Problems and results in Extremal Combinatorics - V

Dedicated to Péter Frankl, Zoltán Füredi, Ervin Györi and János Pach, for their $70th$ birthday

Noga Alon *

Abstract

Frankl, Füredi, Györi and Pach obtained fundamental results in extremal combinatorics, graph theory and discrete geometry. In this paper we describe solutions or partial solutions of several problems in these areas. The problems considered deal with bipartite coverings of graphs, triangle-free graphs of diameter 2, and geometric and combinatorial designs and partical designs. This is a (short) sequel of several previous papers of the same flavour. Each section here is essentially self contained, and can be read separately.

AMS subject classification: 05C35, 05B05
Keywords: Bipartite coverings, Block designs, Triangle-free graphs

1 Introduction

The conference Summit280, celebrating the 70th birthday of Péter Frankl, Zoltán Füredi, Ervin Györi and János Pach, took place in Budapest in July 2024. The remarkable contributions of these four researchers to Extremal Combinatorics, Graph Theory and Discrete Geometry had a profound impact on the development of modern discrete mathematics. Their results, problems, papers and lectures form an ideal representation of the influential Hungarian school of Combinatorics. The present short article contains several recent solutions or partial solutions of several problems in extremal combinatorics and geometry that originated in this school. These are described in the following three sections, each of which can be read separately. As the title of the paper suggests, this is a (short) sequel of four previous papers [1], [2], [3], [4] of the same flavour.

In section 2 we consider bipartite coverings of graphs, extending old results of Hansel and of Katona and Szemerédi, and slightly strengthening recent work of Kim and Lee.

In Section 3 we prove that for any $0 < \varepsilon < 1/6$ and $n > n_0(\varepsilon)$, one can add to any triangle-free graph on n vertices with maximum degree at most $n^{1/2-\varepsilon}$ less than $3n^{2-\varepsilon}$

*Princeton University, Princeton, NJ, USA and Tel Aviv University, Tel Aviv, Israel. Email: nalon@math.princeton.edu. Research supported in part by NSF grant DMS-2154082.

edges, transforming it to a triangle-free graph with diameter 2. This settles an open problem of Erdős and Gyárfás and is closely related to a problem of Erdős and Pach and to a result of Füredi and Seress.

In Section 4 we prove that there is a partial design with n vertices in which each block is of size at least $\Omega(\sqrt{n})$, so that every set that intersects all blocks contains at least $\Omega(\log n)$ points of one of them. We also show that the number of sequences $n \geq x_1 \geq x_2 \geq x_m \geq 2$ so that there is a block design on n elements with blocks of sizes $x_1, x_2, \ldots, x_m$ is at least $2^{\Omega(n^{1/2} \log n)}$, significantly larger than the number of such sequences that correspond to geometric designs that can be realized by points and lines in the Euclidean plane. This settles two problems of Erdős.

2 Biparite coverings of graphs and multigraphs

2.1 Results

A bipartite covering $\mathcal{H} = \{H_1, \ldots, H_m\}$ of a graph G on the set of vertices $[n] = \{1, 2, \ldots, n\}$ is a collection of bipartite graphs H_i on $[n]$, so that each edge of G belongs to at least one of them. Note that each H_i is not necessarily a subgraph of G, the only assumption is that it is a bipartite subgraph of the complete graph on $[n]$. The capacity $\text{cap}(\mathcal{H})$ of the cover is the sum $\sum_i |V(H_i)|$ of the numbers of vertices of these bipartite graphs. A known result of Hansel [15] is that the capacity of any bipartite covering of the complete graph K_n on n vertices is at least $n \log_2 n$. This bound is tight when n is power of 2.

In [20] Kim and Lee consider the analogous problem, where the complete graph K_n is replaced by the complete multigraph K_n^λ in which every pair of distinct vertices is connected by λ parallel edges. A bipartite covering here is a collection of bipartite graphs so that each edge belongs to at least λ of them. They prove that the capacity of each bipartite covering of K_n^λ is at least

$$\max\{2\lambda(n-1), n[\log n + \lfloor(\lambda-1)/2\rfloor \log(\frac{\log n}{\lambda}) - \lambda - 1]\},$$

where all logarithms here and in the rest of this section are in base 2. They also establish an upper bound: there exists a bipartite covering of K_n^λ of capacity at most

$$n(\log(n-1) + (1+o(1))\lambda \log\log n).$$

This shows that for $\lambda = (\log n)^{1-\Omega(1)}$ the smallest possible capacity is $n \log n + \Theta(n\lambda \log\log n)$ but leaves an additive gap of $\Omega(\lambda n \log\log n)$ between the upper and lower bounds.

The proofs in [20] proceed by studying a more general problem regarding graphons. Our first contribution in this section is a shorter combinatorial proof of the results above, slightly improving the bounds. Let $\text{cap}(n, \lambda)$ denote the minimum possible capacity of a bipartite covering of K_n^λ.

