4.8) gives $s(L_{n-1}^n) \geq n$. The identity matrix I_n gives the equality.

The case $k = n$ needs another lemma. If M is an $m \times n$ matrix let $G(M)$ denote the graph whose vertices are the rows of M, two vertices are connected with an edge iff the set A of columns where the two rows are equal is non-empty. The edge is *labeled* by A.

Lemma 4.6 *[3] Let M be a matrix and let $A_1, \ldots, A_r$ be the labels along a circuit of G(M). Then*

$$\bigcap_{\substack{i=1 \\ i \neq j}}^{r} A_i - A_j = \emptyset \quad (1 \leq j \leq r). \tag{4.10}$$

Proof. Suppose that, on the contrary, (4.10) is non-empty, that is there is a column, say the uth one, which is an element of all A_i but A_j. Let the vertices of the circuit be $K-1, \ldots, k_r$ in such a way that the edge (k_i, k_{i+1}) is labeled by A_i $(1 \leq i < r)$ and $(k_r, K-1)$ is labeled by A_r. From $u \in A_{j+1}$ it follows that the k_{j+1}st and k_{j+2}nd entries of the uth column are equal. The same holds for the k_{j+2}nd and k_{j+3}rd entries, etc. Consequently, the k_{j+1}st, k_{j+2}nd, ..., k_rth, k_1st, ...,k_jth entries in the uth column are all equal. This leads to $u \in A_j$ contradicting the assumption, thus proving the lemma.

Now suppose that the $m \times n$ matrix M realizes $\binom{X}{n} = \{X\}$. By Lemma 4.1 there is an edge in $G(M)$ labeled with A for any $(n-1)$-element subset of X. $G(M)$ has n different edges of this kind. These edges cannot form a circuit because the $(n-1)$-element subsets cannot satisfy (4.10), the lemma is applicable. $G(M)$ has at least $n+1$ vertices: $s(L_n^n) \geq n+1$. The following construction gives the equality:

$$\begin{pmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

The proof is complete. □

Substituting $k = 3$ into (4.8) we obtain

$$s(L_3^n) \geq n.$$

$s(L_3^3) = 4 > 3$ is proved and $s(L_3^6) > 6$ can be verified by checking all the cases. We conjecture that the above inequality is sharp for all other cases:

Conjecture 1 *$s(L_3^n) = n$ for $n \geq 7$.*

We are able to reduce this conjecture in the case $n = 3r + 1$ for another conjecture concerning a ceratin kind of resolvable Steiner triple systems:

Conjecture 2 *There is a system of 3-element subsets of an n (=3r+1)-element set {1,2, . . . ,n } satisfying the following conditions:*

(1) Any pair of elements is contained in exactly two 3-sets.

(2) The family of 3-sets can be divided into n subfamilies where the ith subfamily is a partition of {1,2, . . . ,n } - { i }.

(3) Exactly one pair of members of two different subfamilies meet in 2 elements.

We show the construction of an $n \times n$ matrix M representing $L_3^n (n = 3r + 1)$ if the family in Conjecture 2 exists. We write zeros in the main diagonal. The ith row jth entry will be l if i is an element of the lth triple in the j sub-family. It follows by condition (3) that for any two columns of M there are two rows equal in these columns. The rows are, of course, different due to the zeros. The first condition of Lemma 4.1 is satisfied.

Condition (1) implies that any two rows agree in exactly two entries. Hence there are no two rows equal in any given triple of columns. The second condition of Lemma 4.1 is also satisfied, M really represents L_3^n.

Conjecture 2 follows for n (mod 1) or 4 (mod 12) from the following result of Hanani [16], [17]. There exists a Steiner system $S(4, 2, n)$ for these n's. (I.e. we have a 4-uniform subsystem $\mathcal{S}$ on n-element set V such that for every two $v_1, v_2 \in V$ there exists exactly one member $S \in \mathcal{S}$ such that $\{V - 1, v_2\} \subset S$.) Consider the 4-uniform set-system $\mathcal{S}$ over $\{1, 2, \ldots, n\}$ and replace every member $S \in \mathcal{S}$ with 4 3-element subsets. The obtained set-system $\mathcal{F}$ meets the condition of Conjecture 2, where the ith subfamily

$\mathcal{F}_1 = \{S - \{i\} : i \in S \in \mathcal{S}$.

Therefore the following theorem is proved:

Theorem 4.7 *[3]*

$$s(L_3^n) \geq n,$$

$$s(L_3^n) = n \text{ for } n = 12k + 1 \text{ for } n = 12k + 4.$$

Corollary 1 $n \leq s(L_3^n) \leq n + 8$.

Proof. It follows from Theorem 4.7 and the inequality $s(L_3^n) \leq s(L_3^{n+1})$.

