A new type of coding problem

Graham Brightwell

Department of Mathematics, LSE
Houghton St., London WC2A 2AE, UK.
G.R.Brightwell@lse.ac.uk

Gyula O.H. Katona*

Rényi Institute
Budapest, Reáltanoda u. 13–15, 1053
Hungary
ohkatona@renyi.hu

Keywords: graph, code, distance, design, hamiltonian cycle

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 $d(\{A_1, B_1\}, \{A_2, B_2\}) = \min\{|A_1 - A_2| + |B_1 - B_2|, |A_1 - B_2| + |B_1 - A_2|\}$. It is known ([2]) that this is really a distance on the space of such pairs and that the family of all k–element subsets of X can be paired (with one exception if their number is odd) in such a way that the distance of the pairs is at least k. Here we answer questions arising for distances larger than k.

1 Introduction

Let X be a finite set of n elements, $1 < k < n$ an integer. Unordered disjoint pairs $\{A, B\}$ of k-element sets (that is, $|A| = |B| = k, A \cap B = \emptyset$) will be considered. Define the *distance*

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

between two such pairs. It has been verified in [2] that it is really a distance, that is, it satisfies the triangle inequality. We say that a set $\mathcal{C}$ of such pairs 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

$$k\text{–element subset may occur only once in the pairs } \{A, B\} \in \mathcal{C} \text{ as } A \text{ or } B. \qquad (1)$$

The following theorem was proved in [2].

*Research was supported by the Hungarian National Foundation for Scientific Research, grant number T029255.

Theorem 1

$$C'(n,k,k) = \left\lfloor \frac{1}{2}\binom{n}{k} \right\rfloor.$$

It is obvious that one cannot choose more pairs using any k-element set at most once, so the theorem actually states that this many pairs can be constructed with pairwise distance k and satisfying (1). Theorem 1 is a sharpening of a theorem of [1] where $\lfloor \frac{1}{2}\binom{n}{k} \rfloor$ pairs were constructed under the condition that

$$\max\{|A_1 - A_2|, |B_1 - B_2|\}, \max\{|A_1 - B_2|, |B_1 - A_2|\} \geq \frac{k}{2}.$$

The method of the proofs of the constructions uses Hamiltonian type theorems.

It is quite natural to ask if one can choose $\lfloor \frac{1}{2}\binom{n}{k} \rfloor$ pairs with pairwise difference at least $k+1$. The answer is negative. In Section 2 we will give an upper estimate on $C(n,k,d)$ which will be less than $\lfloor \frac{1}{2}\binom{n}{k} \rfloor$ for $k < d$. Section 3 contains lower estimates on $C(n,k,d)$.

2 An upper estimate

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

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

holds.

Proof: Let $\mathcal{C}$ be a family of pairs of disjoint k-element subsets of X such that $d(C,C') \geq d$ for all $C, C' \in \mathcal{C}$ and count the number of pairs (C,D) where $C = \{A,B\} \in \mathcal{C}$, D is a $k - \lfloor\frac{d}{2}\rfloor$-element subset of X and D is a subset of one of either A or B.

First, let us fix a $C = \{A,B\} \in \mathcal{C}$. There are exactly

$$2\binom{k}{k-\lfloor\frac{d}{2}\rfloor} = 2\binom{k}{\lfloor\frac{d}{2}\rfloor}$$

appropriate Ds, therefore the total number of counted pairs (C,D) is

$$|\mathcal{C}|2\binom{k}{\lfloor\frac{d}{2}\rfloor}. \tag{2}$$

On the other hand, if D is fixed then suppose that $C_1 = \{A_1, B_1\}, C_2 = \{A_2, B_2\} \in \mathcal{C}$ and $D \subset A_1, A_2$. Since $|A_1 - A_2| \leq \lfloor\frac{d}{2}\rfloor$ therefore $|B_1 - B_2|$ must be at least $\lceil\frac{d}{2}\rceil$, that is, $|B_1 \cap B_2| \leq k - \lceil\frac{d}{2}\rceil$. Consequently the possible Bs are subsets of the $n - k + \lfloor\frac{d}{2}\rfloor$-element $X - D$ and they cannot cover the same $k - \lceil\frac{d}{2}\rceil + 1$-element set. Hence the number of possible Bs is at most

$$\frac{\binom{n-k+\lfloor\frac{d}{2}\rfloor}{k-\lceil\frac{d}{2}\rceil+1}}{\binom{k}{k-\lceil\frac{d}{2}\rceil+1}}.$$

