Problems and results in Extremal Combinatorics - III

Noga Alon *

Dedicated to Adrian Bondy, for his 70th-birthday

Abstract

Extremal Combinatorics is one of the most active topics in Discrete Mathematics, dealing with problems that are often motivated by questions in other areas, including Theoretical Computer Science, Geometry and Game Theory. This paper contains a collection of problems and results in the area, including solutions or partial solutions to open problems suggested by various researchers. The topics considered here include questions in Extremal Graph Theory, Combinatorial Geometry and Combinatorial Number Theory. This is not a comprehensive survey of the area, and is merely a collection of various extremal problems, which are hopefully interesting. The choice of the problems is inevitably biased, and as the title of the paper suggests, it is a sequel of two previous paper [8], [9] of the same flavour. Each section of this paper is essentially self contained, and can be read separately.

1 Introduction

Extremal Combinatorics deals with the problem of determining or estimating the maximum or minimum possible value of an invariant of a combinatorial object that satisfies certain requirements. Problems of this type are often related to other areas including Computer Science, Information Theory, Number Theory and Game Theory. This branch of Combinatorics has been very active during the the last few decades, see, e.g., [14], [26], and their many references.

This paper contains a collection of problems and results in the area, including solutions or partial solutions to open problems suggested by various researchers. The questions considered include problems in Extremal Graph Theory, Combinatorial Geometry and Combinatorial Number Theory. This is not meant to be a comprehensive survey of the area, but rather a collection of several extremal problems, which are hopefully interesting. The techniques used include combinatorial, probabilistic, geometric and algebraic tools. Each section of this paper is essentially self contained, and can be read separately. This paper is dedicated to Adrian Bondy, a leading graph theorist who has written some of the very best books and survey articles on the subject, including his comprehensive article [15], which is the first chapter of the Handbook of Combinatorics. It is therefore natural to start with an observation settling a problem described in that chapter.

*Sackler School of Mathematics and Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel and School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540. Email: nogaa@tau.ac.il. Research supported in part by a USA-Israeli BSF grant, by an ISF grant, by the Israeli I-Core program and by the Oswald Veblen Fund.

1.1 Regular tournaments with many directed Hamilton cycles

A tournament T on n vertices is an orientation of the complete graph K_n. It is regular if all the vertices have the same outdegree $(n-1)/2$ (and hence also the same indegree $(n-1)/2$.) A Hamilton cycle in T is a simple directed cycle containing all its vertices. It is clear that there exists a tournament on n vertices with at least $(n-1)!/2^n$ Hamilton cycles, as this is the expected number of such cycles in a random tournament. Thomassen asked in 1990 whether or not there is a regular tournament on n vertices with at least that many Hamilton cycles. This question appears as Problem 7.12 in Bondy's Chapter [15]. The following observation shows that the answer is "yes". The proof for n which is 1 or 3 modulo 6 is a bit simpler, hence we describe the proof for this case and only comment on the remaining case $n \equiv 5 (mod\ 6)$.

Theorem 1.1 *For any integer $n > 3$ which is 1 or 3 modulo 6 there is a regular tournament on n vertices with more than $(n-1)!/2^n$ Hamilton cycles.*

Proof: Take a Steiner Triple System on n vertices and orient each triangle of it cyclically, where each direction is chosen randomly among the two possible options. This clearly gives a regular tournament. Fix an undirected Hamilton cycle among the $(n-1)!/2$ such cycles in the underlying undirected complete graph. If it contains no two edges of the same triangle of the Steiner Triple System, then the probability it becomes a directed Hamilton cycle is exactly $1/2^{n-1}$. In any other case the probability is bigger. Hence, by linearity of expectation, the expected number of directed Hamilton cycles is bigger than $(n-1)!/2^n$, completing the proof. $\square$

Note that in fact, using the method in [1] one can show, using Brun's sieve, that the expected number of directed Hamilton cycles in the random regular tournament above is bigger by a factor of roughly $e = 2.71828..$ than the number $(n-1)!/2^n$, but this is not needed for establishing the above claim. It is also possible to prove the statement of the claim for $n \equiv 5(\ mod\ 6)$ by considering a slightly more complicated model of random regular tournaments. We omit the details.

2 Larger homometric sets in graphs

For a graph $G = (V, E)$ and a subset U of V, the *profile* of U is the multiset of pairwise distances (in G) between the vertices in U. Two disjoint subsets of V are *homometric* if their profiles are identical. It is known that for the cycle of length $2n$ and for any partition of it into two disjoint sets, these two sets are homometric (see [29]). In music theory, for $n = 6$, this statement had been known for a long time as the Hexachordal Theorem. In the twelve-tone scale, any set of six notes determines the same multiset of differences, see [10] and its references for some additional information. More generally, as shown in [2], any subset of n vertices in a vertex transitive graph with $2n$ vertices and its complement are homometric.

Following [10], denote by $h(n)$ the largest possible integer h so that any connected graph G with n vertices contains two disjoint homometric subsets, each of size h. Albertson, Pach and Young proved that there exists an absolute constant $c > 0$ so that for every $n > 3$

$$c\frac{\log n}{\log \log n} \le h(n) < n/4.$$

The upper bound has been slightly improved by Axenovich and Özkahya [12] who showed that for infinitely many values of n, $h(n) \leq n/4 - c \log\log n$. There are better estimates for trees, see [12], [20], but the best known lower bound for the function $h(n)$ is still only $c\frac{\log n}{\log\log n}$. The following theorem provides a quadratic improvement of this lower bound. Note that this is still very far from the known upper bound, which is linear in n.

Theorem 2.1 *There is an absolute constant $c > 0$ so that*

$$h(n) \geq c\frac{(\log n)^2}{(\log\log n)^2}.$$

Proof: Throughout the proof we make no attempt to optimize the absolute constants involved. To simplify the presentation we omit all floor and ceiling signs whenever these are not crucial, and assume that n is sufficiently large whenever this is needed. Let $G = (V, E)$ be a connected graph on n vertices. Define

$$k = \frac{1}{112}\frac{(\log n)^2}{(\log\log n)^2} \quad \text{and} \quad r = \frac{1}{14}\frac{\log n}{\log\log n}.$$

To prove the theorem we show that G must contain two disjoint homometric sets, each of size k. If the diameter of G is at least $2k-1$ then it contains two vertices u, v and a shortest path $u = u_1, u_2 \ldots, u_{2k} = v$ between them. In this case the distance between u_i and u_j is $|i-j|$ for all i, j and the two sets $\{u_1, u_2, \ldots, u_k\}$ and $\{u_{k+1}, u_{k+2}, \ldots, u_{2k}\}$ are homometric. Thus we may and will assume that the diameter of G is at most $2k-2$.

Claim: The graph G contains $\frac{\sqrt{n}}{k}$ pairwise disjoint sets of vertices S_i, each of size k, where the distance (in G) between any two vertices of each set S_i is at most $2r$.

