A coding problem for pairs of subsets

Béla Bollobás* Zoltán Füredi† Ida Kantor‡ G. O. H. Katona§ Imre Leader¶

March 6, 2014

Keywords: Transportation distance, packings, codes, designs, difference sets, randomized constructions.

Abstract: Let X be an n–element finite set, $0 < k \leq n/2$ an integer. Suppose that $\{A_1, A_2\}$ and $\{B_1, B_2\}$ are pairs of disjoint k-element subsets of X (that is, $|A_1| = |B_1| = |A_2| = |B_2| = k, A_1 \cap A_2 = \emptyset, B_1 \cap B_2 = \emptyset$). Define the distance of these pairs by $d(\{A_1, A_2\}, \{B_1, B_2\}) = \min\{|A_1 - B_1| + |A_2 - B_2|, |A_1 - B_2| + |A_2 - B_1|\}$. Let $C(n, k, d)$ be the maximum size of a family of pairs of *disjoint* subsets, such that the distance of any two pairs is at least d.

Here we establish a conjecture from [2] concerning the asymptotic formula for $C(n, k, d)$ by using the randomized packing theorem of J. Kahn [10] for k, d are fixed and $n \to \infty$. Also, an infinite number of exact results are given by using special difference sets of integers. Finally, the questions discussed above are put into a more general context and a number of coding theory type problems are proposed.

1 The transportation distance

Let X be a finite set of n elements, if it is convenient we identify it with the set $[n] := \{1, 2, \ldots, n\}$. The family of the k-sets of an underlying set X is denoted by $\binom{X}{k}$. For $0 < k \leq n/2$ let $\mathcal{Y}$ be the family of unordered disjoint pairs $\{A_1, A_2\}$ of k-element subsets of X (that is, $|A_1| = |A_2| = k, A_1 \cap A_2 = \emptyset$). We define the *transportation distance* d on $\mathcal{Y}$ by

$$d(\{A_1, A_2\}, \{B_1, B_2\}) = \min\{|A_1 - B_1| + |A_2 - B_2|, |A_1 - B_2| + |A_2 - B_1|\}. \tag{1}$$

In fact, this is an instance of a more general notion. Whenever (Z, ρ) is a metric space, we can define a metric $\rho^{(s)}$ on Z^s by

$$\rho^{(s)}((x_1, \ldots, x_s), (y_1, \ldots, y_s)) = \min_{\pi \in S_s} \sum_{i=1}^{s} \rho(x_i, y_{\pi(i)}). \tag{2}$$

*Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WB, UK, and Department of Mathematical Sciences, University of Memphis, Memphis TN 38152, USA.

†Alfréd Rényi Institute of Mathematics, 13–15 Reáltanoda Street, 1053 Budapest, Hungary.
E-mail: `z-furedi@illinois.edu`. Research supported in part by the Hungarian National Science Foundation OTKA 104343, and by the European Research Council Advanced Investigators Grant 267195.

‡Computer Science Institute of Charles University, Malostranské nám. 25, 118 00 Praha 1. Czech Republic.
E-mail: `ida@iuuk.mff.cuni.cz`. Research supported by GAČR grant number P201/12/P288 and partially done while this author visited the Rényi Institute.

§Rényi Institute, Hungarian Academy of Sciences, Budapest, Reáltanoda u. 13–15, 1053 Hungary.
E-mail: `ohkatona@renyi,mta.hu`. Research was supported by the Hungarian National Foundation OTKA T029255. This work was done while this author visited the University of Memphis.

¶Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WB, UK.
E-mail`I.Leader@dpmms.cam.ac.uk`. This work was done while this author visited the University of Memphis.

It is not hard to verify that $\rho^{(s)}$ satisfies triangle inequality, i.e., it really is a metric. The transportation distance defined above is obtained by taking $s = 2$, Z to be the set of k-elements subsets of X and ρ is half of their symmetric difference.

The minimization problem (2) (where ρ can be an arbitrary metric) is one of the fundamental combinatorial optimization problems, a so called *assignment problem*, a special case of a more general *Monge-Kantorovich transportation problem* (see, e.g., the monograph [17]).

Transportation distance between finite sets of the same cardinalities is one of the interesting measurements among many different ways to define how two sets differ from each other. In [1], Ajtai, Komlós and Tusnády considered the assignment problem from a different perspective, and determined with *high probability* the transportation distance between two sets of points randomly chosen in a unit square.

