Pairs of disjoint q-element subsets far from each other[*†]

Hikoe Enomoto
Department of Mathematics, Keio University
3-14-1 Hiyoshi, Kohoku-Ku, Yokohama, 223 Japan,
enomoto@math.keio.ac.jp

Gyula O.H. Katona
Alfréd Rényi Institute of Mathematics, HAS,
Budapest P.O.B. 127 H-1364 Hungary
ohkatona@renyi.hu

July 27, 2001

Abstract

Let n and q be given integers and X a finite set with n elements. The following theorem is proved for $n > n_0(q)$. The family of all q-element subsets of X can be partitioned into disjoint pairs (except possibly one if $\binom{n}{q}$ is odd), so that $|A_1 \cap A_2| + |B_1 \cap B_2| \le q$, $|A_1 \cap B_2| + |B_1 \cap A_2| \le q$ holds for any two such pairs $\{A_1, B_1\}$ and $\{A_2, B_2\}$. This is a sharpening of a theorem in [2]. It is also shown that this is a coding type problem, and several problems of similar nature are posed.

1 Introduction

The following theorem was proved in [2].

Theorem 1.1 *Let $|X| = n$ and $2k > q$. The family of all q-element subsets of X can be partitioned into unordered pairs (except possibly one if $\binom{n}{q}$ is odd), so that paired q-element subsets are disjoint and if A_1, B_1 and A_2, B_2 are two such pairs with $|A_1 \cap A_2| \ge k$, then $|B_1 \cap B_2| < k$, provided $n > n_0(k, q)$.*

[*] AMS Subject classification Primary 05B30, Secondary 05C45, 94B99. Keywords: design, Hamiltonian cycle, code.

[†] The work was supported by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (B), 10440032, the Hungarian National Foundation for Scientific Research grant numbers T029255, DIMACS and UVO-ROSTE 875.630.9

The main aim of the present paper is to give a sharpening of this theorem. Define the *closeness* of the pairs $\{A_1, B_1\}$ and $\{A_2, B_2\}$ by

$$\gamma(\{A_1, B_1\}, \{A_2, B_2\}) = \max\{|A_1 \cap A_2| + |B_1 \cap B_2|, |A_1 \cap B_2| + |B_1 \cap A_2|\} \qquad (1.1)$$

It is obvious that $|A_1 \cap A_2| \geq k$ and $|B_1 \cap B_2| \geq k$ imply $\gamma((A_1, B_1), (A_2, B_2)) \geq 2k$ for sets satisfying $A_1 \cap B_1 = A_2 \cap B_2 = \emptyset$, therefore the following theorem is really a sharpening of Theorem 1.1 .

Theorem 1.2 *Let $|X| = n$. The family of all q-element subsets of X can be partitioned into disjoint pairs (except possibly one if $\binom{n}{q}$ is odd), so that $\gamma(\{A_1, B_1\}, \{A_2, B_2\}) \leq q$ holds for any two such pairs $\{A_1, B_1\}$ and $\{A_2, B_2\}$, provided $n > n_0(q)$.*

The proof of Theorem 1.1 was based on a Hamiltonian type theorem. Here we will need another theorem of the same type. Two edge-disjoint (non-directed) simple graphs $G_0 = (V, E_0)$ and $G_1 = (V, E_1)$ will be given on the same vertex set where $|V| = N,\ E_0 \cap E_1 = \emptyset$. Let r denote the minimum degree in G_0. The edges of the second graph are labelled by positive integers. The label of $e \in E_1$ is denoted by $l(e)$. Denote the number of edges of label i starting from the vertex v by

$$d(v, i) = |\{e \in E_1 : v \in e, l(e) = i\}|.$$

Let s be the maximum degree in G_1, that is,

$$s = \max_{v \in V}\{\sum_i d(v, i)\}. \qquad (1.2)$$

Another parameter t is defined by

$$t = t(q) = \max_{v \in V}\{\sum_i d(v, i) \max_{w \in V}\{\sum_{q+1-i \leq j} d(w, j)\}\}. \qquad (1.3)$$

A 4-tuple (x, y, z, v) of vertices is called *heavy* C_4 iff $(x, y), (z, v) \in E_0,\ (y, z), (x, v) \in E_1, l((y, z)) + l((x, v)) \geq q + 1$. After these definitions we are able to formulate our theorem.

Theorem 1.3 *Suppose, that*

$$2\,r - 4\,t - s - 1 > N. \qquad (1.4)$$

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 a heavy C_4.