Theorem 2.1 (Lower bound)**.** *For positive integers $n \geq 2$ and λ,*

$$cap(n, \lambda) \geq \max\{2\lambda(n-1), n[\log n + \lfloor(\lambda-1)/2\rfloor \log(\frac{2\log n}{\lambda - 1})]\}$$

Theorem 2.2 (Upper bound)**.** *Let $k(n, \lambda)$ denote the minimum length of a binary error correcting code with distance at least λ which has at least n codewords. Then $cap(n, \lambda) \leq n \cdot k(n, \lambda)$. Therefore*

1. *For any $n \geq 2$*
$$cap(n, 2) \leq n(\lceil \log n \rceil + 1) < n(\log n + 2).$$

2. *For any n and $\lambda \leq 0.5 \log n$*
$$cap(n, \lambda) \leq n[\log n + (\lambda - 1)(\log(\frac{\log n}{\lambda - 1}) + 4)]$$

3. *For any $0 < c < 1/2$, and for $\lambda \geq c\frac{\log n}{1 - H(c)}$ where $H(x) = -x \log x - (1-x)\log(1-x)$ is the binary entropy function,*
$$cap(n, \lambda) \leq \frac{\lambda}{c} n.$$

4. *For any fixed λ there are infinitely many values of n so that*
$$cap(n, \lambda) \leq n[\log n + \lfloor(\lambda-1)/2)\rfloor \log\log n + 2].$$

Katona and Szemerédi [21] proved the following generalization of the result of Hansel, dealing with the capacity of bipartite coverings of general graphs.

Theorem 2.3 ([21])**.** *Let G be a graph on the set of vertices $[n]$ and let $d_1, d_2, \ldots, d_n$ denote the degrees of its vertices. Then the capacity of any bipartite covering of G is at least*

$$\sum_{i=1}^{n} \log(\frac{n}{n - d_i}).$$

Our second contribution here is the following strengthening of this result.

Theorem 2.4. *Let G be a graph on the set of vertices $[n]$. For each vertex i let α_i denote the maximum size of an independent set of G that contains the vertex i. Then the capacity of any bipartite covering of G is at least*

$$\sum_{i=1}^{n} \log(\frac{n}{\alpha_i}).$$

Since it is clear that $\alpha_i \leq n - d_i$ for every i, this is indeed a strengthening of the Katona-Szemerédi result (Theorem 2.3). The binomial random graph $G = G(n, 0.5)$ is one example for which Theorem 2.4 is strictly stronger than Theorem 2.3. Indeed, with high probability for $G = G(n, 0.5)$, $d_i = (1/2 + o(1))n$ for every i and $\alpha_i = (2 + o(1)) \log n$ for every i. Therefore the lower bound of Theorem 2.3 for this G is typically $(1 + o(1))n$, whereas the lower bound provided by Theorem 2.4 is $n \log n - (1+o(1))n \log\log n$. This is tight since the chromatic number of $G = G(n, 0.5)$ is, with high probability, $\chi_n = (1+o(1))\frac{n}{2\log n}$ implying that G admits a bipartite covering consisting of $\lceil \log \chi_n \rceil$ spanning bipartite graphs, and the corresponding capacity is $n \log n - (1 + o(1))n \log\log n$.

The rest of this section contains the (short) proofs of the results above.

2.2 Complete multigraphs: the lower bound

Let $\mathcal{H} = \{H_1, \ldots, H_m\}$ be a bipartite covering of the complete multigraph K_n^λ on the set of vertices $[n] = \{1, 2, \ldots, n\}$. We first prove that $\text{cap}(\mathcal{H}) \geq 2\lambda(n-1)$. Let n_i denote the number of vertices of H_i. Since it is bipartite the number of its edges is at most $n_i^2/4$. As the edges of all these graphs cover each of the $n(n-1)/2$ edges of the complete graph on $[n]$ at least λ times, and since $n_i \leq n$ for all i, it follows that

$$\frac{n}{4}\sum_{i=1}^{m} n_i \geq \sum_{i=1}^{m} \frac{n_i^2}{4} \geq \lambda n(n-1)/2.$$

This implies that $\text{cap}(\mathcal{H}) = \sum_{i=1}^{m} n_i \geq 2\lambda(n-1)$, as needed.

Note that for any even n this inequality is tight for infinitely many (large) values of λ. In particular it is tight for $\lambda = \frac{n}{4n-4}\binom{n}{n/2}$, and if there is a Hadamard matrix of order n then it is tight for $\lambda = n/2$ as well. In addition, if for some fixed n it is tight for λ_1 and λ_2 then it is also tight for their sum $\lambda = \lambda_1 + \lambda_2$.

We next prove the second inequality, that

$$\text{cap}(\mathcal{H}) \geq n[\log n + \lfloor(\lambda-1)/2\rfloor \log(\frac{2\log n}{\lambda-1})].$$

Without loss of generality assume that each of the bipartite graphs H_i in $\mathcal{H}$ is a complete bipartite graph, and let $L_i, R_i \subset [n]$ denote its two color classes. For each vertex $j \in [n]$ let A_j denote the set of indices i for which the vertex j belongs to the vertex class L_i of H_i and let B_j be the set of indices i for which $j \in R_i$. Let $x_j = |A_j| + |B_j|$ be the total number of bipartite graphs H_i that contain the vertex j. Note that $x_j \geq \lambda$ for each j, as any edge incident with j must be covered at least λ times.

Put $r = \lfloor(\lambda-1)/2\rfloor$ and let $v = (v_1, v_2, \ldots, v_m\}$ be a uniform random binary vector of length m. For each j, $1 \leq j \leq n$, let E_j denote the event that the number of indices i that belong to A_j for which $v_i = 1$ plus the number of indices i that belong to B_j for which $v_i = 0$ is at most r. It is clear that the probability of E_j is exactly the probability that the binomial random variable $B(x_j, 1/2)$ is at most r, which is

$$p(x_j, 1/2) = \frac{\sum_{q=0}^{r} \binom{x_j}{q}}{2^{x_j}}.$$

Note, crucially, that the events E_j are pairwise disjoint. This is because for every two distinct vertices j and j' there are at least $\lambda > 2r$ indices i for which j and j' belong to the two distinct vertex classes of H_i. Therefore