□

The main result of Demetrovics, Füredi and Katona is the following theorem:

Theorem 4.8

$$\sqrt{2}(\frac{1}{k-1})^{(k-1)/2} n^{(k-1)/2} < s(L_k^n) < 2^{3k/2} n^{(k-1)/2} \ (2 \leq k < n). \tag{4.11}$$

Proof. The left-hand side of (4.11) follows easily from (4.8), we only need to give a construction for the right-hand side.

Let p be a prime number. We will show that there exists a set of D of cardinality $2\lfloor\sqrt{p}\rfloor$ such that any integer satisfies

$$i \equiv d_1 - d_2 \pmod{p} \tag{4.12}$$

for some members d_1, d_2 of D. We define D as

$$D = \{0, 1, 2, \ldots, a-1, 2a, 3a, \ldots, (a-1)a\}$$

where $a = \lceil\sqrt{p}\rceil$. For any i satisfying $0 \leq i < p$ we will express it in the form $i = al + r$ $(0 \leq r < a)$. If $1 \leq l \leq a-2$ and $0 < r < a$, then $d_1 = (l+1)a$ and $d_2 = a - r$ obviously satisfy (4.12). If $i = al$ $(2 \leq l \leq a-1)$, then $d_1 = al$ and $d_2 = 0$ are adequate. $a = 3a - 2a$ and the rest can be expressed as a difference of zero and one of the numbers $1, 2, \ldots, a-1$. (Here we suppose that $3 \leq a - 1$. Otherwise we would have $p \leq 9$ and these cases can be checked separately.)

For the cardinality of D we have

$$|D| = 2a - 2 = 2(\lceil\sqrt{p}\rceil - 1) = 2\lfloor\sqrt{p}\rfloor.$$

Let P be defined in the following way:

$$P = \{c_{k-1}x^{k-1} + c_{k-2}x^{k-2} + \cdots + c_1x + c_0 : c_0, \ldots, c_{k-1} \in D, c_{k-1} = 0 \text{ or } 1\}.$$

Note that

$$|P| = 2^k\lfloor\sqrt{p}\rfloor^{k-1}.$$

Let M be a $|P| \times p$ matrix. Let us associate its rows with elements of P. Define the jth entry of the row that is associated with $z(x) \in P$ as $z(j) \pmod p$ $(0 \leq j \leq p-1, 0 \leq z(j) \leq p-1)$. We now prove that M represents L_k^p. For this it is sufficient to show (by Lemma 4.2) that M represents $\binom{X}{k}$ (where $|X| = p$). Here we can use Lemma 4.1, we only have to verify its conditions with $\mathcal{K} = \binom{X}{k}, \mathcal{K}^{-1} = \binom{X}{k-1}$.

Suppose now that the rows associated with $z_1(x)$ and $z_2(x)$ have k equal entries:

$$z_1(t_i) \equiv z_2(t_i) \pmod p \ (0 \leq t_1 < \cdots < t_k < p).$$

Then looking at the polynomial $z_1(x) - z_2(x)$ of degree $\leq k-1$ we note that it has k different roots. This is a contradiction, proving that z_1 and z_2 are really the same.

Now, if we choose the integers $0 \leq t_1 < \cdots < t_{k-1} < p$ arbitrarily, there should exist two different rows containing equal entries in the t_1st, t_2nd, ..., t_{k-1}st places. Consider the polynomial

$$w(x) = (x - t_1)(x - t_2)\ldots(x - t_{k-1}) = x^{k-1} + a_{k-2}x^{k-2} + \cdots + a_1 x + a_0.$$

To a_i $(0 \leq i \leq k-2)$ we are able to find two elements c_i and c_i' of D such that $a_i \equiv c_i - c_i' (\text{mod } p)$. Then $w(x) = z(x) - z'(x)$ will hold where

$$z(x) = x^{k-1} + c_{k-2}x^{k-2} + \cdots + c_1 x + c_0 \text{ and}$$

$$z'(x) = c'_{k-2}x^{k-2} + \cdots + c'_1 x + c'_0.$$

$z(x)$ and $z'(x)$ are obviously different members of P, also it is easy to see that $z(t_i) \equiv z'(t_i) (\text{mod } p)$ holds for every i. As both conditions of Lemma 4.1 have been verified, M really represents L_k^p. This proves

$$s(L_k^n) \leq 2^k p^{(k-1)/2}.$$

For the general case, given an arbitrary n we will choose a prime number p satisfying $n \leq p \leq 2n$. p exists by Chebyshev's theorem. Then we will construct a matrix representing L_k^p and discard $p - n$ columns. The matrix so obtained represents L_k^n. Hence

$$s(L_k^n) \leq 2^k p^{(k-1)/2} \leq 2^k (2n)^{(k-1)/2} \leq 2^{3k/2} n^{(k-1)/2}.$$