Section 2 contains the proofs. In Section 3 we will pose a general question to find the maximum number of elements whose paiwise distance is at least d in a finite "space" furnished with a "distance". Theorem 1.2 is the solution of this question in a special case.

2 Proofs

The proof of Theorem 1.3 is based on Dirac's famous theorem on a sufficient condition for existence of a Hamiltonian cycle and on Lemma 2.2.

Theorem 2.1 (Dirac [3]) *If G is a simple graph on N vertices and all degrees of G are at least $\frac{N}{2}$, then G has a Hamiltonian cycle.*

Lemma 2.2 *Let G_0, G_1, r, s, t and N satisfy (1.4). Assume that there is a Hamiltonian path from a to b in G_0. Then there exist c, $c \neq a$, and d, $d \neq b$, adjacent vertices along the path, such that c is between a and d on the path, $(a, d) \in E_0$, $(b, c) \in E_0$, (a, d, b, c) is not a heavy C_4, and if (x, y) is an edge of the path, then neither (a, d, x, y) nor (b, c, x, y) is a heavy C_4.*

Proof of Lemma 2.2

We call a vertex $x \in V$ *a-bad* (*b-bad*) if there exists an edge (y, z) of the Hamiltonian path such that (a, x, y, z) ((b, x, y, z), respectively) is a heavy C_4.

Let t_a be the number of a-bad vertices and t_b be that of the b-bad vertices. Now, t_a is bounded from above by the number of four-tuples (a, z, y, x) such that (y, z) is an edge of the path, $(a, z), (y, x) \in E_1$ and $l((a, z)) + l((y, x)) \geq q + 1$ holds. There are $d(a, i)$ choices for (a, z) of label i. The vertex y can be chosen in two different ways along the path, finally the number of choices for (y, x) with label j is $d(y, j)$. Therefore the number of these paths can be upperbounded by

$$2\{\sum_i d(a, i) \max_{y \in V}\{\sum_{q+1-i \leq j} d(y, j)\}\} = 2t$$

(see (1.3)). We obtained

$$t_a, t_b \leq 2t. \tag{2.1}$$

The number of pairs $\{c, d\}$ $(a \neq c, d \neq b)$ which are neighbours along the path, c is between a and d is $N - 3$. At least $r - 2$ of these pairs satisfy $(a, d) \in E_0$ and at least $r - 2$ of them satisfy $(c, b) \in E_0$. (The number of edges in E_0 starting from a (b) is at least r, three of these edges do not count here: the two edges along the path and eventually $\{a, b\}$.) Consequently, there are at least $2r - N - 1$ pairs having both of the edges in E_0.

The pair $\{c, d\}$ satisfies the conditions of the lemma if it is chosen from the above $2r - N - 1$ ones, d is not a-bad, c is not b-bad and $(d, b) \notin E_1$. (This last condition implies that (a, d, b, c) is not a heavy C_4.) The number of pairs $\{c, d\}$ for which at least one of these conditions does not hold is at most $t_a + t_b + s$. Therefore if $2r - N - 1 > t_a + t_b + s$ holds then the existence of the pair in the lemma is proved. By (2.1) this is reduced to (1.4). ∎

Proof of Theorem 1.3

Let us suppose indirectly, that $2\,r - 4\,t - s - 1 > N$, but the required Hamiltonian cycle does not exist. We say that K *contains a heavy* C_4 if there exists a heavy C_4 whose E_0 edges are edges of K, where K stands for a path or a cycle in G_0.

If $E_1 = \emptyset$, then t and s are zero, the condition of Dirac's theorem holds for G_0, thus it contains a Hamiltonian cycle. Furthermore no heavy C_4 could exist. So, we may assume that E_1 is non-empty. Let us drop edges one-by-one from E_1 until a required Hamiltonian cycle appears. Consider the last dropped edge (u, v). Dropping it, a Hamiltonian cycle containing no heavy C_4 appears. This means, that there was a Hamiltonian cycle C in G_0 before, which contained such heavy C_4s only that used the edge $(u, v) \in E_1$. Let the neighbours of v along C be w and z. A heavy C_4 using the edge (u, v) must use either (w, v) or (z, v). Thus, the path of $N - 1$ vertices from w to z obtained by deleting the vertex v from C contains no heavy C_4.