$$\sum_{j=1}^{n} p(x_j, 1/2) \leq 1.$$

The desired lower bound for $\text{cap}(\mathcal{H}) = \sum_{j=1}^{n} x_j$ can be deduced from the last inequality by a convexity argument. We proceed with the details. Note, first, that for every x, $p(x, r) \geq (x/r)^r 2^{-x}$. Therefore

$$\sum_{j=1}^{n} (x_j/r)^r 2^{-x_j} \leq 1. \tag{1}$$

Recall also that for each j, $x_j \geq \lambda \geq 2r + 1$. Consider the function $f(x) = (x/r)^r 2^{-x}$. A simple computation shows that for $r = 1$ its second derivative is $(\ln 2)2^{-x}[x \ln 2 - 2]$ which is positive for all $x \geq 2r + 1 = 3$. For $r \geq 2$ the second derivative of $f(x)$ is

$$(x/r)^{r-2} 2^{-x} [((x/r) \ln 2 - 1)^2 - 1/r].$$

It is not difficult to check that this is positive for all $x \geq 2r + 1$. This shows that $f(x)$ is convex in the relevant range. Therefore, by (1) together with Jensen's Inequality, if we denote $x = \text{cap}(\mathcal{H}) = \sum x_j$ we get

$$n(\frac{x}{nr})^r 2^{-x/n} \leq 1$$

implying that

$$x \geq n[\log n + r \log(\frac{x}{nr})].$$

Since $x/n \geq \log n$ this shows that

$$x \geq n[\log n + r \log(\frac{\log n}{r}] \geq n[\log n + \lfloor (\lambda - 1)/2 \rfloor \log(\frac{2 \log n}{\lambda - 1})].$$

This completes the proof of Theorem 2.1. □

2.3 Complete multigraphs, the upper bound

In this subsection we prove Theorem 2.2. Put $k = k(n, \lambda)$ and let $A = (a_{ij})$ be the k by n binary matrix whose columns are n of the codewords of a binary code of length k with minimum distance (at least) λ. For each i, $1 \le i \le k$, let H_i be the complete bipartite graph on the classes of vertices $L_i = \{j : a_{ij} = 0\}$ and $R_i = \{j : a_{ij} = 1\}$. It is easy to see that these bipartite graphs cover every edge of the complete graph on $[n]$ at least λ times. The capacity of this covering is at most kn, establishing the first part of the theorem. The subsequent items in the theorem follow by considering appropriate known error correcting codes, see, e.g. [22].

For the first item simply take the code consisting of all 2^{k-1} codewords with even Hamming weight. Since $2^{k-1} \ge n$ for $k = \lceil \log n \rceil + 1$ the claimed result follows. The second and third items follow from the Gilbert-Varshamov bound which gives that the maximum cardinality of a binary code with length k and distance λ is at least

$$\frac{2^k}{\sum_{i=0}^{\lambda-1} \binom{k}{i}}.$$

This quantity is at least $2^k(\frac{ek}{\lambda-1})^{-(\lambda-1)}$, implying the second item. For any $\lambda = ck \le k/2$ this quantity is also at least $2^{(1-H(c)k}$, where $H(x)$ is the binary entropy function. This yields the third item.

The fourth follows by considering an appropriate augmented BCH code. For any k which is a power of 2 and for any d this is a (linear) binary code of length k with

$$n = \frac{2^k}{2k^{d-1}}$$

codewords and minimum distance $2d$. For $d - 1 = \lfloor (\lambda - 1)/2 \rfloor$, $2d \ge \lambda$ and

$$k = \log n + 1 + (d-1)\log k \le \log n + \lfloor (\lambda-1)/2 \rfloor \log\log n + 2.$$

This completes the proof of Theorem 2.2. □

2.4 General graphs

In this subsection we prove Theorem 2.4. We need the following simple lemma.

Lemma 2.5. *Let $E_i, i \in I$ be a finite collection of events in a (discrete) probability space. Suppose that for every point x in the space, if $x \in E_i$ then the total number of events E_j in the collection that contain x is at most a_i. Then*

$$\sum_{i \in I} \frac{Prob(E_i)}{a_i} \le 1. \tag{2}$$

It is worth noting that the above holds (with the same proof) for any probability space, the assumption that it is discrete here is merely because this is the case we need, and it slightly simplifies the notation in the proof.

Proof. Let x be an arbitrary point of the space, and let $p(x)$ denote its probability. Suppose it belongs to r of the events E_i, let these be $E_{i_1}, \dots E_{i_r}$. By the definition of the numbers a_i it follows that $a_{i_j} \geq r$ for all $1 \leq j \leq r$. Therefore the total contribution of the point x to the sum in the left-hand-side of (2) is

$$\sum_{j=1}^{r} \frac{p(x)}{a_{i_j}} \leq \sum_{j=1}^{r} \frac{p(x)}{r} \leq p(x).$$

The desired result follows by summing over all points x in the space. $\square$

Proof of Theorem 2.4: Let G be a graph on the set of vertices $[n]$, let α_i denote the maximum cardinality of an independent set of G containing the vertex i, and let $\mathcal{H} = \{H_1, H_2, \dots, H_m\}$ be a bipartite covering of G.

As in the proof of Theorem 2.1 we may and will assume, without loss of generality, that each of the bipartite graphs H_i in $\mathcal{H}$ is a complete bipartite graph. Let $L_i, R_i \subset [n]$ denote its two color classes. For each vertex $j \in [n]$ let A_j denote the set of indices i for which the vertex j belongs to the vertex class L_i of H_i and let B_j be the set of indices i for which $j \in R_i$. Let $x_j = |A_j| + |B_j|$ be the total number of bipartite graphs H_i that contain the vertex j. Our objective is to prove a lower bound for the capacity of $\mathcal{H}$, which is exactly the sum $\sum_{j=1}^{n} x_j$.