The theorem is proved. □

The method of Theorem 4.8 gives only a good estimate in case of a small k. For instance a much better estimate is known if $k = n/2$. It is proved ([3]) that:

$$s(\mathcal{K}) \leq |\mathcal{K}^{-1}| + 1.$$

holds for any Sperner-family. From this follows:

$$s(L^n_{n/2}) = s\left(\binom{X}{n/2}\right) \leq \binom{n}{n/2} + 1 = 2^{n+o(n)}.$$

5 Partitions acting as access structures

5.1 Introduction

Let $[n] = \{1, 2, \ldots, n\}$ be a finite set, $\mathcal{P} = (P_1, P_2, \ldots, P_m)$ a partition of $[n]$. We will consider the maximum number of such partitions satisfying certain conditions. We say that a partition *covers* a given subset A of $[n]$, if A is a subset of one of the classes P_i.

The following problem was motivated by the theory of relational databases. Find the maximum number $f(n)$ of partitions satisfying the following two conditions:

(i) for any two partitions there is at least one 2-subset of $[n]$, which is covered by both of them,

(ii) each 2-subset of $[n]$ is covered by at most two different partitions.

Let m be the number of partitions. For every pair $(\mathcal{P}_1, \mathcal{P}_2)$ of partitions take a 2-subset of $[n]$ existing by (i): $\gamma(\mathcal{P}_1, \mathcal{P}_2)$. By (ii) $\gamma(\mathcal{P}_1, \mathcal{P}_2) \neq \gamma(\mathcal{Q}_1, \mathcal{Q}_2)$ unless $(\mathcal{P}_1, \mathcal{P}_2) = (\mathcal{Q}_1, \mathcal{Q}_2)$. Hence the following inequality must hold:

$$\binom{m}{2} \leq \binom{n}{2}.$$

This implies $m \leq n$ and $f(n) \leq n$. It was conjectured and partially proved in [3] that equality holds for all $n \geq 7$. It was finally proved in [1]. There are many interesting related results surveyed in [5].

If we specify equality in both conditions, that is

(i') for any two partitions there exists exactly one 2-subset of $[n]$, which is covered by both of them,

(ii') each 2-subset of $[n]$ is covered by exactly two different partitions,

then we arrive at the definition of *orthogonal double covers*, a concept which has also been studied extensively along with its generalizations.

In some investigations exploring the interrelation of database theory and secret sharing [11] (see Section 2) we arrived to the following generalization of the problem above, to which we were able to find a solution in [10]. Our more general conditions are the following:

(iii) for any two partitions there is a k-element subset, which is covered by both of them,

(iv) each 2-subset of $[n]$ is covered by at most two different partitions.

A set of m partitions on the underlying set $[n]$ satisfying (iii) and (iv) will be called an (n, m, k)-*pamily*.

For given n, k let $f(n, k)$ be the maximum m for which an (n, m, k)-pamily exists. For every pair $(\mathcal{P}_1, \mathcal{P}_2)$ of partitions take a k-element set covered by both partitions, this exists by (iii). It determines a set $\Gamma(\mathcal{P}_1, \mathcal{P}_2)$ of $\binom{k}{2}$ 2-subsets of $[n]$ covered by both of them. By (iv) $\Gamma(\mathcal{P}_1, \mathcal{P}_2)$ and $\Gamma(\mathcal{Q}_1, \mathcal{Q}_2)$ are disjoint unless $(\mathcal{P}_1, \mathcal{P}_2) = (\mathcal{Q}_1, \mathcal{Q}_2)$. Hence the following inequality must hold:

$$\binom{m}{2}\binom{k}{2} \leq \binom{n}{2}. \tag{5.1}$$

The discriminant of this quadratic inequality in m can be upperbounded in the following way.

$$\frac{1}{4} + 2\frac{n(n-1)}{k(k-1)} \leq 2\frac{n^2}{k(k-1)}$$

if $k(k-1) \leq 8n$. By (5.1) we obtain

$$m \leq \frac{1}{2} + n\sqrt{\frac{2}{k(k-1)}}.$$

We believe that this is the asymptotically correct upper bound (for fixed k and $n \to \infty$), but we are able to construct only weaker lower bounds.

Theorem 5.1 *Let $2 \leq k$. Then:*

$$f(n,k) \leq n\sqrt{\frac{2}{k(k-1)}} + \frac{1}{2},$$

$$\left.\begin{array}{c} k+3 \text{ if } (k+1)(2k+3) \leq n \\ \dfrac{n}{3k+4} \text{ if } 5k^2 + \dfrac{37}{2}k + 16 \leq n \end{array}\right\} \leq f(n,k).$$

Remark. $k + 3 \geq \frac{n}{3k+4}$ holds for small ns.