Although the transportation distance is an important notion, especially from algorithmic point of view, and there are monographs and graduate texts about this topic, see, e.g., [17], it is not mentioned in the 668 page *Encyclopedia of Distances* [4]; so even the simplest packing problems are still unsolved. Packings of sets in spherical spaces with large transportation distance are considered in [7].

2 Packings and codes

Given a metric space (Z, ρ) and a distance $h > 0$, the *packing number* $\delta(Z, \geq h)$ is the maximum number of elements in Z with pairwise distance at least h.

A (v, k, t) packing $\mathcal{P} \subseteq \binom{[v]}{k}$ is a family of k-sets with pairwise intersections at most $t - 1$ (here $v \geq k \geq t \geq 1$). In other words, every t-subset is covered at most once. Its maximum size is denoted by $P(v, k, t)$. Obviously,

$$P(v,k,t) \leq \binom{v}{t} / \binom{k}{t}. \tag{3}$$

If here equality holds then $\mathcal{P}$ is called a Steiner system $S(v, k, t)$, or a t-*design* of parameters v, k, t and $\lambda = 1$ (for more definitions concerning symmetric combinatorial structures esp., difference sets, etc. see, e.g., the monograph by Hall [9]). More generally, for a set K of integers, a family $\mathcal{P}$ on v elements is called a (v, K, t)-design (packing) if every t-subset of $[v]$ is contained in exactly one (at most one) member of $\mathcal{P}$ and $|P| \in K$ for every $P \in \mathcal{P}$.

Determining the packing number is a central problem of Coding Theory, it is essentially the same problem as finding the rate of a large-distance error-correcting code.

If equality holds in (3) then every i-subset of $[v]$ is contained in $\binom{v-i}{t-i}/\binom{k-i}{t-i}$ members of $\mathcal{P}$ for $i = 0, 1, \ldots, t-1$. We say that v, k, and t satisfy the *divisibility conditions* if these t fractions are integers. It was recently proved by Keevash [12] that for any given k and t there exists a bound $v_0(k, t)$ such that these trivial necessary conditions are also sufficient for the existence of a t-design.

$$\text{An } S(v,k,t) \text{ exists if } v, k, \text{ and } t \text{ satisfy the divisibility conditions and } v > v_0(k,t). \tag{4}$$

This implies Rödl's theorem[16], that for given k and t as $v \to \infty$

$$P(v,k,t) = (1 + o(1))\binom{v}{t} / \binom{k}{t}. \tag{5}$$

Even more, (4) implies that here the error term is only $O(v^{t-1})$. The case $t = 2$ was proved much earlier by Wilson [18]. For this case he also proved the following more general version. For a finite K there exists a bound $v_0(K, 2)$ such that for $v > v_0(K, 2)$

$$\text{a } (v, K, 2) \text{ design exists if } v \text{ and } K \text{ satisfy the generalized divisibility conditions,} \tag{6}$$

namely, g.c.d.$(\binom{k}{2} : k \in K)$ divides $\binom{v}{2}$ and g.c.d.$(k - 1 : k \in K)$ divides $v - 1$.

3 Packing pairs of subsets

In this paper, we concentrate on the space $\mathcal{Y}$ of pairs of *disjoint* subsets. We say that a set $\mathcal{C} \subset \mathcal{Y}$ of such pairs is a 2-(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 a 2-(n,k,d)-code. Enomoto and Katona in [5] proposed the problem of determining $C(n,k,d)$. For the origin of the problem see [3]. Connections to Hamilton cycles in the Kneser graph $K(n,k)$ are discussed in [11]. The problem makes sense only when $d \le 2k \le n$. It is obvious, that a maximal 2-$(n,k,1)$ code consists of all the pairs, $C(n,k,1) = |\mathcal{Y}| = \frac{1}{2}\binom{n}{k}\binom{n-k}{k}$. A 2-$(n,k,2k)$ code consists of mutually disjoint k-sets, hance $C(n,k,2k) = \lfloor n/2k \rfloor$.

In Section 5 we present a method for the determination the exact value of $C(n,k,2k-1)$ for infinitely many n. However, we were able to complete the cases $k = 2,3$ only, the cases of pairs and triple systems.