Proof of Claim: We prove the existence of sets S_i with the required properties. Suppose we already have pairwise disjoint sets $S_1, \ldots, S_p$, where $0 \leq p < \frac{\sqrt{n}}{k}$ so that each S_i is of cardinality k and for each i the distance in G between any two vertices of S_i is at most $2r$. We proceed by proving the existence of another set $S_{p+1} \subset V - (\cup_{i=1}^p S_i)$ satisfying the required properties. Put $Y = \cup_{i=1}^p S_i$ and note that $|Y| < \sqrt{n}$. Consider an auxiliary graph G' obtained by adding to G a new vertex y connected by edges to all vertices of Y. Since the diameter of G is at most $2k-2$, the distance in G' between y and any other vertex of G is at most $2k-1$. Let V_j be the set of all vertices of distance exactly j from y. Then $V_0 = \{y\}, V_1 = Y$, $V_j = \emptyset$ for all $j \geq 2k$ and $\cup_{j=1}^{2k-1} V_i = V$. As $|Y| < \sqrt{n}$ and $|V| = n$, this implies that there is some i, $0 \leq i < \lceil (2k-2)/r \rceil$, so that

$$\frac{|\cup_{j=1}^{ir+r+1} V_j|}{|\cup_{j=1}^{ir+1} V_j|} \geq (\sqrt{n})^{r/(2k)}. \tag{1}$$

Indeed, if all these ratios are smaller than the right hand side of (1), their product, which is $|V|/|Y| \geq \sqrt{n}$ is too small.

Fix an i so that (1) holds and note that by our choice of parameters this implies that

$$\frac{|\cup_{j=1}^{ir+r+1} V_j|}{|\cup_{j=1}^{ir+1} V_j|} \geq k+1. \tag{2}$$

Consider now a BFS tree in G' rooted at the new vertex y. The level sets in this tree are exactly the sets V_j. By (2) the total number of vertices in the r levels $ir+2, ir+3, \ldots ir+r+1$ is at least

k times larger than the number of vertices in level number $ir+1$. Thus there must be a vertex v in level $ir+1$ which has at least k descendents in these r levels. Let S_{p+1} be a set of k of these descendents. Clearly S_{p+1} does not contain any vertex of $Y = V_1$ and thus does not intersect any of the previous sets $S_1, \ldots, S_p$. In addition, the distance between any of these descendents and v is at most r, and hence the distance between any two vertices of S_{p+1} is at most $2r$. This completes the proof of the claim. □

Returning to the proof of the Theorem observe that the profile of each set S_i is a multiset of exactly $K = \binom{k}{2}$ integers, each between 1 and $2r$. The number of possibilities for such collections of numbers is exactly

$$\binom{K+2r-1}{2r-1} < \binom{k^2}{2r} < n^{3/7} < \frac{\sqrt{n}}{k}.$$

Thus, by the pigeonhole principle, two of the sets S_i must have the same profile, completing the proof. □

3 Steiner systems with distinct domination numbers

3.1 The main result

Let $D = (V, E)$ be a Steiner Triple System (STS, for short) on a set V of v vertices. Thus E is a collection of $v(v-1)/6$ triples of V, so that each pair of vertices of V belongs to exactly one triple. Let $\gamma(D)$ denote the domination number of the incidence graph of D. This is the smallest possible cardinality of a set $V' \cup E'$ such that $V' \subset V, E' \subset E$ and
(i) for each $u \in V - V'$ there is an $e \in E'$ containing it, and
(ii) for each $e \in E - E'$ there is a $u \in V'$ which belongs to e.

Call such a set $V' \cup E'$ a dominating set of D, and call $\gamma(D)$ the domination number of D. This parameter was defined by Goldberg, Rajendraprasad and Mathew, who raised the following conjecture.

Conjecture 3.1 ([22]) *For every integer v, all the Steiner Triple Systems on v vertices have the same domination number.*

The conjecture has been verified for all $v \leq 15$ by a computer search, see [22] for the values of $\gamma(D)$ for all STS D of size at most 15.

Here we disprove the general conjecture, establishing the following

Theorem 3.2 *For any $\epsilon > 0$ there are infinitely many values of v so that there are two Steiner Triple Systems D_1 and D_2, each on v vertices, satisfying $\gamma(D_1) \leq 3v/4$ and $\gamma(D_2) \geq (1-2\epsilon)v$.*

3.2 The proof

The proof of Theorem 3.2 is not long, but applies several results, which we state below. A variant of the proof can be given based on known results for completion of partial Steiner Triple Systems, but we prefer the version given below, which is based on slightly simpler tools. Before stating these tools we describe a simple relation between the domination number of an STS and its independence number. The independence number of an STS is the maximum cardinality of a set of vertices that contains no edge.

Claim 1 *For any STS $D = (V, E)$ on v vertices with domination number $\gamma = \gamma(D)$ and independence number $\beta = \beta(D)$: $v - \beta \le \gamma \le v - \lfloor \frac{\beta}{2} \rfloor$.*

Proof: To prove the upper bound let $V' \subset V$ be the complement of a maximum independent set in V. Thus $|V'| = v - \beta$ and any $e \in E$ contains at least one vertex in V'. Next cover the vertices in $V - V'$ by $\lceil \beta/2 \rceil$ pairs, and for each such pair take an edge in E containing both its members. Let E' be the collection of these $\lceil \beta/2 \rceil$ edges, and note that $V' \cup E'$ is a dominating set of cardinality $v - \lfloor \frac{\beta}{2} \rfloor$.

To prove the lower bound, let $V' \cup E'$ be a dominating set of minimum cardinality, where $V' \subset V$ and $E' \subset E$. For each edge $e \in E'$, take an arbitrary vertex $u \in e$ and add it to V', thus getting a set V" of at most $|V'| + |E'| = \gamma$ vertices. Since $V' \cup E'$ is dominating, $V - V$" must be an independent set. Therefore $\beta \ge |V - V"| \ge v - \gamma$, completing the proof. $\square$

We need the following well known theorem of Wilson.

Theorem 3.3 ([37]) *For any integer k there is a $v_0 = v_0(k)$ so that for every integer $v > v_0$ for which $v - 1$ is divisible by $k - 1$ and $\binom{v}{2}$ is divisible by $\binom{k}{2}$, the complete graph K_v on v vertices can be decomposed into pairwise edge disjoint copies of K_k. That is: there is a collection of subgraphs of K_v, each being a copy of K_k, so that each edge of K_v belongs to exactly one copy.*

We also use the following.

Lemma 3.4 *(i) For any integer v of the form $v = 2^n - 1$ there is an STS $D = (V, E)$ on v vertices with independence number $\beta(D) = (v + 1)/2$.*

(ii) For any $\epsilon > 0$ there is an $r_0 = r_0(\epsilon)$ so that for any $r > r_0$ there is an STS D on $k = 3^r$ vertices with independence number satisfying $\beta(D) \le 2\epsilon k$.

Proof: (i) Let the vertices be the set of all nonzero binary vectors of length n, where three of them x, y, z form a triple iff their sum modulo 2 is the zero vector. This is easily seen to be an STS, and the set of all $2^{n-1} = (v + 1)/2$ vectors with first coordinate being 1 forms an independent set. Thus $\beta(D) \ge (v + 1)/2$. However, $\beta(D)$ cannot be bigger. Indeed, let X be an independent set in a STS $D = (V, E)$ on v vertices. Pick $x \in X$. For any $y \in X - \{x\}$ the pair xy is contained in a unique edge $\{x, y, y'\}$, where $y' \notin X$, as X is independent. Since no pair of points lies in two edges, the mapping $f(y) = y'$ is bijective, implying that $|V - X| \ge |X| - 1$, that is, $|X| \le (v + 1)/2$.
(ii) Here the vertices are all ternary vectors of length r, where three of them form a triple if their sum modulo 3 is the zero vector. Equivalently, $\{x, y, z\}$ form a triple iff they are an arithmetic progression of size 3 (in Z_3^n). Here, too, it is easy to see that this is an STS. The claim about its independence number was first proved in [13], see also [19] for a related proof and [30], [16] for better quantitative estimates. $\square$

Finally we will use the known results about the existence of nearly perfect matchings in simple uniform regular hypergraphs. The basic result was proved by Rödl using his nibble method [31], here it is convenient to use the following subsequent result. Recall that a hypergraph is simple iff no two of its edges share two vertices.