Let $v = (v_1, v_2, \dots, v_m\}$ be a uniform random binary vector of length m. For each j, $1 \leq j \leq n$, let E_j denote the event that $v_i = 0$ for every index i that belongs to A_j and $v_i = 1$ for every index i that belongs to B_j. Note that the probability of E_j is exactly 2^{-x_j}. Note also that if some point $v = (v_1, v_2, \dots, v_m\}$ belongs to the events $E_j, j \in J$, then the set of vertices $J \subset [n]$ is an independent set of G. Indeed, if some two vertices in J are adjacent, then the edge connecting them belongs to at least one of the graphs H_i implying that one of these vertices belongs to L_i whereas the other lies in R_i and showing that they can't both satisfy the requirement given by v_i. It thus follows that any point v that lies in E_j belongs to at most α_j of the events $E_{j'}$. Therefore, by Lemma 2.5

$$\sum_{j=1}^{n} 2^{-x_j - \log \alpha_j} = \sum_{j=1}^{n} \frac{2^{-x_j}}{\alpha_j} \leq 1.$$

By the arithmetic-geometric means inequality this implies

$$n 2^{-(\sum_{j=1}^{n} x_j + \sum_{j=1}^{n} \log \alpha_j)/n} \leq 1,$$

giving

$$2^{\sum_{j=1}^{n} x_j} \geq n^n 2^{-\sum_{j=1}^{n} \log \alpha_j} = 2^{\sum_{j=1}^{n} (\log n - \log \alpha_j)}.$$

Therefore

$$\sum_{j=1}^{n} x_j \geq \sum_{j=1}^{n} \log(\frac{n}{\alpha_j}),$$

completing the proof. $\square$

3 Triangle-free graphs of diameter 2

The following problem was raised by Erdős and Gyárfás ([12], see also [9], Problem number 134).

Problem 3.1. *Let $\varepsilon, \delta > 0$ be two fixed positive reals, and suppose n is large as a function of ε, δ. Let G be a triangle-free graph on n vertices with maximum degree smaller than $n^{1/2-\varepsilon}$. Can G be made into a triangle-free graph with diameter 2 by adding at most δn^2 edges ?*

They proved that the conclusion holds if the maximum degree is at most $\log n / \log\log n$. It is mentioned in [12] that Simonovits showed that this does not necessarily hold if the maximum degree is $C\sqrt{n}$ for some large fixed C. In fact, maximum degree $(1+o(1))\sqrt{n/2}$ suffices as shown by the incidence graph of the lines and points of a projective plane of order p. This is a bipartite $(p+1)$-regular (triangle-free, of course) graph on $n = 2(p^2+p+1)$ vertices, where p is a prime power. Any two vertices of the same vertex class in this graph have a common neighbor. Therefore one cannot add any edge connecting two vertices of the same vertex class without creating a triangle. It follows that in order to reduce the diameter to 2 one must add all missing edges between pairs of nonadjacent vertices that do not lie in the same vertex class. The number of these missing edges is $(1/4 - o(1))n^2 = \Omega(n^2)$.

We describe a short proof of the following, which settles the Erdős-Gyárfás problem in a strong form. Here and in what follows we do not make any effort to optimize the absolute constants.

Theorem 3.2. *Let $G = (V, E)$ be a triangle-free graph with n vertices and maximum degree $d \le c(n)\sqrt{n}$, where*

$$2\frac{(\log n)^{1/3}}{n^{1/6}} \le c = c(n) \le \frac{1}{10}$$

and n is sufficiently large. Then one can add to G at most $2.5cn^2$ edges and get a triangle-free graph of diameter 2.

Note that by taking $c(n) = n^{-\varepsilon}$ the above theorem implies that if in the Erdős-Gyárfás problem the maximum degree is at most $n^{1/2-\varepsilon}$ then it suffices to add at most $O(n^{2-\varepsilon})$ edges.

A related problem of Erdős and Pach dealing with triangle-free graphs appears right before the problem above in [12], see also [9], Problem 133.

Problem 3.3. *Let $f(n)$ denote the smallest integer for which there is a triangle-free graph G on n vertices, diameter 2 and maximum degree $f(n)$. What is the order of growth of $f(n)$?*

Erdős and Pach conjectured that $f(n)/\sqrt{n}$ tends to infinity as n tends to infinity.

A tight asymptotic expression for $f(n)$ is still open. Clearly $f(n) \geq (1-o(1))\sqrt{n}$ as any graph with maximum degree f and diameter 2 can have at most $1+f+f(f-1) = f^2+1$ vertices. We suspect that $f(n) = (1+o(1))\sqrt{n}$. It is known that for $r \in \{1,2,3,7\}$ and possibly also for $r = 57$ there is an r-regular triangle-free graph of diameter 2 with r^2+1 vertices. Although the Hoffman-Singleton Theorem [18] asserts that there are no such graphs for additional values of r, it is possible that there are triangle-free graphs of diameter 2, maximum degree r and $(1-o(1))r^2$ vertices for infinitely many values of r.