Theorem 1. *If* $n \equiv 1 \bmod 8$ *and* $n > n_0$ *then* $C(n,2,3) = \frac{n(n-1)}{8}$.
If $n \equiv 1, 19 \bmod 342$ *and* $n > n_0$ *then* $C(n,3,5) = \frac{n(n-1)}{18}$.

The following theorem was proved in [2]. Let $d \le 2k \le n$ be integers. Then

$$C(n,k,d) \le \frac{1}{2}\frac{n(n-1)\cdots(n-2k+d)}{k(k-1)\cdots\lceil\frac{d+1}{2}\rceil \cdot k(k-1)\cdots\lfloor\frac{d+1}{2}\rfloor}. \tag{7}$$

Quisdorff [15] gave a new proof and using ideas from classical coding theory he significantly improved the upper bound for small values of n (for $n \le 4k$). For completeness, in Section 6 we reprove (7) in an even more streamlined way.

Concerning larger values of n one can build a 2-(n,k,d) code from smaller ones using the following observation. If $|(A_1 \cup A_2) \cap (B_1 \cup B_2)| \le 2k-d$ holds for the disjoint pairs $\{A_1, A_2\} \in \mathcal{Y}$, $\{B_1, B_2\} \in \mathcal{Y}$ then $d(\{A_1, A_2\}, \{B_1, B_2\}) \ge d$. Take a $(2k-d+1)$-packing $\mathcal{P}$ on n elements and choose a 2-$(|P|,k,d)$-code on each members $P \in \mathcal{P}$. We obtain

$$\sum_{P \in \mathcal{P}} C(|P|,k,d) \le C(n,k,d). \tag{8}$$

This gives

$$P(n,p,2k-d+1)C(p,k,d) \le C(n,k,d). \tag{9}$$

Fix p (and k, t and d) then Rödl's theorem (5) gives $(1+o(1))\binom{n}{2k-d+1}\binom{p}{2k-d+1}^{-1}C(p,k,d) \le C(n,k,d)$. Rearranging we get, that the sequence $C(n,k,d)/\binom{n}{2k-d+1}$ is essentially nondecreasing in n, for any fixed p (and k, t and d)

$$C(p,k,d)/\binom{p}{2k-d+1} \le (1+o(1))C(n,k,d)/\binom{n}{2k-d+1}.$$

Since, obviously, $C(2k,k,d) \ge 1$ we obtain that $\lim_{n\to\infty} C(n,k,d)/\binom{n}{2k-d+1}$ exists, it is positive, it equals to its supremum, and finite by (7).

It was conjectured ([2], Conjecture 8) that the upper estimate (7) is asymptotically sharp. We prove this conjecture in Section 7.

Theorem 2.

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

4 The case $d = 2$, the exact values of $C(n,k,2)$

Besides the cases mentioned in the previous Section (the cases $d = 1$, $d = 2k$ and $(k,d) \in \{(2,3),(3,5)\}$) we can solve one more case easily, namely if $d = 2$. Since $C(n,2k,2)] = |\mathcal{Y}| = \frac{1}{2}\binom{2k}{k}$ the construction (9) gives $P(n,2k,2k-1)\frac{1}{2}\binom{2k}{k} \le C(n,k,2)$. Then the recent result of Keevash (4) gives the lower bound in the following Proposition. The upper bound follows from (7).

Proposition 3. *$C(n,k,2)=\binom{n}{2k-1}\frac{1}{4k}\binom{2k}{k}$ for all $n>n_0(k)$ whenever the divisibility conditions of* (4) *hold.* □

5 The case $d=2k-1$, the exact values of $C(n,k,2k-1)$

The distance $\delta(a,b)$ of two integers mod m $(1\le a,b\le m)$ is defined by

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

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

We say that the pair $S=\{s_1,\dots,s_k\},T=\{t_1,\dots,t_k\}\subset\{1,\dots,m\}$ of disjoint sets is *antagonistic* mod m if

(i) all the $k(k-1)$ integers $\delta(s_i,s_j)(i\ne j)$ and $\delta(t_i,t_j)(i\ne j)$ are different,

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

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

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

Problem 4. *Is there a pair of disjoint, antagonistic k-element sets* mod $2k^2+1$*?*

We have an affirmative answer only in three cases.