The total number of pairs (C,D) cannot exceed

$$\binom{n}{k-\lfloor\frac{d}{2}\rfloor}\frac{\binom{n-k+\lfloor\frac{d}{2}\rfloor}{k-\lceil\frac{d}{2}\rceil+1}}{\binom{k}{k-\lceil\frac{d}{2}\rceil+1}}. \tag{3}$$

(2) $\leq$ (3) leads to Theorem 2 by appropriate cancellations. $\square$

Corollary 3 *If* $2 \leq k \leq n/2$ *then*

$$C(n,k,k+1) < \lfloor\frac{1}{2}\binom{n}{k}\rfloor.$$

Proof: Using Theorem 2 it is sufficient to prove

$$\frac{1}{2}\frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots\lceil\frac{k+2}{2}\rceil\cdot k(k-1)\cdots\lfloor\frac{k+2}{2}\rfloor} < \frac{1}{2}\frac{n(n-1)\cdots(n-k+1)}{k!} - \frac{1}{2}. \tag{4}$$

It will be proved in the form

$$1 < \frac{n(n-1)\cdots(n-k+1)}{k!}\left(1-\frac{\lfloor\frac{k}{2}\rfloor!}{k(k-1)\cdots(\lceil\frac{k}{2}\rceil+1)}\right). \tag{5}$$

Observe that

$$2 \leq \frac{n}{k} < \frac{n-1}{k-1} < \ldots < \frac{n-k+1}{1}$$

and

$$\frac{1}{2} \geq \frac{\lfloor\frac{k}{2}\rfloor}{k} > \frac{\lfloor\frac{k}{2}\rfloor-1}{k-1} > \ldots > \frac{1}{\lceil\frac{k}{2}\rceil+1}.$$

Using these inequalities in (5) we arrive to the stronger inequality

$$1 < 2^k\left(1-\frac{1}{2^{\lfloor\frac{k}{2}\rfloor}}\right)$$

which is trivially true for $2 \leq k$. (5), (4) and the corollary are proved. $\square$

3 Lower estimates

Let $1 < v < u < n$ be integers. The family $\mathcal{P}$ is called an (n,u,v) *packing family* if it consists of u-element subsets of an n-element underlying set X and every v-element subset of X is contained in at most one member of $\mathcal{P}$. The class of all (n,u,v) packing families is denoted by $\mathbb{P}(n,u,v)$. Introduce the notation

$$m(n,u,v) = \max\{|\mathcal{P}| : \mathcal{P} \in \mathbb{P}(n,u,v)\}.$$

The inequality

$$m(n,u,v) \leq \frac{\binom{n}{v}}{\binom{u}{v}} \tag{1}$$

is obvious for a (n, u, v) packing family $\mathcal{P}$ (1) holds with equality iff every v-element subset is contained in exactly one member of the family $\mathcal{P}$. In this case $\mathcal{P}$ is called an (n, u, v) *Steiner family.*

The celebrated theorem of Rödl [5] (see also [3]) states that there are families asymtotically achieving the upper estimate (1), that is

$$\frac{m(n,u,v)\binom{u}{v}}{\binom{n}{v}} \to 1$$

for fixed u, v when n tends to infinity.

Proposition 4 *Let $d \le 2k \le n$ be integers. Then*

$$C(2k,k,d)\frac{n(n-1)\cdots(n-2k+d)}{(2k)(2k-1)\cdots d}(1+o(1)) \le C(n,k,d) \tag{3}$$

holds, where $o(1)$ may depend on k and d.

Proof: Take a family $\mathcal{P} \in \mathbb{P}(n, 2k, 2k-d+1)$ with size

$$|\mathcal{P}| = (1+o(1))\frac{\binom{n}{2k-d+1}}{\binom{2k}{2k-d+1}} = (1+o(1))\frac{n(n-1)\cdots(n-2k+d)}{(2k)(2k-1)\cdots d}$$

which exists by [5]. Let A_1, B_1 and A_2, B_2 be partitions (into k-element sets) of two different members of $\mathcal{P}$, that is, $A_1 \cap B_1 = A_2 \cap B_2 = \emptyset, |A_1| = |A_2| = |B_1| = |B_2| = k$ and $A_1 \cup B_1, A_2 \cup B_2$ are in $\mathcal{P}$. Their intersection has at most $2k-d$ elements, hence we have

$$|A_1 \cap A_2| + |B_1 \cap B_2|, |A_1 \cap B_2| + |B_1 \cap A_2| \le |(A_1 \cup B_1) \cap (A_2 \cup B_2)| \le 2k-d.$$

This implies

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

$$\min\{k - |A_1 \cap A_2| + k - |B_1 \cap B_2|, k - |A_1 \cap B_2| + k - |B_1 \cap A_2|\} \ge d.$$

Take the maximum number $C(2k, k, d)$ of such partitions with distance at least d in each member of $\mathcal{P}$. This construction proves (3). □

Now we give a lower estimate on $C(2k, k, d)$ for some cases. The method is a modification of the method used by Sloane and Graham [4] proving lower bounds for constant weight codes.