Our upper and lower bounds are relatively far apart for small k, e.g. when k is fixed and n tends to infinity. They are much closer when k is about $\sqrt{n}$. This case is treated in the following theorem in an asymptotic form.

Theorem 5.2 *Let $n(k)$ be a function of k and suppose that*

$$\lim_{k\to\infty} \frac{k^2}{n(k)} = \lambda.$$

Then

$$\left.\begin{array}{c}\frac{2}{3}(\lambda+1) \text{ if } 0<\lambda\leq\frac{1}{2}\\ 1 \text{ if } \frac{1}{2}<\lambda<1\end{array}\right\}\leq\lim_{k\to\infty}\frac{f(n(k),k)}{\frac{n(k)}{k}}\leq\sqrt{2}.$$

5.2 Proofs

We start with a well-known lemma from number theory, that we prove for the sake of completeness.

Lemma 5.3 *Let p be a prime. If both $a+c\equiv b+d \pmod p$ and $ac\equiv bd \pmod p$ hold then the element sets $\{a,c\}$ and $\{b,d\}$ are equal $\pmod p$.*

Proof. The case $p=2$ is trivial therefore we can assume that p is odd. Suppose first that $a+c\equiv b+d\equiv 0 \pmod p$. Then $c\equiv -a \pmod p$ and $d\equiv -b \pmod p$ hold and the second condition becomes $-a^2\equiv -b^2 \pmod p$ which implies $b\equiv \pm a \pmod p$. Then the first condition gives $c\equiv -a \pmod p, d\equiv \mp a\equiv -b \pmod p$, proving the statement for this case.

The general case can be reduced to the above one. Let $e\equiv a+c \pmod p$, and subtract $e/2$ from all four numbers. (Here $e/2$ denotes the unique integer whose double is congruent to $e \pmod p$.) Then

$$a-\frac{e}{2}+c-\frac{e}{2}\equiv b-\frac{e}{2}+d-\frac{e}{2} \pmod p$$

$$(a-\frac{e}{2})(c-\frac{e}{2})\equiv(b-\frac{e}{2})(d-\frac{e}{2})\equiv ac-(a+c)\frac{e}{2}+\left(\frac{e}{2}\right)^2\equiv bd-(b+d)\frac{e}{2}+\left(\frac{e}{2}\right)^2$$

are consequences of the conditions, the modified numbers satisfy the conditions of the first case, therefore the pairs $\left(a-\frac{e}{2}, c-\frac{e}{2}\right)$ and $\left(b-\frac{e}{2}, d-\frac{e}{2}\right)$ (mod p) are the same. □

The following lemma gives our main construction.

Lemma 5.4 *Let p be an odd prime. Then there exists a $(p(p-1), p+1, p-2)$-pamily.*

Proof. *The construction.* Let the underlying set X be the set of ordered pairs (i, j) where i and j are integers mod p ($0 \leq i, j \leq p-1$) and $i \not\equiv -j^2$ (mod p). Then $|X| = p^2 - p = p(p-1)$ is obvious. First we define the sets $Y(a, b) \subset X$ ($1 \leq a, b \leq p+1$). Start with the case when one of a or b is $p+1$. Then $Y(p-j, p+1) = Y(p+1, p-j)$ ($0 \leq j \leq p-1$) contains all the pairs $(i, j) \in X$ ($0 \leq i \leq p-1$). $Y(p+1, p+1) = \emptyset$. Suppose now $1 \leq a, b \leq p$ and define $Y(a, b)$. The element $(i, j) \in X$ is in $Y(a, b)$ if $j \neq p-a, j \neq p-b$ and

$$ab + aj + bj \equiv i \pmod{p} \tag{5.2}$$

holds. Since (5.2) is symmetric in a and b, we have $Y(b, a) = Y(a, b)$.

Let us show that $1 \leq a, b_1, b_2 \leq p, a \neq b_1, a \neq b_2, b_1 \neq b_2$ implies $Y(a, b_1) \cap Y(a, b_2) = \emptyset$. Otherwise they have a common element $(i, j) \in X$ where $j \neq p-a, p-b_1, p-b_2$ and (5.2) holds for both b_1 and b_2. However $j \neq p-a$ implies that b is uniquely determined by (5.2), the contradiction verifies the statement.

This statement is also true when $b_2 = p + 1$, since $Y(a, b_1)$ contains no element (i, j) with $j = p - a$ while the elements of $Y(a, p + 1)$ contain only such elements. Finally, $Y(p + 1, b_1) \cap Y(p + 1, b_2) = \emptyset$ is trivial.