Proposition 5. *There is a pair of disjoint, antagonistic k-element sets* mod $2k^2+1$ *when $k=1,2,3$.*

Proof: We simply give such k-element sets in these cases. It is easy to check that they satisfy the conditions.

$k=1$: $S=\{1\},T=\{2\}$.

$k=2$: $S=\{1,8\},T=\{2,3\}$.

$k=3$: $S=\{1,5,19\},T=\{2,13,15\}$. □

Lemma 6. *If there is a pair of disjoint, antagonistic k-element sets* mod m *then $C(m,k,2k-1)\ge m$.*

Proof: Let (S,T) be the antagonistic pair. The shifts $S(u)=\{a+u \bmod m : s\in S\},T(u)=\{s+u \bmod m : s\in T\}(0\le u<m)$ will serve as pairs of disjoint subsets of X.

Suppose that $S(u)$ and $S(v)$ $(u\ne v)$ have two elements in common: $s_1+u=s_2+v\ne s_3+u=s_4+v$ where $s_1,s_2,s_3,s_4\in S,(s_1,s_2)\ne(s_3,s_4)$. The difference is $s_1-s_2=s_3-s_4$ contradicting (i). One can prove in the same way that $T(u)$ and $T(v)$ $(u\ne v)$ and $S(u)$ and $T(v)$, respectively, have at most one element in common. In other words the intersection of any pair from the sets $S(u),T(u),S(v),T(v)$ has at most one element.

Suppose now that both $S(u)\cap S(v)$ and $T(u)\cap T(v)$ are non-empty for some $u\ne v$. Then $s_1+u=s_2+v,t_1+u=t_2+v$ holds for some $s_1,s_2\in S,t_1,t_2\in T$. This leads to $v-u=s_1-s_2=t_1-t_2$, contradicting (i), again.

Finally suppose that both $S(u)\cap T(v)$ and $T(u)\cap S(v)$ are non-empty for some $u\ne v$. Then $s_1+u=t_1+v,t_2+u=s_2+v$ is true for some $s_1,s_2\in S,t_1,t_2\in T$. Here $v-u=s_1-t_1=t_2-s_2$ is obtained, contradicting either (ii) or (iii) (the latter one, if $s_1-t_1=t_1-s_1$ is obtained).

This proves that the distance of the pairs $(S(u),T(u))$ and $(S(v),T(v))$ $(u\ne v)$ is at least $2k-1$. □

Corollary 7. *Suppose that there is Steiner family $\mathcal{S}(n, 2k^2+1, 2)$ and a disjoint, antagonistic pair of k-element subsets* mod $2k^2+1$ *then*

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

Proof: The upper bound $C(n, k, 2k-1) \le n(n-1)/2k^2$ is a corollary of (7).

The lower estimate is obtained from (9). By Lemma 6 one can choose $2k^2+1$ pairs of disjoint k-subsets with distance $2k-1$ in a set of $2k^2+1$ elements. This can be done in each of the members of $\mathcal{S}(n, 2k^2+1, 2)$. Since the members have at most one common element, the distance of two pairs in distinct members of $\mathcal{S}(n, 2k^2+1, 2)$ will have distance at least $2k-1$. Therefore all the

$$|\mathcal{S}(n, 2k^2+1, 2)|(2k^2+1) = \frac{\binom{n}{2}}{\binom{2k^2+1}{2}}(2k^2+1) = \frac{n(n-1)}{2k^2}$$

pairs have distance at least 1. $\square$

Proof of Theorem 1. We only need lower bounds, i.e., constructions. The case $k = 3$ follows from Wilson's theorem (4) of the existence of $S(n, 19, 2)$, Proposition 5 and Corollary 7.

Similarly, the case $k = 2$ for $n \equiv 1, 9 \bmod 72$ follows in the same way using Steiner systems $S(n, 9, 2)$ and the fact $C(9, 2, 3) = 9$ from Corollary 7. However, one can see that $C(17, 3, 2) = 34$ and then the results follows from Wilson's theorem (6) of the existence of $S(n, \{9, 17\}, 2)$ for all large $n \equiv 1 \bmod 8$ and costruction (8).