3.1 The proof

Proof of Theorem 3.2. Let $G = (V, E)$, d and $c = c(n)$ be as in the statement of the theorem. Throughout the proof we assume, whenever this is needed, that $n > n_0$ where n_0 is a sufficiently large constant. We first apply (a variant of) the triangle-free process for $m = c^2 n^{3/2}$ steps as follows. Starting with $G = G_0$, in each step i for $1 \leq i \leq m$ let G_i be obtained from G_{i-1} by adding a random edge chosen uniformly among all pairs of nonadjacent vertices of G_{i-1} that are both of degree smaller than $2c\sqrt{n}$ and that do not have a common neighbor. Note that by construction the maximum degree of G_m (and hence of all the graphs during the process) is at most $2c\sqrt{n}$. In addition, by construction, G_m (and all graphs during the process) are triangle-free.

Claim : With high probability G_m does not contain an independent set of size $5cn$.

Proof: Fix an independent set U of $5cn$ vertices of $G = G_0$. We estimate the probability that it stays independent in G_m. Since the maximum degree in each G_i is at most $2c\sqrt{n}$, the number of pairs of vertices in U that have a common neighbor is at most

$$n\binom{2c\sqrt{n}}{2} < 2c^2n^2.$$

In addition, the total number of vertices whose degrees have been increased already to $2c\sqrt{n}$ is at most $2cn$ (since the total number of edges added is at most $c^2n^{3/2}$ so the graph consisting of all added edges can have at most $2cn$ vertices of degree at least $c\sqrt{n}$). If follows that in every G_i during the process there are at least

$$\binom{|U| - 2cn}{2} - 2c^2n^2 > 2c^2n^2$$

pairs of vertices of U that are of degree smaller than $2c\sqrt{n}$ and that do not have a common neighbor. Each such pair can be chosen as the selected random edge in each step, and the probability none of these edges have been chosen during the process is at most

$$(1-4c^2)^m \leq e^{-4c^4n^{3/2}}.$$

There are at most

$$\binom{n}{5cn} \leq 2^{H(5c)n} < 2^{10c\log(1/c)n}$$

possible sets U, where H is the binary entropy function. Our choice of c ensures that

$$2^{10c\log(1/c)n}e^{-4c^4n^{3/2}} = o(1).$$

The assertion of the claim thus follows by the union bound.

Returning to the proof of the theorem, fix a graph G_m satisfying the conclusion of the claim and add to it, repeatedly, edges to make it a maximal (with respect to containment) triangle-free graph. In other words, as long as there is a pair of nonadjacent vertices with no common neighbor, add the edge connecting them. This creates a triangle-free graph G' of diameter 2, and its independence number is at most that of G_m, which is smaller than $5cn$, by the claim. This implies that the maximum degree in G' is smaller than $5cn$, as the set of all neighbors of any vertex is an independent set. Therefore G' contains at most $2.5cn^2$ edges, completing the proof of Theorem 3.2. □

Problem 3.3 has been solved up to a constant factor already in the 80s by Hanson and Seyffarth [16] who proved that $f(n) \leq 2\sqrt{n}$. Additional constructions appear in [10], [17]. The problem of deciding whether or not $f(n) = (1 + o(1))\sqrt{n}$ remains open. The constructions in all three papers above are Cayley graphs of Abelian groups, and it is easy to see that using such a construction cannot provide graphs with maximum degree smaller than $(\sqrt{2} + o(1))\sqrt{n}$. A better upper estimate of $(2/\sqrt{3} + o(1))\sqrt{n}$ is described by Füredi and Seress in [14]. See also [5] for some related constructions.

4 Blocking partial designs and block-compatible sequences

4.1 Results

In this section we consider two related open problems of Erdős on block designs. Recall that a family of subsets $A_1, A_2 \ldots, A_m$ of a finite set X is a (pairwise balanced) block design if every pair of distinct elements of X is contained in exactly one of the subsets A_i. It is a partial design if every pair of distinct elements of X is contained in at most one of the subsets A_i (equivalently, if $|A_i \cap A_j| \leq 1$ for all $1 \leq i < j \leq m$.)

The first problem deals with partial designs and appears in [13], see also [9], problem number 664.

Problem 4.1. *Is it true that for every fixed positive constant $c < 1$ there is a finite constant $C = C(c)$ so that the following holds. For every m and n and for every family of subsets $\{A_1, A_2, \ldots, A_m\}$ of $[n] = \{1, 2, \ldots, n\}$ that satisfies $|A_i| > c\sqrt{n}$ for all $1 \leq i \leq m$, and $|A_i \cap A_j| \leq 1$ for all $1 \leq i < j \leq m$, there is a subset $B \subset [n]$ so that $0 < |B \cap A_i| \leq C$ for all $1 \leq i \leq m$?*

The second problem appears in [11], page 35, see also [9], problem number 732.

Problem 4.2. *Call a sequence* $n \geq x_1 \geq x_2 \geq \ldots \geq x_m \geq 2$ block-compatible *for* n *if there is a pairwise balanced block design* $A_1, A_2, \ldots, A_m$ *of* m *subsets of* $[n]$ *such that* $|A_i| = x_i$ *for* $1 \leq i \leq m$*. Is there an absolute constant* $c > 0$ *so that for all large* n *there are at least* $e^{cn^{1/2}\log n}$ *sequences that are block-compatible for* n *?*

We show that the answer to the first problem is "no" and the answer to the second is "yes". The proofs are short, based on appropriate modifications of the family of lines of a projective plane which form a block design with $m = n = q^2 + q + 1$ subsets of cardinality $q + 1 = (1 + o(1))\sqrt{n}$ of a set of size $n = q^2 + q + 1$. It is well known that such a plane exists for any prime power q.

4.2 Proofs

Throughout the proofs we assume, whenever this is needed, that the parameter n is sufficiently large. All logarithms are in base 2, unless otherwise specified. To simplify the presentation we omit all floor and ceiling signs whenever these are not crucial. We make no serious attempt to optimize the absolute constants that appear in the proofs.