Lemma 3.5 ([3]) *Any simple k-uniform, D-regular hypergraph on N vertices, with $k > 3$, contains a matching covering all vertices but at most $O(ND^{-1/(k-1)})$.*

Proof of Theorem 3.2: Given $\epsilon > 0$, let r be a large odd integer and define $k = 3^r$. By Lemma 3.4, part (ii) if r is sufficiently large there is an STS F on k vertices satisfying $\beta(F) \leq \epsilon k$. Note that since r is odd, $k = 3^r \equiv 3 \pmod 4$ and hence $(k-1)/2 = (3^r - 1)/2$ is odd. Let $\phi = \phi(\binom{k}{2})$ be the Euler function of $K = \binom{k}{2}$, that is, the cardinality of the multiplicative group Z_K^* of the residues modulo K which are relatively prime to K. Let n be a large integer so that ϕ divides $n-1$. Finally, put $v = 2^n - 1$. Note that as K is odd, $2 \in Z_K^*$ and as ϕ divides $n-1$, 2^{n-1} is 1 modulo K, that is, $K = \binom{k}{2}$ divides $2^{n-1} - 1$. Thus $k - 1 = 2\frac{k-1}{2}$, divides $v - 1 = 2\frac{v-1}{2} = 2(2^{n-1} - 1)$. Also, $K = \binom{k}{2}$ divides $\binom{v}{2} = (2^n - 1)(2^{n-1} - 1)$. Therefore, by Wilson's Theorem (Theorem 3.3 above) if n is sufficiently large then K_v can be decomposed into copies of K_k. Let D_2 be a STS on a set of v vertices obtained by substituting a copy of the STS F in each copy of K_k in this decomposition. Formally, if the set of vertices of K_v is V, then for each copy of K_k in the decomposition whose set of k vertices is $V_K \subset V$, take a copy of F on the set of vertices V_K and take all its edges to be edges of D_2. The triples of the STS D_2 are thus all the triples in the

$$\binom{v}{2} / \binom{k}{2}$$

copies of F described above.

Consider the k-uniform hypergraph on the set of vertices of K_v whose edges are all vertex sets of the complete graphs K_k in the decomposition. This hypergraph is simple and regular of degree $D = (v-1)/(k-1)$. Therefore, if n (and hence v and D) are sufficiently large then by Lemma 3.5 there is a matching M of this hypergraph covering at least $(1-\epsilon)v$ vertices. Note that this matching corresponds to a collection of pairwise disjoint copies of F in D_2.

We claim that $\beta(D_2) < 2\epsilon v$. Indeed, any independent set in D_2 can contain at most ϵk vertices in each copy of F in the matching M, and hence altogether at most an ϵ fraction of the vertices covered by this matching. As there are at most ϵv other vertices, the assertion of the claim follows.

By Lemma 3.4, part (i) there is also an STS D_1 on v vertices satisfying $\beta(D_1) = (v+1)/2$. By Claim 1 we thus conclude that $\gamma(D_1) \leq (3v-1)/4$ and $\gamma(D_2) \geq (1-2\epsilon)v$, completing the proof. $\square$

4 On complete decomposition graphs

An H decomposition of a graph G is a partition of the edges of G into parts each of which is an isomorphic copy of the graph H. The intersection graph of the decomposition is the graph whose vertices are the parts where two are connected iff they share at least one common vertex of G. Jamison [25] asked for the largest possible $s = s(d,k)$ so that the complete graph on s vertices is the intersection graph of an $H(k)$ decomposition of some d-regular graph, where $H(k)$ is a matching of size k. He showed that

$$s(d,k) \leq 2k(d-1) + 1, \tag{3}$$

proved that $s(3,2) = 9$, that is, equality holds in (3) for $d = 3, k = 2$, and asked to determine the cases in which equality holds. Here we answer this question and prove the following.

Theorem 4.1 *For every $d \geq 3, k \geq 2$ equality holds in (3) if and only if there is a K_d-decomposition of the complete graph on $2k(d-1)+1$ vertices.*

In particular, $s(3,k) = 4k+1$ iff there is a Steiner Triple system on $4k+1$ vertices, that is, iff k is 0 or 2 modulo 3, and for any fixed d and sufficiently large k, $s(d,k) = 2k(d-1)+1$ iff d divides $2k-4k^2$.

Proof: Put $s = 2k(d-1)+1$ and suppose there is a K_d-decomposition of K_s. This means that there is a collection $B_1, B_2 \ldots, B_m$ of subsets of a set S of size s, where $|B_p| = d$ for each p and every pair of distinct vertices $i, j \in S$ are contained together in a unique block B_p. Construct a d-regular graph G on the set of vertices $M = \{1, 2, \ldots, m\}$ as follows. For each vertex $v \in S$ there are exactly $(s-1)/(d-1) = 2k$ blocks B_p containing v. Split them arbitrarily into k disjoint pairs, and for each pair B_p, B_q, let pq be an edge of G. Let M_v denote the matching of size k consisting of all these k edges. We have thus constructed s matchings, and the graph G consists of the edges of all of them. Note that the matchings are pairwise edge-disjoint, as there are no distinct u, v contained in two blocks B_p, B_q. Note also that G is d-regular, as each vertex p gets a contribution of one edge from each element $v \in B_p$. In addition, each pair of distinct matchings M_v, M_u has exactly one common vertex-the unique p so that $u, v \in B_p$. This shows that the existence of the design (that is, the K_d-decomposition of K_s) implies that $s(d,k) \geq 2k(d-1)+1$ which, in view of (3), shows that equality holds.

Conversely, assuming that equality holds in (3) we prove that there exists a K_d-decomposition of K_s. By assumption there is a d-regular graph G and a decomposition of its set of edges into matchings $M_1, M_2, \ldots, M_s$, $(s = 2k(d-1)+1)$, each matching of size k, so that every pair of matchings share at least one common vertex. Simple counting implies that each pair of matchings share exactly one common vertex (since altogether the vertices saturated by a matching M are incident with at most $2k(d-1)$ other edges, and if two of these edges belong to the same matching then M would not have enough room to share a vertex with all the other matchings). Put $S = \{1, 2, \ldots, s\}$. For each vertex v of G define a block $B_v \subset S$ as follows:

$$B_v = \{i : \ v \text{ is saturated by the matching } M_i\}.$$

Clearly each block is of size exactly d, as the graph is d regular. Moreover, every pair of elements $i, j \in S$ lie in a unique block, since by the above discussion for every two matchings M_i, M_j there is a unique vertex v saturated by both of them. It follows that the construction above is indeed a K_d-decomposition of K_s. This completes the proof of the first sentence in the statement the theorem. The assertion of the second sentence follows by the well known fact that there exists a Steiner triple system on v vertices iff v is either 1 or 3 modulo 6, and by Wilson's Theorem [37] on the existence of balanced incomplete block designs. (We omit the simple computation showing that the divisibility condition is indeed that d should divide $2k-4k^2$.)