The construction for $C(17, 2, 3)$ is similar to the proof of Lemma 6. The 9 pairs there are defined as $\{\{x+1, x+8\}, \{x+2, x+3\}\} : x \in Z_9\}$. These correspond to a perfect edge decomposition of K_9 into C_4's with side lengths 1, 3, 4, and 2. For $n = 17$ we take the pairs $\{\{x, x+7\}, \{x+2, x+6\}\} : x \in Z_{17}\}$ and $\{\{y, y+11\}, \{y+7, y+8\}\} : y \in Z_{17}\}$ which correspond to C_4's of side lengths $(2, 5, 1, 6)$ and $(7, 4, 3, 8)$, respectively. $\square$

Note that the method gives that $C(n, 1, 1) = \frac{n(n-1)}{2}$ when $n \equiv 1, 3 \bmod 6$. This, however, is trivial for all n.

6 A new proof of the upper estimate

The upper estimate in (7) was proved in [2]. We give a new, more illuminating proof here.

Given a pair $\{A, B\}$ of disjoint k-element sets let $\mathcal{P}(\{A, B\}, u, v)$ denote the family of pairs $\{U, V\}$ where $|U| = u, |V| = v$ and $U \subseteq A, V \subseteq B$ or vice versa. We have

$$|\mathcal{P}(\{A, B\}, u, v)| = 2\binom{k}{u}\binom{k}{v}.$$

Suppose first $u < v$. Then the total number of pairs $\{U, V\}, U \cap V = \emptyset, |U| = u, |V| = v$ in an n-element set is

$$\binom{n}{u}\binom{n-u}{v}.$$

Let $\{A_1, B_1\}, \{A_2, B_2\}$ be two pairs with distance at least d, and $u \le v$ be two nonnegative integers such that $u + v = 2k - d + 1$. By definition (1), $\mathcal{P}(\{A_1, B_1\}, u, v)$ and $\mathcal{P}(\{A_2, B_2\}, u, v)$ are disjoint. We have

$$C(n, k, d) \le \frac{\binom{n}{u}\binom{n-u}{v}}{2\binom{k}{u}\binom{k}{v}} = \frac{n(n-1)\ldots(n-2k+d)}{2k(k-1)\ldots(k-u+1)k(k-1)\ldots(k-v+1)} \tag{10}$$

for every pair u, v that satisfies the above requirements. If $u = v$, then equality (10) holds by similar arguments.

The numerator does not depend on u, and the denominator is maximized when u and v are as close as possible, i.e., for $u = 2k - \lceil \frac{d-1}{2} \rceil$ and $v = 2k - \lfloor \frac{d-1}{2} \rfloor$. Substituting these values, we obtain the upper estimate in (7). $\square$

7 Nearly perfect selection

Let $\mathcal{W}$ be the family of pairs $\{U,V\}$ such that $U,V\subseteq[n]$, $U\cap V=\emptyset$, and $|U|+|V|=2k-d+1$ holds. Note that $|\mathcal{W}|=\frac{1}{2}\sum_{0\le u\le 2k-d+1}\binom{n}{u}\binom{n-u}{(2k-d+1)-u}$. For a pair $\{A,B\}$ of disjoint k-element sets, let $\mathcal{P}(\{A,B\})$ denote the family of pairs $\{U,V\}\in\mathcal{W}$ for which $U\subseteq A$ and $V\subseteq B$, or vice versa.

Lemma 8. *$d(\{A_1,B_1\},\{A_2,B_2\})\le d-1$ holds if and only if $\mathcal{P}(\{A_1,B_1\})\cap\mathcal{P}(\{A_2,B_2\})\ne\emptyset$.*

Proof: Suppose that $\{U,V\}\in\mathcal{P}(\{A_1,B_1\})\cap\mathcal{P}(\{A_2,B_2\})$, say $U\subset A_1\cap A_2$ and $V\subset B_1\cap B_2$. Then $|A_1-A_2|\le k-|U|,|B_1-B_2|\le k-|V|$ imply $|A_1-A_2|+|B_1-B_2|\le 2k-|U|-|V|=d-1$ proving the statement. The other case is analogous.