The following result settles Problem 4.1

Theorem 4.3. *Let* q *be a (large) prime power and put* $m = n = q^2 + q + 1$*. Then there is a partial design consisting of* m *subsets* $A_1, A_2, \ldots, A_m$ *of an* n *element set* P*, so that* $|A_i| > 0.4\sqrt{n}$ *for all* $1 \leq i \leq m$*,* $|A_i \cap A_j| \leq 1$ *for all* $1 \leq i < j \leq m$*, and for any subset* B *of* P *that has a nonempty intersection with all sets* A_j*, there is some* $1 \leq i \leq m$ *so that* $|B \cap A_i| \geq 0.1 \log n$*.*

Proof. Let P be the set of $n = q^2 + q + 1$ points of a projective plane of order q, and let $L_1, L_1, \ldots, L_m \subset P$ be the sets of points of the m lines of this plane. Thus each point lies in $q + 1$ lines and each line is of size $q + 1$. For each $1 \leq i \leq m$ let A_i be a random subset of L_i obtained by picking every point of L_i, randomly and independently, to lie in A_i with probability $1/2$. Note that the choices of distinct subsets A_i here are independent. By the standard estimates for Binomial distributions (c.f., e.g., [6], Appendix A) together with the union bound, with high probability (that is, with probability tending to 1 as q (or n) tend to infinity) the following conditions hold:

1. Each set A_i is of cardinality $(1/2 + o(1))\sqrt{n}$.
2. Each point lies in $(1/2 + o(1))\sqrt{n}$ of the sets A_i.
3. $|A_i \cap A_j| \leq 1$ for all $1 \leq i < j \leq m$.

Claim 4.4. *With high probability there is no subset* $B \subset P$ *of cardinality at most* $0.3\sqrt{n}\log n$ *that intersects all the sets* A_i*.*

Proof: Let B be a fixed set of cardinality at most $0.3\sqrt{n}\log n$. Consider the set of pairs $S = \{(b, L_i) : b \in B, 1 \le i \le m, b \in L_i\}$. Clearly $|S| = |B|(q+1)$. If $J = J(B)$ is the set of all indices i so that $|B \cap L_i| \ge 0.4 \log n$ then $|S| \ge |J| 0.4 \log n$. Thus

$$|J| \le \frac{|B|(q+1)}{0.4 \log n} \le (3/4 + o(1))n < 0.8n.$$

Therefore there are at least $0.2n$ lines L_i that contain less that $0.4 \log n$ points of B. For each such line L_i, the probability that A_i contains no point of B is at least $2^{-0.4 \log n} = n^{-0.4}$. Therefore, the probability that there is no such line (that is, that B intersects all these sets A_i) is at most

$$(1 - 1/n^{0.4})^{0.2n} \le e^{-0.2n^{0.6}}.$$

Since the total number of possible sets B as above is smaller than

$$n^{0.3\sqrt{n}\log n} = 2^{0.3\sqrt{n}(\log n)^2} = o(2^{0.2n^{0.6}}) \ (= o(e^{0.2n^{0.6}}) \ ),$$

the union bound implies the assertion of the claim.

Returning to the proof of the theorem, fix a choice of the sets A_i that satisfy the conditions 1, 2, 3 above and the assertion of Claim 4.4. If a set B intersects all subsets A_i then it must satisfy $|B| > 0.3\sqrt{n}\log n$. Since each point is contained in $(1/2 + o(1))\sqrt{n}$ of the sets A_i this implies, by averaging, that the intersection of B with some set A_i is of size at least $|B|(1/2 + o(1))\sqrt{n}/n > 0.1 \log n$. This completes the proof. □

The next result settles problem 4.2.

Theorem 4.5. *Let q be a large prime power and put $n = q^2 + q + 1$. Let $S = (x_1 \ge x_2 \ge x_3 \ge \ldots \ge x_m)$ be any sequence of integers satisfying*

$$q + 1 \ge x_1 \ge x_2 \ge x_3 \ldots \ge x_n \ge 3,$$

$$m = n + \sum_{i=1}^{n} [\binom{q+1}{2} - \binom{x_i}{2}],$$

and $x_i = 2$ for all $n < i \le m$. Then S is block-compatible for n. Therefore, the number of sequences that are block-compatible for n is at least

$$\binom{n+q-2}{q-2} = 2^{(0.5+o(1))n^{1/2}\log n}.$$

Proof. Let P and $L_1, L_2, \ldots, L_n \subset P$ be, as in the proof of Theorem 4.3, the set of points of a projective plane of order q and the sets of points of the n lines of P. For each $1 \le i \le n$ let X_i be a subset of cardinality x_i of L_i. Consider the block design consisting of the n blocks X_i together with the following additional

$$\sum_{i=1}^{n} [\binom{q+1}{2} - \binom{x_i}{2}]$$

blocks of cardinality 2: for each $1 \leq i \leq n$, every pair of distinct elements of L_i which is not contained in X_i. This is clearly a block design and the ordered sequence of cardinalities of its blocks is the sequence S. This completes the first part of the proof. For the estimate note that the number of possibilities for the subsequence $x_1 \geq x_2 \geq \ldots \geq x_n \geq 3$ in the above construction is the number of ordered sequences of $q-1$ nonnegative integers whose sum is n, which is

$$\binom{n+q-2}{q-2}.$$

□

4.3 Remarks

- It is easy to see that the estimate in Theorem 4.3 is tight up to constant factors, for every partial design in which all blocks are of sizes $\Theta(\sqrt{n})$. Indeed the following simple fact can be proved by choosing the set B randomly and by applying the union bound and the standard estimates for Binomial distribution.