The classes of the ath partition are the sets $Y(a, b)$ where $1 \leq b \leq p + 1, b \neq a$, which are disjoint. We only need to prove that they cover X. This is trivial when $a = p+1$, so $a \leq p$ can be supposed. Then the elements $(i, p-a)$ are covered by $Y(a, p+1)$, therefore it is sufficient to consider the pairs (i, j) where $j \neq p - a$. For given a, i, j (5.2) has a (unique) solution b and then $(i, j) \in Y(a, b)$ holds, proving that the classes $Y(a, b)$ $(1 \leq b \leq p + 1, b \neq a)$ really determine a partition. (Here the solution for b may be equal to a, this is why $(Y(a, a)$ is also needed.)

Proof of the fact that this pamily satisfies (iii). Consider two partitions, the ath and bth ones $(a \neq b)$. It should be verified that there are two classes in these partitions having intersection of size at least $p - 2$. The set $Y(a, b)$ is a class in both the ath and the bth partitions. We will show that its size is at least $p - 2$.

Let $1 \leq a < b \leq p$. We have to show that X has at least $p - 2$ pairs $(i, j) \in Y(a, b)$. These elements are defined by (5.2). j cannot be $p - a$ and $p - b$, all other $p - 2$ values can be chosen. However, for given a, b, j (5.2) uniquely determines an i. By the definition of $Y(a, b)$ this is equivalent to the statement that there are $p - 2$ solutions of (5.2) in X for any fixed a, b. This is obvious, since j can be chosen in $p - 2$ ways avoiding the cases $j = p - a, p - b$, and (5.2) uniquely determines i. The only problem could be

that the so obtained i satisfies $i \not\equiv -j^2 \pmod{p}$ and the solution (i, j) is not in X. However this and (5.2) would imply $(a+j)(b+j) = ab+aj+bj+j^2 \equiv 0 \pmod{p}$ contradicting the assumptions $a + j \not\equiv 0 \pmod{p}$ and $b + j \not\equiv 0 \pmod{p}$.

Compare now the ath ($1 \leq a \leq p$) and the $(p+1)$st partitions. $Y(a, p+1)$ is a class in both partitions, we need to show $|Y(a, p + 1)| \geq p - 2$. It is s actually equal to $p - 1$ since this is the number of elements $(i, p - a)$ where $i \not\equiv -(p - a)^2 \pmod{p}$.

Proof of the fact that this pamily satisfies (iv). Suppose $1 \leq a < b \leq p$ and choose a pair of classes from the ath and bth partitions, respectively, with intersection of size at least 2, that is, $|Y(a, c) \cap Y(b, d)| \geq 2$ holds for some $1 \leq c, d \leq p + 1$. Suppose temporarily that $c, d \leq p$. Let (i, j) and (u, v) be two distinct elements of the intersection. We intend to prove that $c = b, d = a$ follows.

By the definition of $Y(a, c)$ and $Y(b, d)$ the following congruencies must hold.

$$ac + aj + cj \equiv i \pmod{p} \tag{5.3}$$

$$ac + av + cv \equiv u \pmod{p} \tag{5.4}$$

$$bd + bj + dj \equiv i \pmod{p} \tag{5.5}$$

$$bd + bv + dv \equiv u \pmod{p}. \tag{5.6}$$

If $v \equiv j \pmod{p}$ then $(i, j) = (u, v)$ by (5.3) and (5.4), therefore

$$v \neq j \tag{5.7}$$

can be supposed. The differences (5.3)-(5.5) and (5.4)-(5.6) are

$$ac - bd + j((a + c) - (b + d)) \equiv 0 \pmod{p} \tag{5.8}$$

and

$$ac - bd + v((a + c) - (b + d)) \equiv 0 \pmod{p}.$$

The difference of these two congruencies is

$$(j - v)((a + c) - (b + d)) \equiv 0 \pmod{p}.$$

Hence we have

$$a + c \equiv b + d \pmod{p} \tag{5.9}$$

by (5.7). Moreover, (5.8) leads to

$$ac \equiv bd \pmod{p}. \tag{5.10}$$

Lemma 5.3, (5.9) and (5.10) give $\{a, c\} = \{b, d\}$, that is, $c = b, d = a$.

The same conclusion is needed when one or more of a, b, c, d is equal to $p + 1$.

Let $c = p + 1, a, b, d \leq p$ and suppose that $(i, j), (u, v) \in Y(a, p + 1) \cap Y(b, d)$. Then $j = v = p - a$. (5.5) and (5.6) make $i = u$, that is, (i, j) and (u, v) are the same, the contradiction settles this case.

Let $c = d = p + 1, a, b \leq p$. The contradiction is obtained trivially: $i = p - a = p - b$.