Conversely, if the distance is at most $d-1$ then either $|A_1-A_2|+|B_1-B_2|\le d-1$ or $|A_1-B_2|+|B_1-A_2|\le d-1$ must hold. Suppose that the first one is true. Then $|A_1\cap A_2|+|B_1\cap B_2|\ge 2k-d+1$ follows. Take $U=A_1\cap A_2$ and a $V\subseteq B_1\cap B_2$ such that $|V|=2k-d+1-|U|$. Then $\mathcal{P}(\{A_1,B_1\})\cap\mathcal{P}(\{A_2,B_2\})\ne\emptyset$ holds, as claimed. □

We can view the sets $\mathcal{P}(\{A,B\})$ as the edges of a hypergraph on the vertex set $\mathcal{W}$. Let us call this hypergraph $\mathcal{H}$. Then a 2-(n,k,d)-code corresponds to a *matching* in $\mathcal{H}$.

In his celebrated paper [16], Rödl established (5) in the following way. He viewed the t-element sets as vertices of a $\binom{k}{t}$-uniform hypergraph $\mathcal{H}_n$ whose edges correspond to the k-element subsets of $[n]$. Equality (5) is in fact a statement about the existence of an almost perfect matching in $\mathcal{H}_n$. Using the same key proof idea, a powerful generalization by Frankl and Rödl [6] guarantees the existence of almost perfect matchings in hypergraphs satisfying certain more general conditions. Various generalizations and stronger versions versions were later proved, e.g., by Pippenger and Spencer [14].

A function $t:E(\mathcal{H})\to\mathbb{R}$ is a *fractional matching* of the hypergraph $\mathcal{H}$ if $\sum_{e\in E(\mathcal{H});x\in e}t(e)\le 1$ holds for every vertex $x\in V(\mathcal{H})$. The *fractional matching number*, denoted $\nu^*(\mathcal{H})$ is the maximum of $\sum_{e\in E(\mathcal{H})}t(e)$ over all fractional matchings. If $\nu(\mathcal{H})$ denotes the maximum size of a matching in $\mathcal{H}$, then clearly

$$\nu(\mathcal{H})\le\nu^*(\mathcal{H}).$$

Kahn [10] proved that under certain conditions, asymptotic equality holds. Both the hypotheses and the conclusion are in the spirit of the Frankl–Rödl theorem.

Given a hypergraph $\mathcal{H}$ with vertex set $[n]$, a fractional matching t and a subset $W\subseteq[n]$, define $\bar{t}(W)=\sum_{W\subseteq e\in E(\mathcal{H})}t(e)$ and $\alpha(t)=\max\{\bar{t}(\{x,y\}):x,y\in V(\mathcal{H}),x\ne y\}$. In other words, $\alpha(t)$ is a fractional generalization of the codegree. We say that $\mathcal{H}$ is *s-bounded* if each of its edges has size at most s.

Theorem 9 ([10])**.** *For every s and every $\varepsilon>0$ there is a δ such that whenever $\mathcal{H}$ is an s-bounded hypergraph and t a fractional matching with $\alpha(t)<\delta$, then*

$$\nu(\mathcal{H})>(1-\varepsilon)t(\mathcal{H}).$$

Proof of Theorem 2. In the light of Lemma 8 it suffices to verify the conditions of Theorem 9 and to produce a fractional matching t of the hypergraph $\mathcal{H}$ of the desired size.

Define a constant weight function $t:E(\mathcal{H})\to\mathbb{R}$ by

$$t(e)=\frac{\lceil\frac{d-1}{2}\rceil!\lfloor\frac{d-1}{2}\rfloor!}{n^{d-1}}.$$

For a vertex $x=\{U,V\}\in\mathcal{W}$ with $|U|=u$ and $|V|=v$ we have

$$\deg(\{U,V\})=\binom{n-u-v}{k-u}\binom{n-k-v}{k-v}\le\frac{n^{d-1}}{(k-u)!(k-v)!}\le\frac{n^{d-1}}{\lceil\frac{d-1}{2}\rceil!\lfloor\frac{d-1}{2}\rfloor!}$$

hence t is indeed a fractional matching. Note that $t(\mathcal{H})$ is is asymptotically equal to the quantity in the statement of the Theorem 2.