  Proposition 4.6. *For any two positive constants $c_1 < c_2$ there are two positive constants $C_1(c) < C_2(c)$ so that the following holds. For every m and n and for every family of subsets $\{A_1, A_2, \ldots, A_m\}$ of $[n] = \{1, 2, \ldots, n\}$ that satisfies $c_1\sqrt{n} \leq |A_i| \leq c_2\sqrt{n}$ for all $1 \leq i \leq m$, and $|A_i \cap A_j| \leq 1$ for all $1 \leq i < j \leq m$, there is a subset $B \subset [n]$ so that $C_1 \log n < |B \cap A_i| \leq C_2 \log n$ for all $1 \leq i \leq m$.*

  In fact the conclusion holds, of course, even without any assumption on the sizes of the intersections of pairs of blocks A_i.

- Problem 4.1 for block designs (and not for partial designs) was also asked by Erdős in [11]. This remains open although we suspect that the answer here is negative as well. We suggest the following conjecture which, if true, would establish this negative answer.

  Conjecture 4.7. *Let q be a (large) prime power, put $n = q^2 + q + 1$, let P be the set of n points of a projective plane of order q and let $L_1, L_2, \ldots, L_n$ be the sets of points of its lines. Let R be a random subset of P obtained by picking each point of P randomly and independently to lie in R with probability $1/2$. Then with high probability the smallest cardinality of a subset B of R that intersects all the subsets $L_1 \cap R, L_2 \cap R, \ldots, L_n \cap R$ satisfies $|B|/q > f(q)$ for some function $f(q)$ tending to infinity as q tends to infinity. In fact, this may even be true with $f(q) = \Omega(\log q)$.*

  This conjecture remains open, although related results have been proved in [7], [8] using the container method. The parameters in these papers are very different and it seems that a proof here, if true, would require additional ideas.

- As mentioned by Erdős in [11], the $e^{cn^{1/2}\log n}$ lower bound for the number of block-compatible sequences for n is tight up to the absolute constant c. For completeness we include a brief proof of the upper bound. It is clear that if $n \geq x_1 \geq \ldots \geq x_m \geq 2$ is block-compatible for n, then $m \leq \binom{n}{2} < n^2/2$. Therefore, the number of choices of all x_i which are, say, at most $2\sqrt{n}$ is smaller than

$$\binom{n^2/2 + 2\sqrt{n}}{2\sqrt{n}} < 2^{4\sqrt{n}\log n}.$$

  Next, observe that for each block-compatible sequence x_i for n, the number of indices i with $x_i > 2\sqrt{n}$ is smaller than $\sqrt{n}$. Indeed, otherwise the size of the union of a set of blocks $A_i, i \in I$ with $|A_i| = x_i > 2\sqrt{n}$ and $|I| = \sqrt{n}$ in a block design realizing the sequence is at least

$$\sum_{i \in I} |A_i| - \sum_{i,j \in I, i<j} |A_i \cap A_j| > \sqrt{n} \cdot 2\sqrt{n} - \binom{\sqrt{n}}{2} > n,$$

  which is impossible. Thus there are less than $\sqrt{n}$ such large x_i, and the number of choices for those is at most

$$\binom{n + \sqrt{n}}{\sqrt{n}} < 2^{\sqrt{n}\log n}.$$

  This gives the required $2^{O(\sqrt{n}\log n)}$ upper bound for the total number of block-compatible sequences for n.

- Call a sequence $n \geq x_1 \geq x_2 \geq \ldots \geq x_m \geq 2$ *line-compatible for* n if there is a set P of n points in the Euclidean plane R^2 so that for the family $L_1, L_2, \ldots, L_m$ of all lines in R^2 determined by the points of P, $|L_i \cap P| = x_i$ for $1 \leq i \leq m$. Note that every line-compatible sequence for n is also block-compatible for n, but the converse is not true. Erdős conjectured in [11] (see also [9], problem 733) that the number of sequences which are line compatible for n is only $2^{O(n^{1/2})}$. This upper bound was proved by Szemerédi and Trotter in [24]. Note that in view of Theorem 4.5 this is much smaller than the number of block-compatible sequences for n.

  Indeed, there are far more block designs on n points than designs that can be described by the lines determined by a set of points in the plane. This is demonstrated by the following result.

Proposition 4.8. *1. The number of hypergraphs on n labelled vertices whose edges form a block design is $2^{\Theta(n^2\log n)}$.*

2. The number of hypergraphs whose vertices are n labelled points in R^2 and whose edges are the sets of points contained in the lines determined by the points is only $2^{\Theta(n\log n)}$.

Proof. The lower bound in the first part follows from the known lower bound for the number of Steiner triple systems on n points, which, as proved in [19], is $2^{(1+o(1)n^2 \log n/6}$. (This is the case for n congruent to 1 or 3 modulo 6. For the other values of n we can include some blocks of size 2 as needed.) To prove the upper bound let $x_1, x_2, \ldots x_m$ be the sizes of the blocks. Then $\sum_i \binom{x_i}{2} = \binom{n}{2}$ and thus $\sum x_i \le n^2$. The number of choices of the number of blocks m and their sizes x_i is at most $2^{O(\sqrt{n} \log n)}$. Given those, the number of ways to choose subsets of cardinalities $x_1, x_2, \ldots, x_m$ in $[n]$ is at most

$$\binom{n}{x_1}\binom{n}{x_2}\cdots\binom{n}{x_m} < n^{\sum_i x_i} \le n^{n^2} = 2^{n^2 \log n}.$$

The lower bound in the second part is proved by considering all possible labelings of the hypergraph consisting of a matching of $\lfloor n/3 \rfloor$ pairwise disjoint edges of size 3 and additional $\binom{n}{2} - \lfloor n/3 \rfloor$ edges of size 2. It is easy to realize this hypergraph by points in the plane and the lines they determine by placing 3 points on each of $\lfloor n/3 \rfloor$ parallel lines where no other line contains more than two points.