Let us first consider the simplest case of $C(2k, k, 3) = C(2k, k, 4)$.

Theorem 5

$$|\mathcal{N}| \le C(2k,k,3)$$

where $\mathcal{N}$ is the family of all k-element subsets A of $X = \{1, 2, \dots, 2k\}$ such that

$$\sum_{i \in A} i \equiv 0 \pmod{2k+1}.$$

PROOF: Since $1+2+\ldots+2k = k(2k+1) \equiv 0 \pmod{2k+1}$ holds, $A \in \mathcal{N}$ implies $X - A \in \mathcal{N}$, too. $\mathcal{N}$ consist of complementing pairs of k-element subsets of X.

Suppose that $A, B \in \mathcal{N}, |A \cap B| = k - 1$ holds.

$$\sum_{i \in A} i \equiv \sum_{i \in B} i \pmod{2k+1}$$

implies

$$\sum_{i \in A-B} i \equiv \sum_{i \in B-A} i \pmod{2k+1}.$$

Here $A - B$ and $B - A$ are 1-element sets, therefore they must be equal. Hence $A = B$, that is two different members of $\mathcal{N}$ cannot have $k-1$ common elements. They cannot have exactly one common element either, since this would imply that A and $X - B \in \mathcal{N}$ have $k-1$ common elements, a contradiction. $\square$

It seems that $|\mathcal{N}|$ cannot be much smaller than

$$\frac{1}{(2k+1)}\binom{2k}{k}.$$

We are quite sure that this is known, but we were unable to find the appropriate reference.

Suppose now that $q = 2k+1$ is a prime power. We can prove an analogous lower bound for $C(2k, k, d)$ only in this case. Let $X = \{\omega_1, \ldots, \omega_{q-1}\}$ be the set of all non-zero elements of the finite field $GF(q)$. Let $d = 2\delta$ and define $\mathcal{N}_0(k, \delta)$ as the family of all k-element subsets A of X such that

$$\sum_{i_1 < \ldots < i_\rho \in A} \omega_{i_1} \cdots \omega_{i_\rho} = 0 \tag{4}$$

holds for every integer $1 \leq \rho < \delta$.

Let us see that $A \in \mathcal{N}_0(k, \delta)$ implies the same for $X - A$. Introduce the notation

$$s(B, u, v) = \sum \omega_j^u \omega_{i_1} \cdots \omega_{i_v}$$

for all $B \subset \{1, \ldots, q-1\}, 0 \leq u, 0 \leq v < |B|$ where the sum is taken for all $v+1$ different elements $j, i_1 < \ldots < i_v$ of B. It is obvious that $s(B, 0, v)$ is $(|B| - v)$ times the sum of all products of v distinct ω_is with indeces from B. On the other hand $s(B, 1, v) = \frac{(v+1)}{|B|-v} s(B, 0, v)$ holds and $s(B, u, 0)$ is the sum of the uth powers of ω_is with indices from B.

$$s(B, u, 0)\frac{s(B, 0, v)}{|B| - v} = s(B, u, v) + s(B, u+1, v-1) \ (1 \leq u, 1 \leq v < |B|) \tag{5}$$

is obviously true.

Let ε be a primitive root of the field. Then

$$s(X, u, 0) = \varepsilon^{0u} + \varepsilon^{1u} + \varepsilon^{2u} + \ldots + \varepsilon^{(q-1)u} = \frac{\varepsilon^{qu} - 1}{\varepsilon^u - 1} = 0 \tag{6}$$

holds for $1 \leq u < q$.