The hypergraph $\mathcal{H}$ is not regular but s-bounded with $s = \frac{1}{2}\sum_u \binom{k}{u}\binom{k}{(2k-d+1)-u}$. Here s does not depend on n. For $x, y \in V(\mathcal{H}) = \mathcal{W}$ let $\deg(x, y)$ denote the codegree of $x = \{U, V\}$ and $y = \{U', V'\}$, i.e., the number of hyperedges $\mathcal{P}(\{A, B\})$ that contain both x and y. If $U \cup V = U' \cup V'$ (they partition the same $(2k-d+1)$-element set) then the codegre $\deg(x, y) = 0$. Otherwise, $|U \cup U' \cup V \cup V'| \geq 2k-d+2$ and $(U \cup U' \cup V \cup V') \subset (A \cup B)$ imply that

$$\deg(\{U, V\}, \{U', V\}) = O(n^{d-2}).$$

Hence $\alpha(t) = \deg(\{U, V\}, \{U', V\}) \cdot t(e) = o(1)$ and Kahn's theorem completes the proof. $\square$

8 s-tuples of sets, q-ary codes

Let $\mathcal{Y}^{(s)}$ be the family of s-tuples of pairwise disjoint k-element subsets of $[n]$. A natural definition of a metric on $\mathcal{Y}^{(s)}$ was already mentioned in the introduction, in equation (2). With ρ being half the symmetric difference, the distance is defined as

$$\rho^{(s)}(\{A_1, \dots, A_s\}, \{B_1, \dots, B_s\}) = \min_{\pi \in S_s} \sum_{i=1}^{s} |A_i \setminus B_{\pi(i)}|.$$

Let $C_s(n, k, d)$ denote the maximum size of a subfamily $\mathcal{S}$ of $\mathcal{Y}^{(s)}$ such that any two elements in $\mathcal{S}$ have distance at least d. The proofs presented in Sections 7 and 6 can be easily adapted to determining $C_s(n, k, d)$, as well. The proof of the lower and the upper bounds in Theorem 10 is completely analogous to the proofs of equation (7) and Theorem 2.

Theorem 10.

$$\lim_{n\to\infty} \frac{C_s(n, k, d)}{n^{sk-d+1}} = \frac{1}{s!} \frac{\lceil \frac{d-1}{s} \rceil! \lceil \frac{d-2}{s} \rceil! \dots \lceil \frac{d-s}{s} \rceil!}{(k!)^s}. \qquad \square$$

Let $\mathcal{Y}_q$ be the set of q-ary vectors of length n and weight k (weight is the number of nonzero entries). Let $A_q(n, d, k)$ be the maximum size of a subset $\mathcal{C} \subseteq \mathcal{Y}_q$ such that $\rho'(u, v) \geq d$ whenever $u, v \in \mathcal{C}$. Here ρ' is the Hamming distance.

With a slightly more technical proof along the same lines, the following can be proven.

Theorem 11. *Fix $q \geq 2$, k and d. If d is odd, then, as $n \to \infty$,*

$$A_q(n, d, k) \sim \frac{n^{k-\frac{d-1}{2}} (q-1)^{k-\frac{d-1}{2}} \left(\frac{d-1}{2}\right)!}{k!}.$$

If $d \geq 2$ is even, then, as $n \to \infty$,

$$A_q(n, d, k) \sim \frac{n^{k-\frac{d}{2}+1} (q-1)^{k-\frac{d}{2}+1} \left(\frac{d}{2}-1\right)!}{k!}. \qquad \square$$

To use random methods constructing codes is not a new idea. The best known general bounds for the *covering radius* problems are obtained in this way, see, e.g., [8, 13].

We can also consider pairs (or more generally s-tuples) of q-ary vectors of weight k. For simplicity, we will only state the results for pairs here. Define the set $\mathcal{Y}_q^{(2)}$ of pairs $\{u, v\}$ of vectors such that

- $u, v \in \{0, 1, \dots, q-1\}^n$
- each of u and v has exactly k nonzero entries
- the supports of u and v are disjoint (i.e. $u_i = 0$ for all i such that $v_i \neq 0$, and $v_i = 0$ for all i such that $u_i \neq 0$).

Define distance between these pairs by

$$\delta(\{u,v\},\{w,z\}) = \min\{\rho'(u,w)+\rho'(v,z), \rho'(u,z)+\rho'(v,w)\}$$

where ρ' is again the Hamming distance.