We prove the upper bound in the second part by using the result of [23] on the number of zero patterns of a sequence of polynomials.

If $f = (f_1, \ldots, f_a)$ is a sequence of polynomials in b variables over a field K, then the *zero pattern* of f evaluated at the point $u \in K^b$ is the set

$$Z_f(u) = \{i \in [a] : f_i(u) = 0\}.$$

Let Z_f denote the total number of distinct zero patterns that appear as u ranges over K^b. We need the following result of Rónyai, Babai, and Ganapathy [23].

Theorem 4.9 ([23])**.** *Let $f = (f_1, \ldots, f_a)$ be a sequence of polynomials in b variables over a field K, and let d_i denote the degree of f_i. Then*

$$Z_f \le \binom{b + \sum_{i=1}^{a} d_i}{b}.$$

In our case the field is R, each point can be described by 2 real variables, so $b = 2n$, and for each set of 3 points there is a degree 2 polynomial that vanishes iff they lie on a line. The zero pattern of these $\binom{n}{3}$ real polynomials of degree 2 in $2n$ variables determines all the lines, and therefore, by Theorem 4.9, the number of possible hypergraphs here is at most

$$\binom{2n + (n^3/6)2}{2n} = 2^{(4+o(1))n \log n}.$$

This completes the proof. ☐

Acknowledgment I thank Ishay Haviv and Wojciech Samotij for helpful comments.

References

[1] Noga Alon, Problems and results in extremal combinatorics, I, Discrete Math. 273 (2003), 31-53.

[2] Noga Alon, Problems and results in extremal combinatorics, II, Discrete Math. 308 (2008), 4460-4472.

[3] Noga Alon, Problems and results in Extremal Combinatorics, III, Journal of Combinatorics 7 (2016), 233–256.

[4] Noga Alon, Problems and results in Extremal Combinatorics, IV, arXiv:2009.12692, 2020.

[5] Noga Alon, Paul Erdös, Ron Holzman and Michel Krivelevich, On k-saturated graphs with restrictions on the degrees, J. Graph Theory 23 (1996), 1-20.

[6] Noga Alon and Joel H. Spencer, *The Probabilistic Method, Fourth Edition*, Wiley, 2016, xiv+375 pp.

[7] József Balogh and Wojciech Samotij, An efficient container lemma, Discrete Anal. 2020, Paper No. 17, 56 pp.

[8] József Balogh and József Solymosi, On the number of points in general position in the plane, Discrete Anal. (2018), Paper No. 16, 20 pp.

[9] Thomas Bloom, Erdős' Open Problems, https://www.erdosproblems.com/all

[10] W. Edwin Clark and John Pedersen, Sum-free sets in vector spaces over GF(2), Journal of Combinatorial Theory, Series A 61 (1992), 222–229.

[11] Paul Erdős, On the combinatorial problems which I would most like to see solved, Combinatorica (1981), 25–42.

[12] Paul Erdős, Some old and new problems in various branches of combinatorics, Discrete Math. 165/166 (1997), 227–231.

[13] Paul Erdős, Some unsolved problems, in: Combinatorics, geometry and probability (Cambridge, 1993) (1997), 1-10.

[14] Zoltán Füredi and Akos Seress, Maximal triangle-free graphs with restrictions on the degrees, J. Graph Theory 18 (1994), no. 1, 11–24.

[15] Georges Hansel, Nombre minimal de contacts de fermeture nécessaires pour réaliser une fonction booléenne symétrique de n variables, C. R. Acad. Sci. Paris 258 (1964), 6037–6040.

[16] Denis Hanson and Karen Seyffarth, k-saturated graphs of prescribed maximum degree, Congressus Numerantium 44 (1984), 127–138.

[17] Ishay Haviv and Dan Levy, Symmetric complete sum-free sets in cyclic groups, Israel J. Math. 227 (2018), 931–956.

[18] Alan J. Hoffman an Robert R. Singleton, Moore graphs with diameter 2 and 3, IBM Journal of Research and Development, 5 (1960), 497–504.

[19] Peter Keevash, Counting Steiner triple systems, European Congress of Mathematics, 459–481, Eur. Math. Soc., Zürich, 2018.

[20] Jaehoon Kim and Hyunwoo Lee, Covering multigraphs with bipartite graphs, arXiv:2304.11691, 2023.

[21] Gyula Katona and Endre Szemerédi, On a problem of graph theory, Studia Sci. Math. Hungar. 2 (1967), 23–28.

[22] F. Jessie MacWilliams and Neil J. A. Sloane, *The Theory of Error-Correcting Codes*, I. North-Holland Mathematical Library, Vol. 16. North-Holland Publishing Co., Amsterdam-New York-Oxford (1977).

[23] Lajos Rónyai, László Babai and Murali K. Ganapathy, On the number of zero-patterns of a sequence of polynomials, J. Amer. Math. Soc. 14 (2001), 717–735.

[24] Endre Szemerédi and William T. Trotter, Extremal problems in discrete geometry, Combinatorica 3 (1983), no. 3-4, 381–392.