Let $a, c \leq p, b = p + 1$ hold. Then $d \leq p$ is a consequence. $j = v = p - d$ follows. (5.3) and (5.4) imply $i = u$, the same contradiction is obtained.

If $a \leq p, c = b = p+1$ then d must be equal to a and the desired $c = b, d = a$ is obtained.

We have seen that if the intersection of a class of the ath and a class of the bth partition ($a \neq b$) has at least two common elements then these classes are $Y(a,b) = Y(b,a)$. This makes it impossible to have 3 distinct partitions containing classes with intersection of size at least 2. □

Summarizing what we obtained: $Y(a,b) = Y(b,a)$ is the only pair of classes in the ath and bth ($a \neq b$) partitions with more then two common elements and $|Y(a,b)| \geq p-2$.

Lemma 5.5 *(a) Let p be an odd prime, $r (0 \leq r \leq \frac{p-1}{2})$ an integer. Then there exists a $(p(p-1-r), p+1, p-2-r)$-pamily.*

(b) Let $p > 5$ be an odd prime. There exists a $(\frac{(p-1)(p-2)}{2}, p+1, \frac{p-5}{2})$-pamily.

Proof. *Proof of (a).* Let $X_i = \{(i,j) : 0 \leq j \leq p-1, i+j^2 \not\equiv 0\}$. Suppose that i is not 0 and $-i$ is not a quadratic residue mod p. Then $|X_i| = p$. Let us see that

$$|X_i \cap Y(a,b)| \leq 1 \tag{5.11}$$

holds in this case. If $1 \leq a, b \leq p, a+b \not\equiv 0 \pmod{p}$ then this follows from the fact that (5.2) has exactly one solution in j when a, b, i are given. On the other hand, if $a+b \equiv 0 \pmod{p}$ then (5.2) becomes $a(-a) \equiv i \pmod{p}$ contradicting the condition that $-i$ is not a quadratic residue. Finally, if $1 \leq a \leq p, b = p+1$ then the label of $Y(a, p+1)$ is determined by j, namely

$a = p - j$, that is, $X_i \cap Y(a, p+1) = \{(i, p-a)\}$, (5.11) holds for this case, too.

As a consequence, the sizes of the non-one-element intersections of the restrictions of the classes on $X - X_i$ are decreased by at most one, they are at least $p - 3$. Deleting r such X_i from X, the intersection sizes will be decreased by at most r.

Proof of (b). The statement of (a) with $r = \frac{p-1}{2}$ results in the existence of a $\left(\frac{p(p-1)}{2}, p+1, \frac{p-3}{2}\right)$-pamily. It is constructed on the set $X' = X - \bigcup X_i$ where i runs on the set of indexes for which $i \neq 0$ and $-i$ is not a quadratic residue. Now X_0 will be deleted. $|X_0| = p - 1$ since $j = 0$ is excluded.

Let us show now that

$$|X_0 \cap Y(a, b)| \leq 1 \tag{5.12}$$

holds. If $1 \leq a, b \leq p, a + b \not\equiv 0 \pmod{p}$ then the reasoning is the same as in the case of (a). However, when $a + b \equiv 0 \pmod{p}$ then $i = 0$ and (5.2) imply $a = b = 0$, contradicting the assumption $a \neq b$. The subcase $1 \leq a \leq p, b = p + 1$ behaves exactly like at (a).

Deleting X_0 from X', the so obtained $X'' = X' - X_0$ has size $\frac{(p-1)(p-2)}{2}$. The restrictions of the $p+1$ partitions on X'' have the property that for any two of these restricted partitions one can find one class in both of them with an intersection of size $\frac{p-5}{2}$. □

Lemma 5.6 *Let p be an odd prime, $s (0 \leq s \leq \frac{p-9}{2})$ an integer. Then there exists a $(\frac{(p-1)(p-2)}{2} - s\frac{p+1}{2} - s^2, p+1, \frac{p-5}{2} - s)$-pamily.*

Proof. Illustrate the proof for $s = 1$. Suppose that $i \neq 0$ and $-i$ is a quadratic residue. Then two different j satisfies $i + j^2 \equiv 0 \pmod{p}$ therefore $|X_i| = p - 2$. The inequality (5.11) and its proof work in the same way as in the case of the previous lemma except when $a + b \equiv 0 \pmod{p}$ holds. Then (5.2) becomes $ab \equiv i \pmod{p}$ and $-a^2 \equiv i$ has a pair of solutions, a and $-a$. (5.11) holds with one exception: $\{a, -a\}$.