5 List coloring of directed line graphs

The list chromatic number $\chi_\ell(H)$ of a graph H is the smallest integer k so that for any assignment of a list of k colors to each vertex of H there is a proper vertex coloring of the graph in which each vertex gets a color from its list. It is clear that $\chi_\ell(H) \geq \chi(H)$ for any graph H, and it is easy and well known that strict inequality may hold. See, e.g., [6] and its references for some background on list coloring.

Theorem 5.1 *Let $G = (V, E)$ be a (finite, loopless) directed graph with maximum total degree D and let $L = L(G)$ be the directed graph whose vertices are the directed edges of G, in which $(x, y) \in E$ is connected by a directed edge to (x', y') iff $y = x'$. If*

$$e(2D-1)(\frac{3}{4})^s < 1 \tag{4}$$

then the list chromatic number of L satisfies $\chi_\ell(L) < s$. Thus $\chi_\ell(L) \leq O(\log D)$.

Remark: If the average degree in G is at least $\Omega(D)$ (and in particular if G is regular Eulerian), then by the main result in [7] (see also [35] for a numerical improvement) the list chromatic number of $L(G)$ is at least $\Omega(\log_2 D)$. The chromatic number of $L(G)$ may be much smaller, and is well known to be $(1 + o(1)) \log_2(\chi(G))$, see, e.g., [5].

Proof of Theorem 5.1: Let G, L, D and s satisfy the assumptions of the theorem. For every vertex (x, y) of $L(G)$ let $S_{x,y}$ be a set of s colors assigned to this vertex. We have to show that there is a proper vertex coloring of $L(G)$ assigning to each vertex a color from its list. Let $S = \cup S_{x,y}$ be the set of all colors. For each vertex $v \in V$ of G, let $S = P_v \cup Q_v$, $P_v \cap Q_v = \emptyset$ be a random partition of the set of all colors into two disjoint sets chosen uniformly and independently among all partitions. We claim that with positive probability, for each $(x, y) \in E$

$$S_{x,y} \cap P_x \cap Q_y \neq \emptyset. \tag{5}$$

To prove this claim we apply the Lovász Local Lemma, proved in [18]. Let $B_{x,y}$ denote the event that (5) fails. It is clear that the probability of this event is $(3/4)^s$. In addition, each such event $B_{x,y}$ is mutually independent of all other events $B_{x',y'}$ besides those with $\{x', y'\} \cap \{x, y\} \neq \emptyset$, since all other events are determined by the random partitions (P_v, Q_v) for $v \notin \{x, y\}$. It follows that each event is mutually independent of all others besides at most $2D - 2$, and the desired claim follows from the Local Lemma.

To complete the proof fix partitions P_v, Q_v for which (5) holds for each $(x, y) \in E$, and color (x, y) by an arbitrary color c from $S_{x,y} \cap P_x \cap Q_y$. This is clearly a proper coloring, since if c is the color of (x, y) and c' is the color of (y, z) then $c \in Q_y$ whereas $c' \in P_y$, which is disjoint from Q_y, implying that $c \neq c'$. This completes the proof. $\square$

6 Intersections of sets with their shifts

Let $[n] = \{1, 2, \ldots, n\}$ and let $A \subset [n]$ satisfy $|A| = m > 1$. For a positive integer i put $A + i = \{a + i : a \in A\}$.

Theorem 6.1 *For any integer k there is some $i > 0$ so that*

$$|A \cap (A + i)| \geq \frac{km(km - n - k + 1)}{(n + k - 1)k(k - 1)}.$$

Remark: The above is nearly tight. Indeed, by a theorem of Singer [33], for any prime power q and for any $d \geq 2$ there is a subset S of cardinality $m = q^{d-1}+q^{d-2}+\ldots+1$ of $[n]$, for $n = q^d+q^{d-1}+\ldots+1$, so that the intersection of S with any shifted copy of itself (and in fact even with any cyclic shifted copy) is of cardinality at most $q^{d-2}+\ldots+1$. In particular note that this gives sets of size $m = q+1$ in $[n]$ for $n = q^2+q+1$ (that is, $m > \sqrt{n}$) so that all intersections are of size at most 1. It follows that the case $m \leq (1-o(1))\sqrt{n}$ is clear (as there are enough prime powers to get close to $\sqrt{n}$ for any n by a number of the form $q+1$). One can use other constructions of difference sets as well. For $m > (1-o(1))\sqrt{n}$ we can choose an optimal k in the theorem above. Without trying to optimize note that it is clear that by taking k so that $n = o(km)$, $k = o(n)$ and $k >> 1$ (for example, $k = n^{3/4}$ will always do), we get that there is always an intersection of size at least $(1+o(1))\frac{m^2}{n}$. For specific values of m and n we can optimize more carefully. In addition, if $m = o(n)$ and $\sqrt{n \log n} = o(m)$ then the $(1+o(1)))\frac{m^2}{n}$ estimate above is tight, as shown by a random subset of cardinality m in $[n]$.

Proof: Define $x = \max\{|A \cap (A+i)| : 1 \leq i \leq k-1\}$. Note that for any $0 \leq i < j \leq k-1$, $|(A+i) \cap (A+j)| \leq x$, since clearly $|(A+i) \cap (A+j)| = |A \cap (A+j-i)|$. For each integer p, $1 \leq p \leq n+k-1$, let d_p denote the number of indices i, $0 \leq i \leq k-1$, so that $p \in A+i$. Clearly $\sum_{p=1}^{n+k-1} d_p = km$, since any set $A+i$ contributes to exactly m d_p-s. By the convexity of the function $f(z) = \binom{z}{2} = z(z-1)/2$ this implies that

$$\sum_{p=1}^{n+k-1} \binom{d_p}{2} \geq (n+k-1)\frac{(\frac{km}{n+k-1})(\frac{km}{n+k-1}-1)}{2} = \frac{km(km-n-k+1)}{2(n+k-1)}.$$

On the other hand the sum $\sum_{p=1}^{n+k-1} \binom{d_p}{2}$ is exactly the sum of the quantities $|(A+i) \cap (A+j)|$, over all pairs i, j with $0 \leq i < j \leq k-1$. Indeed, both these sums count precisely the number of triples (p, i, j) with $1 \leq p \leq n+k-1$, $0 \leq i < j \leq k-1$ and $p \in (A+i) \cap (A+j)$.

By the definition of x the final sum is at most $x\binom{k}{2}$, implying that

$$x\binom{k}{2} \geq \frac{km(km-n-k+1)}{2(n+k-1)},$$

and completing the proof. □

Note that for specific values of m and n we can optimize further by using the fact that in the above proof the value of d_p for small or large indices p cannot be large (namely, $d_p \leq p$ for all p and similarly $d_{n+k-p} \leq p$ for all p). This, however, does not change the asymptotic estimate.