Lemma 2.2 can be applied for the Hamiltonian path obtained from C by deleting the edge (z, v), taking $a = v$ and $b = z$. Replacing the edges (c, d) (provided by Lemma 2.2) and (z, v) with edges (v, d) and (z, c) a new Hamiltonian cycle C' is obtained, which can contain a heavy C_4 only if that heavy C_4 uses the edge (w, v). Now, a second application of Lemma 2.2 with $a = w$ and $b = v$ gives a Hamiltonian cycle C'' containing no heavy C_4, even without dropping the edge (u, v), a contradiction. ∎

Proof of Theorem 1.2

We construct graphs $G_0 = (V, E_0)$ and $G_1 = (V, E_1)$ that satisfy the requirements of Theorem 1.3. The vertex set V consists of the q-element subsets of X, $|V| = \binom{n}{q} = N$. Two q-element subsets are adjacent in G_0 if their intersection is empty, while two q-element subsets are adjacent in G_1 if they have a non-empty intersection. The label of the edge (A, B) is $l((A, B)) = |A \cap B|$. G_0 is regular with degree $r = \binom{n-q}{q} = \frac{1}{q!}n^q + O(n^{q-1})$. In G_1 we have

$$d(v, i) = d(i) = \binom{q}{i}\binom{n-q}{q-i} \; (1 \le i < q). \tag{2.2}$$

By (1.2) and (2.2) we have

$$s = \sum_{i=1}^{q-1} \binom{q}{i}\binom{n-q}{q-i} = \frac{q}{(q-1)!} n^{q-1} + O(n^{q-2}). \tag{2.3}$$

On the other hand (1.3) and (2.2) imply

$$t = \sum_{q+1 \le i+j} d(i)d(j) = \sum_{q+1 \le i+j} \binom{q}{i}\binom{n-q}{q-i}\binom{q}{j}\binom{n-q}{q-j} =$$

$$= n^{q-1} \sum_{i=2}^{q-1} \binom{q}{i}\binom{q}{q+1-i} \frac{1}{(q-i)!}\frac{1}{(i-1)!} + O(n^{q-2}).$$

It is easy to check that $2r - 4t - s - 1 > N = \binom{n}{q} = \frac{1}{q!}n^q + O(n^{q-1})$, provided $n > n_0(q)$.

According to Theorem 1.3, there is a Hamiltonian cycle H in G_0 that does not contain two disjoint edges that span a heavy C_4. Now the required partition of the q-element subsets into disjoint pairs can be obtained by going around H, every other edge will form a good pair. The condition $\gamma(\{A_1, B_1\}, \{A_2, B_2\}) \le q$ can be deduced from (1.1) and from the fact that H contains no heavy C_4. ∎

3 Generalized coding problems

Define

$$\delta(\{A_1, B_1\}, \{A_2, B_2\}) = 2q - \gamma(\{A_1, B_1\}, \{A_2, B_2\}).$$

This is a "distance" in the "space" of all disjoint pairs of q-element subsets of X. Theorem 1.2 answers a coding type question, how many elements can be chosen from this space with large pairwise distances.

In general, let Y be a finite set and $\delta(x, y) \geq 0$ a real-valued symmetric ($\delta(x, y) = \delta(y, x)$) function defined on the pairs $x, y \in Y$. Let $0 < d$ be a fixed integer. A subset $C = \{c_1, \ldots, c_m\} \subset Y$ is called a *code of distance* d if $\delta(c_i, c_j) \geq d$ holds for $i \neq j$. The following (probably too general) question can be asked.

Problem 3.1 *Let Y, $\delta(x, y)$ and the real d be given. Determine the maximum size $|C|$ of a d-distance code.*

$\delta(x, y)$ is called a *distance* if $\delta(x, y) = 0$ iff $x = y$ and the triangle inequality holds:

$$\delta(x, y) \leq \delta(x, z) + \delta(z, y)$$

for any 3 elements of Y. Problem 3.1 can be asked for $\delta(x, y)$ not possessing these conditions, but it is really more natural for distances.

The best known finite set with a distance is when Y is the set of all sequences of length n, the elements taken from a finite set, the distance is the Hamming distance. Problem 3.1 leads to traditional coding theory. There are many known results of this type in geometry, but there Y is infinite.

Our case when $Y = Y_1$ is the set of all disjoint pairs of q-element subsets of X can also be considered as a set of sequences, however the "distance" is not a Hamming distance. Still, it is a distance.