In the following, $A_q^2(n,d,k)$ will denote the maximum size of a subset $\mathcal{C} \subseteq \mathcal{Y}_q^{(2)}$ such that $\delta(\{u,v\},\{w,z\}) \geq d$ for any pair $\{u,v\},\{w,z\}$ of members of $\mathcal{C}$.

Theorem 12. *Fix q, d and k. If d is odd and $q \geq 3$, then, as $n \to \infty$,*

$$A_q^2(n,d,k) \sim \frac{1}{2} \cdot \frac{n^{2k-\frac{d-1}{2}} \cdot (q-1)^{2k-\frac{d-1}{2}} \cdot \lfloor \frac{d-1}{4} \rfloor! \lceil \frac{d-1}{4} \rceil!}{(k!)^2}.$$

If $d \geq 2$ is even and $q \geq 2$, then, as $n \to \infty$,

$$A_q^2(n,d,k) \sim \frac{1}{2} \cdot \frac{n^{2k-\frac{d}{2}+1} \cdot (q-1)^{2k-\frac{d}{2}} \cdot \lfloor \frac{d}{4} \rfloor! \left(\lceil \frac{d}{4} \rceil - 1\right)!}{(k!)^2}. \qquad \square$$

The distance δ used here is twice the distance defined in Section 1, hence the apparent inconsistency of this result for $q = 2$ with Theorem 2.

For $q = 2$ and d odd we have $A_q(n,d,k) = A_q(n,d+1,k)$.

9 Open problems

We believe that for an arbitrary pair of k and d, there are infinitely many n's with equality in equation (7).

References

[1] M. Ajtai, J. Komlós, and G. Tusnády, *On optimal matchings*, Combinatorica, 4 (1984), pp. 259–264.

[2] G. Brightwell and G. O. H. Katona, *A new type of coding problem*, Studia Sci. Math. Hungar., 38 (2001), pp. 139–147.

[3] J. Demetrovics, G. O. H. Katona, and A. Sali, *Design type problems motivated by database theory*, J. Statist. Plann. Inference, 72 (1998), pp. 149–164. R. C. Bose Memorial Conference (Fort Collins, CO, 1995).

[4] M. M. Deza and E. Deza, *Encyclopedia of Distances*, Springer, 2nd ed. 2013.

[5] H. Enomoto and G. O. H. Katona, *Pairs of disjoint q-element subsets far from each other*, Electron. J. Combin., 8 (2001), Research Paper 7, 7 pp. (electronic). In honor of Aviezri Fraenkel on the occasion of his 70th birthday.

[6] P. Frankl and V. Rödl, *Near perfect coverings in graphs and hypergraphs*, European J. Combin., 6 (1985), pp. 317–326.

[7] Z. Füredi, *Packings of sets in spherical spaces with large transportation distance*, in preparation.

[8] Z. Füredi and J-H. Kang, Covering the n-space by convex bodies and its chromatic number, *Discrete Mathematics* **308** (2008), 4495–4500.

[9] M. Hall, *Combinatorial Theory, Second Edition*, Wiley-Interscience, 1998.

[10] J. Kahn, *A linear programming perspective on the Frankl-Rödl-Pippenger theorem*, Random Structures Algorithms, 8 (1996), pp. 149–157.

[11] G. O. H. Katona, *Constructions via Hamiltonian theorems*, Discrete Math., 303 (2005), pp. 87–103.

[12] P. Keevash, *The existence of designs*, arxiv.org 1401.3665.

[13] M. Krivelevich, B. Sudakov, and Van H. Vu, *Covering codes with improved density*, IEEE Trans. Inform. Theory 49 (2003), no. 7, 1812–1815.

[14] N. Pippenger and J. Spencer, *Asymptotic behavior of the chromatic index for hypergraphs*, J. Combin. Theory Ser. A, 51 (1989), pp. 24–42.

[15] Jörn Quistorff, *New upper bounds on Enomoto–Katona's coding type problem*, Studia Sci. Math. Hungar. 42 (2005), pp. 61–72.

[16] V. Rödl, *On a packing and covering problem*, European J. Combin., 6 (1985), pp. 69–78.

[17] C. Villani, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.

[18] R. M. Wilson, *An existence theory for pairwise balanced designs. II. The structure of PBD-closed sets and the existence conjectures*, J. Combinatorial Theory Ser. A, 13 (1972), pp. 246–273.