Therefore deleting X_i from X'' and restricting the permutations on the remaining underlying set, the pairwise intersection of the non-one-element classes of the ath and bth permutations will be decreased only by one except for the case when a satisfies $a^2 \equiv -i \pmod{p}$ and $b = p - a$. In order to preserve the desired intersection property we add a (disjoint) set Z_i' to $X'' - X_i$ and extend the partitions on it. Namely all partitions will have one-element classes on Z_i', except the ath and bth partitions. They have only one class: the whole Z_i'. Its size must be $\frac{p-7}{2}$. The size of the new underlying set is $\frac{(p-1)(p-2)}{2} - (p-2) + \frac{p-7}{2} = \frac{p^2-4p-1}{2}$. We have obtained a $(\frac{p^2-4p-1}{2}, p+1, \frac{p-7}{2})$-pamily.

The same can be done for larger s. the only difference is that the sets Z can be smaller, since the intersections between the partitions became smaller. Take s (non-zero) indexes i ($-i$ is a quadratic residue). Delete all of them from X'' and add the disjoint (to X'' and each other) sets Z_i^s of size $\frac{p-5}{2} - s$. The size of the new underlying set $(X'' \cup_i X_i) - \cup_i Z_i^s$ is

$$\frac{(p-1)(p-2)}{2} - s(p-2) + s\left(\frac{p-5}{2} - s\right) = \frac{(p-1)(p-2)}{2} - s\frac{p+1}{2} - s^2.$$

The ath partition is extended by one-element classes on Z_i^s unless $a^2 + i \equiv 0 \pmod{p}$ holds, then the extension has one class: the whole Z_i^s. It is easy to see that this pamily satisfies the conditions. □

We now recall once more the following theorems concerning the density of prime numbers which will be needed in the proofs.

Theorem 5.7 *Let $0 < \varepsilon$ be a positive real number. If $R > R(\varepsilon)$ then there is a prime number between R and $R(1 + \varepsilon)$.*

Theorem 5.8 (Chebyshev Theorem) *If $R \geq 2$ is a real number then there is a prime number between R and $2R$.*

Proof of Theorem 5.1

Case 1. Suppose $(k + 1)(2k + 3) \leq n$. Lemma 5.5 (a) will be used with $k = p - 2 - r$ where p and r have to be properly found. By the Chebyshev theorem there is a prime number p satisfying $k + 2 \leq p \leq 2k + 3$. Choose $0 \leq r = p - (k + 2)$. Here $\frac{p+1}{2} \leq k + 2$ implies $r = p - (k + 2) \leq p - \frac{p+1}{2} = \frac{p-1}{2}$, therefore r satisfies the condition in Lemma 5.5 (a). The pamily is constructed on a set of $p(p-1-r)$ elements. Here $p(p-1-r) \leq (2k+3)(k+1)$ which is at most n by the condition of the case. The construction of Lemma 5.5 can be placed in $[n]$. Consequently we have $p + 1 \geq k + 3$ partitions in the pamily.

Case 2. Suppose $5k^2 + \frac{37}{2}k + 16 \leq n$. Then Lemma 5.6 will be used with $k = \frac{p-5}{2} - s$ with properly chosen p and s. Since s is non-negative, p must satisfy $2k+5 \leq p$. The condition $s \leq \frac{p-9}{2}$ of the lemma holds by $2 \leq k$. If we

want to use the lemma with the underlying set of size $\frac{(p-1)(p-2)}{2} - s\frac{p+1}{2} - s^2$ then this number cannot exceed n. Replacing s by $\frac{p-5}{2} - k$, the following inequality is needed:

$$\frac{(p-1)(p-2)}{2} - s\frac{p+1}{2} - s^2 = p\left(\frac{3k}{2} + 2\right) - k^2 - \frac{9}{2}k - 4 \leq n.$$

Hence we must have

$$p \leq \frac{n + k^2 + \frac{9}{2}k + 4}{\frac{3k}{2} + 2}.$$

By the Chebyshev inequality there is a prime p satisfying

$$\frac{n + k^2 + \frac{9}{2}k + 4}{3k + 4} \leq p \leq \frac{n + k^2 + \frac{9}{2}k + 4}{\frac{3k}{2} + 2}.$$

The lower bound is in accordance with the condition $2k + 5 \leq p$ when

$$2k + 5 \leq \frac{n + k^2 + \frac{9}{2}k + 4}{3k + 4},$$

but this is equivalent to the condition of the case. □

Proof of Theorem 5.2

The proof of Theorem 5.1 is copied, but Theorem 5.7 is used rather than Theorem 5.8. We show the main ideas only, the details with ε are left to the reader. n, p, r, s are functions of k, but we will not indicate it in the notation.