7 Long paths in graph orientations

For a digraph D, let $\ell(D)$ denote the maximum number of vertices of a directed simple path in D. For an undirected graph G and an integer $j \geq 0$, let $\ell_j(G)$ denote the minimum possible value of $\ell(D)$ when the minimum is taken over all orientations D of G in which every outdegree is at least j. (If there is no such orientation define $\ell(G) = 0$). It is well known that for any graph G $\ell_0(G) = \chi(G)$. This was proved independently by Gallai [21], Roy [32], Hasse [23] and Vitaver [36]. Hod and Naor [24] showed that for any d-regular graph G on n vertices $\ell_1(G) \leq O(\log n / \log\log d)$ and that there are d-regular graphs on n vertices for which $\ell_1(G) \geq \Omega(\log n/\log d)$. They also raised the problem of estimating $\ell_j(G)$ for larger values of j. Here we show that already for $j = 2$ and any fixed $d \geq 4$ there are d regular graphs G on n vertices for which $\ell_2(G) \geq \Omega(n)$.

Theorem 7.1 *For any fixed $d \geq 4$ there is a d-regular graph G on n vertices so that $\ell_2(G) \geq \Omega(n)$, that is, in any orientation of G in which each outdegree is at least 2, the length of the longest directed path is at least $\Omega(n)$.*

To prove the above we establish the following.

Lemma 7.2 *Let $G = (V, E)$ be a graph in which any set of $z \leq z_0$ vertices spans at most $(j+1)z/2-1$ edges, then $\ell_j(G) \geq z_0/2$.*

Proof: Let D be an orientation of G in which every outdegree is at least j. Consider a run of DFS on D. During the algorithm, each vertex is colored white, gray or black. Initially all vertices are white, and in each step at most one vertex changes a color, where white vertices can become gray, and gray can become black. During the algorithm the gray vertices form a directed path, and there are no directed edges from a black vertex to a white vertex. At the end all vertices are black. Consider the algorithm when exactly $x = z_0/2$ vertices are black. Let y be the number of gray vertices at that point. Then the induced subgraph of G on the set of all black and gray vertices contains at least $jx + y - 1$ edges (at least j outgoing edges from each black vertex and at least $y - 1$ edges between gray vertices). If $y < x$ this gives a set of $z = x + y$ vertices with more than $(j + 1)z/2 - 1$ edges, contradiction. $\square$

The lemma and the known results about the distribution of edges in random d-regular graphs that imply that with high probability all induced subgraphs on small linear size sets in such graphs have average degree at most $2 + \delta < 3$, (c.f., e.g., Lemma 6 in [24]), imply the statement of the theorem.

8 The cover number of sign matrices

The ϵ-cover number $N_\epsilon(A)$ of an m by n matrix A is the minimum possible cardinality of a set S of vectors in R^m so that any point in the convex hull of the columns of A is within ℓ_∞-distance at most ϵ from some point of S. This notion is considered in [4], where it is shown, as a corollary of a more general result, that for any n by n sign-matrix A,

$$N_\epsilon(A) \leq n^{O(\log n/\epsilon^2)}.$$

Here we show that this estimate is tight, up to the hidden constant factor in the exponent (when n is at least $1/\epsilon^7$, say). It is convenient to consider m by n matrices, with $m = n^6$ (this clearly only changes the constant, as any such matrix can be embedded in an m by m matrix, and $\log n^6 = 6 \log n$.)

Let A be a random $m = n^6$ by n sign matrix.

Theorem 8.1 *With high probability (whp, for short)*

$$N_\epsilon(A) \geq n^{\Omega(\log n/\epsilon^2)}.$$

To prove the theorem it suffices to show that whp there is a set T of at least $n^{\Omega(\log n/\epsilon^2)}$ vectors in R^m that lie in the convex hull of the columns of A, so that the ℓ_∞ distance between any two of these vectors exceeds ϵ. Let $\mathcal{F}$ be a family of

$$t \geq n^{(1-o(1))\frac{\log n}{\epsilon^2}}$$

subsets of $[n] = \{1, 2, \ldots, n\}$, so that the cardinality of each member of $\mathcal{F}$ is $\frac{2}{\epsilon^2} \log n$, and the intersection of any two distinct members of $\mathcal{F}$ is at most $\frac{1}{\epsilon^2} \log n$. The existence of such an $\mathcal{F}$ is proved by a simple greedy procedure, picking sets one by one and omitting all those that intersect a chosen set by too many elements.

Let A_j denote column number j of A. For each $F \in \mathcal{F}$, let v_F denote the average of the columns A_j for $j \in F$. It suffices to show that whp the ℓ_∞-distance between any pair of vectors v_F and $v_{F'}$, for $F, F' \in \mathcal{F}, F \neq F'$, exceeds ϵ. For this it suffices to show that whp, for any two such sets F, F', there is a row i so that the number of indices j with $j \in F - F'$ and $A_{ij} = 1$ exceeds the number of indices j with $j \in F' - F$ and $A_{ij} = 1$, by more than $\frac{1}{\epsilon} \log n$. Fix a pair of sets $F, F' \in \mathcal{F}$. By standard estimates of Binomial distributions (see, e.g. [11], Appendix A), the probability that for a fixed i the above happens exceeds $\frac{1}{n^5}$, with room to spare. (Indeed, it suffices to ensure that the number of $+1$ entries in $F - F'$ exceeds its expectation by at least $\frac{0.5}{\epsilon} \log n$, and the same occurs to the number of -1 entries in $F' - F$.) Therefore, the probability that there is no row i in which this happens is smaller than

$$(1 - \frac{1}{n^5})^{n^6} \leq e^{-n}$$

and this number is sufficiently small to ensure, by the union bound, that whp there is such a coordinate for any pair of distinct sets $F, F' \in \mathcal{F}$. This completes the proof of the theorem. $\square$

Recall that the Vapnik-Chervonenkis dimension of a sign (or binary) matrix A, denoted by $VC(A)$, is the maximum cardinality of a set J of columns of A so that for each $F \subset J$ there is a row i of A such that for $j \in J$, $A_{ij} = 1$ iff $j \in F$. It will be interesting to decide if the more general result proved in [4], that asserts that

$$N_\epsilon(A) \leq n^{O(d/\epsilon^2)},$$

where $d = VC(A)$, is also tight for all admissible values of n, ϵ and $VC(A)$. It is not difficult to get a lower bound of $n^{\Omega(d/\epsilon)}$. Indeed, let A be the $\binom{n}{d}$ by n sign-matrix in which the rows are indexed by d-subsets of $[n]$, and each row has -1 in all entries corresponding to its set. Let $\mathcal{F}$ be a family of subsets of $[n]$, each of size $\frac{d}{\epsilon}$, where for each two distinct $F, F' \in \mathcal{F}$, $|F - F'| \geq d$ and $|\mathcal{F}| \geq n^{\Omega(d/\epsilon)}$. Then the averages v_F as defined above, for $F \in \mathcal{F}$ satisfy $||v_F - v_{F'}||_\infty > \epsilon$ for all distinct $F, F' \in \mathcal{F}$, as shown by considering the row indexed by a d-set that lies in $F - F'$.

For $d = 1$ the $n^{\Omega(1/\epsilon)}$ estimate is, in fact, tight, as we prove next. The proof applies the known characterization of spaces with VC-dimension 1. We suspect, however, that the behavior for $d > 1$ may well be different, as spaces with VC-dimension 2 are far more complicated. It is more convenient to state and prove the theorem for binary matrices rather than sign matrices and prove first a result for totally unimodular matrices, which may be interesting in its own right.

A matrix A is totally unimodular (TU, for short) if the determinant of any square submatrix of it (of any dimension) lies in $\{0, -1, 1\}$. In particular , all entries of such a matrix must lie in $\{0, -1, 1\}$.