(5) will be applied for $B = A$ several times. Start with the case $u = 1, v = \delta - 2$:

$$s(A, 1, 0)\frac{s(A, 0, \delta-2)}{|A| - (\delta - 2)} = s(A, 1, \delta - 2) + s(A, 2, \delta - 3).$$

Here $s(A, 0, \delta - 2)$ and $s(A, 1, \delta - 2)$ are zero by (4). Consequently $s(A, 2, \delta - 3) = 0$ also holds. Applying (5) with $u = 2, v = \delta - 3$ and using $s(A, 2, \delta - 3) = 0$ the equality $s(A, 3, \delta - 4) = 0$ is obtained. Continuing this procedure we arrive to $s(A, \delta - 1, 0) = 0$. The equations $s(A, u, 0) = 0$ can be obtained in the same way for $1 \leq u \leq \delta - 1$. In other words,

$$\sum_{i \in A} \omega_i^u = 0 \quad (1 \leq u \leq \delta - 1) \tag{7}$$

holds. (6) and (7) imply that

$$s(X - A, u, 0) = s(X, u, 0) - s(A, u, 0) = \sum_{i \in X - A} \omega_i^u = 0$$

also holds for $1 \leq u \leq \delta - 1$. If the previous method is applied backwards for $X - A$, then it leads to the validity of (4) for $X - A$, proving that it is really in $\mathcal{N}_0(k, \delta)$.

We will now see that the symmetric difference of any two members A, B of $\mathcal{N}_0(k, \delta)$ is at least 2δ. Otherwise $A - B = \{r_1, \dots, r_\gamma\}, B - A = \{s_1, \dots, s_\gamma\}$ hold where $\gamma \leq \delta - 1$. Introduce the shorter notations $\alpha_i = \omega_{r_i}, \beta_i = \omega_{s_i}$. It is easy to see (see [4]) that the defining conditions (4) imply the equations

$$\sigma_1 = \sum_i \alpha_i = \sum_i \beta_i,$$

$$\sigma_2 = \sum_{i<j} \alpha_i \alpha_j = \sum_{i<j} \beta_i \beta_j,$$

$$\dots$$

$$\sigma_{\delta-1} = \sum_{i_1 < \dots < i_{\delta-1}} \alpha_{i_1} \cdots \alpha_{i_{\delta-1}} = \sum_{i_1 < \dots < i_{\delta-1}} \beta_{i_1} \cdots \beta_{i_{\delta-1}}.$$

That is, the elementary symmetric functions of the α_is and the β_is agree, therefore $\alpha_1, \dots, \alpha_\gamma, \beta_1 \dots \beta_\gamma$ are all zeros of the polynomial

$$x^\gamma - \sigma_1 x^{\gamma-1} + \sigma_2 x^{\gamma-2} - \dots (-1)^\gamma \sigma_\gamma$$

of order γ. This contradiction proves that the pairwise distance of A and B is at least $d = 2\delta$. Since the same holds for the complements, the complementary pairs of the members of $\mathcal{N}_0(k, \delta)$ are really in distance at least d. The following theorem is proved.

Theorem 6 *If $2k + 1$ is a prime power and $d = 2\delta$ then*

$$\frac{1}{2} |\mathcal{N}_0(k, \delta)| \leq C(2k, k, d)$$

holds.

The size of $\mathcal{N}_0(k, \delta)$ can be determined for small values, but we believe that it cannot be much less than

$$\frac{1}{q^{\delta-1}} \binom{2k}{k},$$

since the defining sums are probably nearly equally distributed among all the $q^{\delta-1}$ possibilities.

4 Open problems

Theorem 2 and Proposition 4 imply the following statement.

Corollary 7

$$c_1(k,d)n^{2k-d+1} \le C(n,k,d) \le c_2(k,d)n^{2k-d+1}.$$

However we think that the upper bound of Theorem 2 is asymptotically correct.

Conjecture 8

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

Actually we believe that, for an arbitrary pair of k and d, there are infinitely many ns with equality in Theorem 2.

The case $d = 1$ is uninteresting. If $d = 2$ then the upper and lower estimates coincide providing the $(n, 2k, 2k-1)$ Steiner family exists. Therefore the first unfinished case is $d = 3$. Even in the case of $k = 2$, the upper and lower estimates significantly differ. The upper estimate is

$$C(n,2,3) \le \frac{n(n-1)}{8}.$$

On the other hand $C(4, 2, 3)$ is obviously 1, therefore our construction gives only the lower bound

$$\frac{n(n-1)}{12}$$

when an $(n, 4, 2)$ Steiner family exists. Can one add $\frac{n(n-1)}{24}$ pairs of disjoint two-element sets (edges) to the Steiner system which preserves the condition that the pairwise distance of the pairs is at least 3?

References

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

[2] H. Enomoto and G.O.H. Katona, Pairs of disjoint q-element subsets far from each other, *Electronic Journal of Combinatorics* (1999) **52**

[3] P. Frankl and V. Rödl, Near perfect coverings in graphs and hypergraphs, *European J. of Combinatorics* (1985) **6** 317-326.

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

[5] V. Rödl, On a packing and covering problem, *European J. of Combinatorics* (1985) **6** 69-78.