Case 1. Suppose $\frac{1}{2} < \lambda < 1$. Lemma 5.5 (a) will be used with $k = p-2-r$ where p has to be properly found. The size of the underlying set of the construction is $p(k + 1)$ cannot be larger than n, that is, $p \leq \frac{n}{k+1}$ must hold. By Theorem 5.7 there is a prime number p satisfying this condition

and $p \sim \frac{n}{k}$. Then $r = k + 2 - p \sim (1-\lambda)\frac{n}{k}$. The conditions $0 \leq r \leq \frac{p-1}{2}$ are satisfied by $\frac{1}{2} < \lambda < 1$. Consequently, the construction of Lemma 5.5 can be placed in $[n]$, we have $p + 1 \sim \frac{n}{k}$ partitions in the pamily.

Case 2. Suppose $0 < \lambda < \frac{1}{2}$. Lemma 5.6 will be used with $k = p - 2 - r$ where p has to be properly found. Let p be a prime such that $p \sim k\frac{2(1+\lambda)}{3\lambda}$. It exists by Theorem 5.7. Then one can find a non-negative integer s such that $s \sim k\frac{1-2\lambda}{3\lambda}$. Since $0 < \lambda \leq \frac{1}{2}$, we have $0 < s < \frac{p-1}{2}$. Then

$$s\left(\frac{p+1}{2} + s\right) \sim k^2 \frac{(1-2\lambda)(2-\lambda)}{9\lambda^2}$$

and

$$\frac{(p-1)(p-2)}{2} - s\frac{p+1}{2} - s^2 \sim k^2 \frac{1}{\lambda}.$$

Use here $k \sim \lambda\frac{n}{k}$ in one of the ks obtaining that the size of the underlying set in the construction is $\sim k\lambda\frac{n}{k}\frac{1}{\lambda} = n$. □

Acknowledgement I am deeply indebted to my supervisor Gyula O.H. Katona for his constant guidance and support during my research.

References

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

[2] H. Davenport, *Multiplicative number theory*, Graduate Text in Mathematics, 3rd edition, Springer, 2000.

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

[4] J. Demetrovics and G.O.H. Katona, Extremal combinatorial problems in relational data base, *Fundamentals of Computation Theory 81, Proc. of the 1981 International FCT-Conference, Szeged, Hungary, 1981, Lecture Notes in Computer Science,* **117** (Springer Verlag, Berlin, 1981) 110-119.

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

[6] P. Erdős and H. Hanani, On a limit theorem in combinatorial analysis, *Publ. Math. Debrecen* **10**(1963) 10-13.

[7] P. Erdős and J. Surányi, *Topics in the Theory of Numbers*, Springer, 2003, for Hungarian readers: *Válogatott fejezetek a számelméletből*, 2-ik kiadás, Polygon, Szeged, 2004.

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

[9] P.M. Gergely, k-elemű halmazok aszimptotikus pakolása, ahol nincs kétszeresen fedett pár, *Matematikai Lapok* **3** (2010), 44-50.

[10] P.M. Gergely, Partitions with certain intersection properties, accepted for publication by *J. of Combinatorial Designs*

[11] P.M. Gergely and G.O.H. Katona, Between relational database models and secret sharing, in preparation

[12] H.-D.O.F. Gronau, M. Grüttmüller, S. Hartmann, U. Leck and V. Leck, On orthogonal double covers of graphs, *Des. Codes Cryptogr.* **27** (2002), 49-91.

[13] 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.

[14] H.-D.O.F. Gronau, R.C. Mullin and A. Rosa, Orthogonal double covers of complete graphs by trees, *Graphs Combin.* **13**(1997) 251-262.

[15] H.-D.O.F. Gronau, R.C. Mullin, A. Rosa and P.J. Schellenberg, Symmetric graph designs, *Graphs Combin.* **16**(2000) 93-102.

[16] H. Hanani, The existence and construction of balanced incomplete block designs, *Ann. Math. Statist.* **32**(1961) 361-386.

[17] H. Hanani, On resolvable balanced incomplete block designs, *J. Combin. Theory (A)* **17**(1974) 275-289.

[18] S. Hartmann, U. Leck and V. Leck, A conjecture on orthogonal double covers by paths. Proceedings of the Thirtieth Southeastern International Conference on Combinatorics, Graph Theory, and Computing (Boca Raton, FL, 1999). *Congr. Numer.* **140**(1999) 187-193.

[19] U. Leck and V. Leck, On orthogonal double covers by trees, *J. Combin. Des.* **5**(1997) 433-441.

[20] U. Leck and V. Leck, Orthogonal double covers of complete graphs by trees of small diameter. Proceedings of the Conference on Optimal Discrete Structures and Algorithms—ODSA '97 (Rostock). *Discrete Appl. Math.* **95**(1999), 377–388.

[21] A. Shamir, How to share a secret, *Communications of the ACM* **22**(1979)) 612613.

[22] R. Wilson, An existence theory for pairwise balanced designs, **13**(1972) 220-245.