Theorem 8.2 *Let A be a TU matrix with n columns. Then for any $\epsilon > 0$,*

$$N_\epsilon(A) \leq \binom{n + \lceil 1/\epsilon \rceil - 1}{\lceil 1/\epsilon \rceil}.$$

Proof: Let $A = A(u, v),\ u \in U, v \in V$. Let $\{z_v : v \in V\}$ denote the column vectors of A, and let C denote their convex hull. Put $k = \lceil 1/\epsilon \rceil$ and let $\mathcal{F}$ denote the set of all averages of k columns of A (with repetitions). Then $|\mathcal{F}| = \binom{n+k-1}{k}$ and it suffices to show that for every vector $z \in C$ there is a vector $y \in \mathcal{F}$ so that $||y - z||_\infty \leq 1/k$. Let $z = \sum_{v \in V} x_v z_v$ be a vector in C, where $x_v \geq 0$ for all v and $\sum_{v \in V} x_v = 1$. Define $x'_v = k x_v$, then

$$kz = \sum_{v \in V} x'_v z_v, \quad x'_v \geq 0 \ \text{ for all } \ v \in V \ \text{ and } \ \sum_{v \in V} x'_v = k \tag{6}$$

Consider the following linear program in the variables $x'_v,\ v \in V$:

$$\lfloor kz(u) \rfloor \leq \sum_{v \in V} x'_v z_v(u) \leq \lceil kz(u) \rceil \ \text{ for all } \ u \in U, \ \ x'_v \geq 0 \ \text{ for all } \ v \in V \ \text{ and } \ \sum_{v \in V} x'_v = k \tag{7}$$

By (6) this has a feasible solution x'_v. Since the matrix A is totally unimodular, the linear program (7) admits an integral solution x^*_v. It follows that

$$|z(u) - \frac{1}{k} \sum_{v \in V} x^*_v z_v(u)| \leq \frac{1}{k} \leq \epsilon$$

for all $u \in U$. Since the vector $y = \frac{1}{k} \sum_{v \in V} x^*_v z_v \in \mathcal{F}$, this completes the proof of the theorem. $\square$

Remark: The assertion of the last theorem is tight, as shown by the identity matrix A with n columns. Indeed, if $\epsilon < 1/k$ then the set of all averages of k columns of A (with repetitions) contains no two members within ℓ_∞ distance smaller than $1/k$, implying that here

$$N_\epsilon(A) \geq \binom{n + k - 1}{k}.$$

For a binary matrix $A = (a_{ij})$, the complement of A is the binary matrix $\overline{A} = (1 - a_{ij})$. Call a binary matrix TU-complement (TUC, for short) if it is the complement of a totally unimodular binary matrix. Finally, call an m by n matrix A a double-TU matrix, if it is possible to split its columns into two disjoint sets, thus splitting A into two matrices A_1 and A_2, where A_1 has m rows and n_1 columns, A_2 has m rows and n_2 columns, with $n = n_1 + n_2$, and A_1 is a TU-matrix while A_2 is a TUC-matrix. We allow one of these matrices to be empty and therefore any TU or TUC matrix is double-TU as well. The following result is proved in [27].

Lemma 8.3 ([27]) *A binary matrix A has VC-dimension 1 if and only if it is possible to reduce it to the empty matrix by repeatedly applying operations of the following two forms:*

(i) Delete a row identical to another one.

(ii) Delete a column with at most one 0 or at most one 1.

Corollary 8.4 *Any binary matrix A with VC-dimension 1 is a double TU-matrix.*

Proof: By the above Lemma and since any binary matrix with at most one row or column is TU (and hence double TU too), it suffices to show that if A is double TU and A' is obtained from A either by adding to it a row identical to one of the rows of A or by adding to it a column with at most one 0 or at most one 1, then A' is also double TU. Suppose, thus, that A is a double TU matrix with n columns. Without loss of generality assume that the matrix A_1 consisting of the first n_1 columns of A is TU, and the matrix A_2 consisting of the last n_2 columns of A is TUC, where $n = n_1 + n_2$. If A' is obtained from A by adding to it a row identical to one of the rows of A, then it is easy to check that the matrix consisting of the first n_1 columns of A' is TU, and the one consisting of the last n_2 columns of A' is TUC, as needed. If A' is obtained from A by adding a column with at most one 1, then we can add this column to A_1 to get A'_1, which is still TU (as the determinant of any square submatrix of A'_1 that contains part of this column can be expanded with respect to this column.) Similarly, if A' is obtained from A by adding a column with at most one 0, then we can add this column to A_2 keeping it TUC. This completes the proof. □

We can now prove the following.

Theorem 8.5 *Let A be a binary matrix with VC-dimension 1 and n columns. Then*

$$N_\epsilon(A) \le n^{O(1/\epsilon)}.$$

Proof: By the above Corollary A is double TU. Assume, without loss of generality, that the matrix A_1 consisting of the first n_1 columns of A is TU, and the matrix A_2 consisting of the last n_2 columns of A is TUC, where $n = n_1 + n_2$. By Theorem 8.2

$$N_{\epsilon/2}(A_1) < n^{2/\epsilon}.$$

Similarly, for the complement B of A_2 we have, by Theorem 8.2,

$$N_{\epsilon/2}(B) < n^{2/\epsilon}.$$

As each vector v in the convex hull of the columns of the complement B is simply the vector $j - u$ for some vector u in the convex hull of A_2, where j is the all 1 vector, it follows that

$$N_{\epsilon/2}(A_2) = N_{\epsilon/2}(B) < n^{2/\epsilon}.$$

Finally note that this implies that

$$N_\epsilon(A) \le n^{O(1/\epsilon)}.$$

This is because for every pair of matrices $\overline{A}_1$ and $\overline{A}_2$ we have

$$N_\epsilon(\overline{A}_1 + \overline{A}_2) \le N_{\epsilon/2}(\overline{A}_1) N_{\epsilon/2}(\overline{A}_2).$$

Indeed, one can just take the pairwise sum of the corresponding $\epsilon/2$-nets and use the triangle inequality. As the cover numbers of the matrices A_1, A_2 do not increase much by appending to them 0 columns to get matrices $\overline{A}_1$ and $\overline{A}_2$ whose sum is A, the desired estimate follows, completing the proof. □

9 Orders with the running intersection property

9.1 The main result

A sequence of subsets $F_1, F_2, \ldots, F_m$ of a finite set X satisfies the **Running Intersection Property** (RIP) if for every $k > 1$ the intersection of F_k with the union of all previous F_j is contained in one of these previous subsets, that is,

$$\text{For every } k > 1 \text{ there is an } i < k \text{ so that } F_k \cap (\cup_{j<k} F_j) \subset F_i. \tag{8}$$

A family of subsets $\mathcal{F}$ of X satisfies RIP^* if there is an ordering of its members that satisfies RIP. For more on the Running Intersection Property see, e.g., [28], [17] and the references therein.

A sequence of subsets $F_1, F_2, \ldots, F_m$ of a finite set X is **expansive** if for every $k > 1$, the cardinality of the intersection of F_k with the union of all previous sets is at least as large as the cardinality of the intersection of any subset that appears after F_k with this union, that is,

$$\text{For every } k > 1,\ |F_k \cap (\cup_{j<k} F_j)| \geq |F_i \cap (\cup_{j<k} F_j)| \text{ for all } i > k. \tag{9}$$

R. Spiegler [34] conjectured that if a family of subsets satisfies RIP^*, then any expansive ordering of its members satisfies RIP. This is proved in the following theorem.