Proposition 3.2 *Let Y_1 be the set of all disjoint pairs $\{A, B\}$ of q-element subsets of an n-element X.*

$$\delta_1(\{A_1, B_1\}, \{A_2, B_2\}) = 2q - \gamma(\{A_1, B_1\}, \{A_2, B_2\}) \tag{3.1}$$

is a distance.

Proof of Proposition 3.2
It is easy to see that $\delta_1(\{A_1, B_1\}, \{A_2, B_2\}) = 0$ iff $\{A_1, B_1\} = \{A_2, B_2\}$. So we really have to prove only the triangle inequality. By (3.1) and (1.1) this is reduced to

$$\max\{|A_1 \cap A_3| + |B_1 \cap B_3|, |A_1 \cap B_3| + |B_1 \cap A_3|\}+$$

$$+ \max\{|A_2 \cap A_3| + |B_2 \cap B_3|, |A_2 \cap B_3| + |B_2 \cap A_3|\} \leq$$

$$2q + \max\{|A_1 \cap A_2| + |B_1 \cap B_2|, |A_1 \cap B_2| + |B_1 \cap A_2|\}. \tag{3.2}$$

Two cases will be distinguished.

Case 1. *Either the first or the second value is larger (or equal) in both terms of the left hand side of (3.2).*

By symmetry it can be supposed that the first values are the larger ones. Then the left hand side of (3.2) is

$$|A_1 \cap A_3| + |B_1 \cap B_3| + |A_2 \cap A_3| + |B_2 \cap B_3|. \tag{3.3}$$

Observe that

$$|A_1 \cap A_3| + |A_2 \cap A_3| \leq |A_3| + |A_1 \cap A_2| = q + |A_1 \cap A_2|.$$

The same holds for the Bs, therefore (3.3) is at most $2q + |A_1 \cap A_2| + |B_1 \cap B_2|$, proving (3.2) for this case.

Case 2. *The first value is larger in the first term, the second one in the second term, or vice versa, on the left hand side of (3.2).*

By symmetry we can suppose that the left hand side of (3.2) is

$$|A_1 \cap A_3| + |B_1 \cap B_3| + |A_2 \cap B_3| + |B_2 \cap A_3|. \tag{3.4}$$

All these intersections are subsets of $A_3 \cup B_3$. Using the fact that $A_i \cap B_i = \emptyset$, only the first and the fourth, the second and the third, resp., intersections can be non-disjoint, the other pairs are disjoint. Therefore no element is in more than two of the intersections in (3.4) and these elements are all either in $A_1 \cap B_2 \cap A_3$ or in $B_1 \cap A_2 \cap B_3$. This gives an upper bound on (3.4):

$$|A_3| + |B_3| + |A_1 \cap B_2 \cap A_3| + |B_1 \cap A_2 \cap B_3| \leq 2q + |A_1 \cap B_2| + |B_1 \cap A_2|,$$

proving (3.2) for this case, too. ∎

The following special case of Problem 3.1 arises now naturally.

Problem 3.3 *Let Y_1, $\delta_1(x, y)$ be the space with distance defined above. Determine the maximum size $|C|$ of a q-distance code.*

Unfortunately, Theorem 1.2 is not a solution, since the condition on the distance permits the existence of a pair $\{A, B_1\}, \{A, B_2\}, B_1 \cap B_2 = \emptyset$, which is excluded in Theorem 1.2 by the unique usage of every q-element subset.

Let us see some other possible special cases of Problem 3.1.

Problem 3.4 *Let Y be set of all permutations of n elements and suppose that δ is the number of inversions between two permutations (number of pairs beeing in different order). Given the integer $0 < d$, determine the largest set of permutations with pairwise distance at least d.*

Problem 3.5 *Let Y be the set of $n \times n$ matrices over a finite field F and suppose that the distance δ between two such matrices is the rank of the difference (entry by entry) of the matrices. Given the integer $0 < d$ determine the largest set of matrices with pairwise distance at least d.*

Finally, let Y be the set of all simple graphs $G = (V, E)$ on the same vertex set V, $|V| = n$. The distance $\delta(G_1, G_2)$ between the graphs $G_1 = (V, E_1), G_2 = (V, E_2)$ is the size of the largest complete graph in $(V, E_1 \circ E_2)$ where $\circ$ is the symmetric difference. Some results on this problem will be published in a forthcoming paper [1].

References

[1] N. Alon, G.O.H. Katona, Codes among graphs, paper under preparation.

[2] 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.

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