Theorem 9.1 *Let $\mathcal{F}$ be a family of subsets of a finite set X. If $\mathcal{F}$ satisfies RIP^*, then any expansive ordering of its members satisfies RIP.*

Note that the above theorem supplies a simple efficient algorithm for checking if a given family $\mathcal{F}$ satisfies RIP^*: we simply produce an expansive ordering of it and check if it satisfies (8).

9.2 The proof

A family of subsets $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ of a finite set satisfies the **Tree Decomposition Property** (TDP) if there is a tree T on the set of vertices $\mathcal{F}$ so that for any $F_i, F_j, F_k \in \mathcal{F}$ where F_i is on the unique path in T from F_k to F_j, $F_k \cap F_j \subset F_i$. If this is the case we say that T, with an arbitrary vertex of it designated as a root, is a **realization for** $\mathcal{F}$. Note that there can be many trees realizing the same family $\mathcal{F}$.

An ordering of the vertices of a rooted tree is called **admissible** if the root appears first, and any other vertex appears after its unique parent in the tree. The following simple lemma appears, in various forms, in the literature, see, e.g., [28], Chapter 2. For completeness we include a short proof.

Lemma 9.2 *Let $\mathcal{F}$ be a family of subsets of a finite set X. Then $\mathcal{F}$ satisfies RIP^* if and only if it satisfies TDP. Moreover, if $\mathcal{F}$ satisfies TDP and T is a rooted tree realizing $\mathcal{F}$, then any admissible ordering of its members satisfies RIP.*

Proof: Assume first that $\mathcal{F}$ satisfies RIP^*. Then there is a sequence $F_1, F_2, \ldots, F_m$ of the members of $\mathcal{F}$ such that (8) holds. Let T be a rooted tree on the set of vertices $\mathcal{F}$, where F_1 is the root, and for each $k > 1$ the unique parent of F_k in the tree is an arbitrarily chosen F_i so that $i < k$ and

$$F_k \cap (\cup_{j<k} F_j) \subset F_i.$$

We claim that T is a realization for $\mathcal{F}$. Indeed, suppose $F_i, F_j, F_k \in \mathcal{F}$, with F_i being on the unique path between F_j and F_k. We have to show that $F_k \cap F_j \subset F_i$. This is proved by induction on the distance in T between F_k and F_j. If the distance is 0 or 1 there is nothing to prove as in this case F_i is either F_j or F_k. Assuming the assertion holds for distance smaller than d, suppose the distance between F_j and F_k is $d \geq 2$. Without loss of generality suppose that $j < k$. Let F_s be the unique parent of F_k in T. By the construction of T, $F_k \cap F_j \subset F_s$. It is also clear that F_s is on the unique path in T from F_k to F_j (since F_j is not a descendent of F_k). Thus either $s = i$, and then the desired result $F_k \cap F_j \subset F_i$ holds, or F_i is on the unique path in T between F_s and F_j. In the latter case, as the distance between F_s and F_j is $d-1$ it follows, by the induction hypothesis, that $F_s \cap F_j \subset F_i$, completing the proof of the claim, as $F_k \cap F_j \subset F_s \cap F_j \subset F_i$.

Conversely, suppose that $\mathcal{F}$ satisfies TDP, and let T be a rooted tree which forms a realization for $\mathcal{F}$. Let F_1 be the root, and let $F_1, F_2, \ldots, F_m$ be an admissible order of the members of $\mathcal{F}$. We complete the proof of the lemma by showing that this ordering satisfies RIP. For $k > 1$, let F_i be the unique parent of F_k in T. Since the order is admissible $i < k$. In addition, F_i lies on the unique path between F_k and F_j for any $j < k$, as no such F_j is a descendent of F_k in T. As T satisfies TDP it follows that $F_k \cap F_j \subset F_i$ for each $j < k$, and hence $F_k \cap (\cup_{j<k} F_j) \subset F_i$, as needed. □

Proof of Theorem 9.1: Let $\mathcal{F}$ be a family of subsets of a finite set X and suppose it satisfies RIP^*. Let $F_1, F_2, \ldots, F_m$ be an expansive order of the members of $\mathcal{F}$. We have to show that this ordering satisfies RIP. To do so we prove the following:

Claim: For every $k \geq 1$ there is a tree that forms a realization for $\mathcal{F}$ so that $F_1, F_2, \ldots, F_k$ is an initial segment in an admissible ordering of the vertices of the tree.

Note that the case $k = m$ of the above lemma implies the assertion of the theorem, as it provides a realization for $\mathcal{F}$ in which the sequence $F_1, F_2, \ldots, F_m$ is admissible, and hence, by Lemma 9.2, this sequence satisfies RIP, as needed.

It remains to prove the claim. This is done by induction on k. The case $k = 1$ follows from Lemma 9.2. Assuming the assertion of the claim for $k-1$, we prove it for k, $k \geq 2$.

By the induction hypothesis there is a tree T on the set of vertices $\mathcal{F}$ so that $F_1, F_2, \ldots, F_{k-1}$ is an initial segment in an admissible ordering of the vertices of the tree. Therefore, each F_j for $j \geq k$ is a descendent in T of at least one of the vertices in the set $\{F_1, F_2, \ldots, F_{k-1}\}$. In particular, this holds for F_k, let F_i be the first vertex in the path from F_k to the root F_1 in T so that $i \leq k-1$. If F_i is the parent of F_k in T, then the tree T satisfies the assertion of the claim for k, establishing the required induction step. We thus assume that this is not the case and the path in T from F_i to F_k is the following: $F_i, G_1, G_2, \ldots, G_s, F_k$, where $G_j \in \{F_{k+1}, F_{k+2}, \ldots, F_m\}$ for all $j, 1 \leq j \leq s$.

Our objective is to transform T into another tree T' that satisfies the assertion of the claim for k. To this end we define several pieces of the tree T, as follows. Let T_0 be the subtree of T rooted at F_1 and consisting of all vertices of T besides G_1 and its descendents. Let T_1 denote the subtree of T rooted at G_1, besides G_2 and its descendents. Similarly, for each $q < s$, let T_q denote the subtree of T rooted at G_q besides G_{q+1} and its descendents. Let T_s be the subtree of T rooted at G_s besides F_k and its descendents. Finally, let T_∞ denote the subtree of T rooted at F_k.

Note that the tree T_0 contains all the vertices $F_1, F_2, \ldots, F_{k-1}$. This is because these vertices form an initial segment in an admissible ordering of the vertices of T, hence none of them can be a descendent of G_1, which is not in this initial segment.

Recall that $F_1, F_2, \ldots, F_m$ is an expansive ordering of the members of $\mathcal{F}$. Therefore,

$$|F_k \cap (\cup_{j<k} F_j)| \geq |G_q \cap (\cup_{j<k} F_j)| \tag{10}$$

for all $1 \leq q \leq s$. On the other hand, T is a realization for $\mathcal{F}$ which satisfies TDP, and as all vertices F_j for $j < k$ lie in T_0, it follows that for each such F_j, $F_k \cap F_j \subset F_i$ and also $F_k \cap F_i \subset G_q$ for all $1 \leq q \leq s$. We conclude that $F_k \cap (\cup_{j<k} F_j) = F_k \cap F_i$ is contained in $F_i \cap G_q$ for all $1 \leq q \leq s$, and by (10) we have

$$F_k \cap (\cup_{j<k} F_j) = F_k \cap F_i = F_i \cap G_1 = F_i \cap G_2 = \ldots = F_i \cap G_s.$$

We can now construct the tree T'. It is obtained from T by reversing the path between G_1 and F_k as follows: starting with the subtree T_0, connect to it the subtree T_∞ by making F_i the parent of F_k. Next, connect the subtree T_s by making F_k the parent of G_s, the subtree T_{s-1} by making G_s the parent of G_{s-1} and so on until the tree T_1 which is connected by letting G_2 be the parent of G_1. Clearly $F_1, F_2, \ldots, F_k$ is an initial segment in an admissible ordering of T', and hence it only remains to check that the tree T' is indeed a realization of $\mathcal{F}$, namely, that for every three vertices along a path in the tree, the subset corresponding to the middle vertex is contained in the intersection of those corresponding to the other two subsets. This is obvious if all three vertices belong to T_0 or if none of them belongs to T_0. The only remaining cases are when the path is between a vertex in T_0 and a vertex not in T_0. In this case, the intersection of the corresponding sets is contained in the common value of $F_k \cap F_i = G_1 \cap F_i = G_2 \cap F_i = \ldots = G_s \cap F_i$ and the desired inclusion in T' follows from the corresponding one in T. This completes the proof of the claim, establishing the assertion of the theorem. $\square$

Acknowledgment: I would like to thank Michael Krivelevich, Troy Lee, Jirka Matoušek, Adi Shraibman and Ran Spiegler for helpful discussions and useful comments on various parts of this article.

References

[1] I. Adler, N. Alon and S. M. Ross, On the Maximum Number of Hamiltonian Paths in Tournaments, Random Structures and Algorithms 18 (2001), 291-296.

[2] T. A. Althuis and F. Göbel, Graph theoretic aspects of music theory, University of Twente, Dept. of Applied Math, Memoranda 1573, 2001.

[3] N. Alon, J. H. Kim and J. Spencer, Nearly perfect matchings in regular simple hypergraphs, Israel J. Math. 100 (1997), 171-187.

[4] N. Alon, T. Lee and A. Shraibman, The cover number of a matrix and its algorithmic applications, Proc. APPROX 2014, 34-47.

[5] N. Alon, Monochromatic directed walks in arc colored directed graphs, Acta Math. Acad. Sci. Hungar. 49 (1987), 163-167.

[6] N. Alon, Restricted colorings of graphs, in "Surveys in Combinatorics", Proc. 14^{th} British Combinatorial Conference, London Mathematical Society Lecture Notes Series 187, edited by K. Walker, Cambridge University Press, 1993, 1-33.

[7] N. Alon, Degrees and choice numbers, Random Structures and Algorithms 16 (2000), 364-368.

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

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

[10] M. O. Albertson, J. Pach and M. E. Young, Disjoint homometric sets in graphs, Ars Mathematica Contemporanea, 4(1):1–4, 2011.

[11] N. Alon and J. H. Spencer, *The Probabilistic Method*, Third Edition, Wiley, New York, 2008.

[12] M. Axenovich and L. Özkahya, On homometric sets in graphs, Australasian Journal of Combinatorics, 55:175–187, 2013.

[13] T. C. Brown and J. P. Buhler, A density version of a geometric Ramsey theorem, J. Combin. Theory Ser. A 32 (1982), no. 1, 20–34.

[14] B. Bollobás, *Extremal Graph Theory*, Academic Press, London, 1978.

[15] J. A. Bondy, Basic Graph Theory: paths and circuits, in : "Handbook of Combinatorics", R.L. Graham, M. Grötschel and L. Lovász, eds, North Holland (1995), Chapter 1, pp. 3-110.

[16] M. Bateman and N. H. Katz, New bounds on cap sets, J. Amer. Math. Soc. 25 (2012), no. 2, 585–613.

[17] R. G. Cowell, A. P. Dawid, S. L. Lauritzen and D. J. Spiegelhalter, Probabilistic networks and expert systems, Springer-Verlag, New York, 2007, xii+321 pp.

[18] P. Erdős and L. Lovász, *Problems and results on* 3*-chromatic hypergraphs and some related questions*, in:"Infinite and Finite Sets" (A. Hajnal et. al. eds), Colloq. Math. Soc. J. Bolyai 11, North Holland, Amsterdam, 1975, pp. 609-627.

[19] P. Frankl, R. Graham and V. Rödl, On subsets of abelian groups with no 3-term arithmetic progression, J. Combin. Theory Ser. A, 45 (1987), pp. 157–161.

[20] R. Fulek and S. Mitroviić, Homometric sets in trees, European J. Combin. 35 (2014), 256–263.

[21] T. Gallai, On directed graphs and circuits. In: Theory of Graphs, Proc. Colloq. Tihany 1966, New York Academic Press, pp. 115–118 (1968).

[22] F. Goldberg, D. Rajendraprasad and R. Mathew, Domination in designs, arXiv:1405.3436v1

[23] M. Hasse, Zur algebraischen Begründung der Graphentheorie. I (in German). Mathematische Nachrichten 28(5.6):275–290 (1965).

[24] R. Hod and A. Naor, Component Games on Regular Graphs, Combinatorics, Probability and Computing, 23 (2014), 75–89.

[25] R. Jamison, Private communication, 2014.

[26] S. Jukna, *Extremal Combinatorics*, Springer-Verlag, Berlin, 2001.

[27] N. Linial, Y. Mansour and R. Rivest, Results on learnability and the Vapnik-Chervonenkis dimension, Information and Computation 90 (1991), 33-49.

[28] S. L. Lauritzen, Graphical models, Oxford Statistical Science Series, 17. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1996. x+298 pp.

[29] P. Lemke, S. S. Skiena and W. D. Smith, Reconstructing sets from interpoint distances, Discrete and Computational Geometry: The Goodman-Pollack Festschrift, vol. 25 of Algorithms Combin., Springer, Berlin, 2003, 597–631.

[30] R. Meshulam, On subsets of finite abelian groups containing no 3-term arithmetic progressions, Journal of Combinatorial theory Ser. A, 71(1995), 168–172.

[31] V. Rödl, On a packing and covering problem, Europ. J. Combinatorics 6 (1985), 69-78.

[32] B. Roy, Nombre chromatique et plus longs chemins d'un graphe (in French). Rev. Francaise Informat. Recherche Opérationnelle 1(5):129–132 (1967).

[33] J. Singer, A theorem in finite projective geometry and some applications to number theory, *Trans. Amer. Math. Soc.* 43 (1938), 377–385.

[34] R. Spiegler, Bayesian networks and missing data imputations, to appear.

[35] D. Saxton and A. Thomason, Hypergraph containers, Inventiones Mathematicae, to appear.

[36] L. M. Vitaver, Determination of minimal coloring of vertices of a graph by means of Boolean powers of the incidence matrix (in Russian). Dokl. Akad. Nauk SSSR 147: 758–759 (1962).

[37] R. M. Wilson, An existence theory for pairwise balanced designs. III. Proof of the existence conjectures, J. Combinatorial Theory Ser. A 18 (1975), 71–79.