Graph-Codes: Questions, Results and Methods

Noga Alon [*]

Abstract

This paper is a presentation of the content of the Takagi Lectures delivered by the author in Tokyo in 2025. The topic discussed is Graph-Codes, a subject motivated by questions in Extremal Combinatorics, Additive Number Theory and Coding Theory. The initial guiding fact is that viewing binary vectors as characteristic vectors of edge-sets of graphs transforms the basic combinatorial questions of Coding Theory into intriguing extremal problems about families of graphs. We discuss some of these questions and describe several results and open problems. The relevant methods combine Combinatorial and Probabilistic tools with techniques from Information Theory, Number Theory and the theory of Combinatorial Designs.

1 Introduction

The study of Graph-Codes starts with the observation that viewing graphs as characteristic vectors of their edge-sets transforms the basic problems of Coding Theory into intriguing questions in Extremal Graph Theory. Special cases of problems of this type have been considered already in 1976 by Simonovits and Sós, and in 2009 by Gowers, but the systematic study of the subject was initiated only a few years ago in [4]. Since then, the subject received a considerable amount of attention by many researchers, but despite this recent intensive study some of the basic problems in the area are still wide open. In this survey we start with a formal description of the subject, present its background, and proceed with the challenging problems in the area, some of the known results and the relevant techniques.

The *symmetric difference* $G_1 \oplus G_2$ of two graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ on the same set of vertices V is the graph $(V, E_1 \oplus E_2)$ where $E_1 \oplus E_2$ is the symmetric difference of E_1 and E_2, that is, the set of all edges that belong to exactly one of the

---

[*] Princeton University, Princeton, NJ, USA. Email: `nalon@math.princeton.edu`. Research supported in part by NSF grant DMS-2154082

two graphs. Put $V = [n] = \{1, 2, \ldots, n\}$ and let $\mathcal{H}$ be a family of graphs on the set of vertices $[n]$ which is closed under isomorphism. A collection of graphs $\mathcal{F}$ on $[n]$ is called an *$\mathcal{H}$-(graph)-code* if it contains no two members whose symmetric difference is a graph in $\mathcal{H}$. For the special case that $\mathcal{H}$ contains all copies of a single graph H on $[n]$ this is called an H-code. The basic extremal question in the subject is that of determining or estimating the maximum possible cardinality of such codes for various families $\mathcal{H}$. Let $D_{\mathcal{H}}(n)$ denote this maximum, and let

$$d_{\mathcal{H}}(n) = \frac{D_{\mathcal{H}}(n)}{2^{\binom{n}{2}}}$$

denote the maximum possible fraction of the total number of graphs on $[n]$ in an $\mathcal{H}$-code. If $\mathcal{H}$ consists of all graphs isomorphic to a single graph H with no isolated vertices (together with $n - |V(H)|$ additional isolated vertices to ensure that this is a family of n-vertex graphs) we denote $d_{\mathcal{H}}(n)$ by $d_H(n)$. Note that if $\mathcal{H}$ consists of all graphs with less than d edges, then $D_{\mathcal{H}}(n)$ is simply the maximum possible cardinality of a binary error-correcting code of length $\binom{n}{2}$ and minimum distance at least d. This motivates the terminology "graph-codes".

The case $\mathcal{H} = \mathcal{K}$ where $\mathcal{K}$ is the family of all cliques is of particular interest. This case is motivated by a conjecture of Gowers raised in his blog post [22] in 2009 and is discussed briefly in the comments of that blog. As explained in [22], the version in which the family is not allowed to contain two distinct graphs whose symmetric difference is a clique with the extra requirement that one of the graphs is a subgraph of the other can be viewed as the first open case of a polynomial density Hales-Jewett Conjecture, a basic problem in Additive Ramsey Theory. This version, as well as the one without the additional requirement above, are both wide open, namely, it is not known whether or not the maximum possible density in any of them is bounded away from zero.

If $\mathcal{H}$ consists of all graphs with independence number at most 2, then $d_{\mathcal{H}}(n) \geq 1/8$ for all $n \geq 3$, as shown by the family of all graphs on $[n]$ containing a triangle on the set of vertices $\{1, 2, 3\}$. An interesting result of Ellis, Filmus and Friedgut [15], settling a conjecture of Simonovits and Sós raised in 1976, asserts that this is tight for all $n \geq 3$. The corresponding result, that $d_{\mathcal{H}'}(n) = 1/2^6$ for all $n \geq 4$, where $\mathcal{H}'$ is the family of all graphs with independence number at most 3, is proved in [9]. The natural extension for independence number at most 4 is established in a very recent work [31].

A more systematic study of the parameters $D_{\mathcal{H}}(n)$ and $d_{\mathcal{H}}(n)$ for various families of graphs $\mathcal{H}$ appears in [4]. The families $\mathcal{H}$ considered in this work include the family of all disconnected graphs, the family of all graphs that are not 2-connected, the family of all non-Hamiltonian graphs and the family of all graphs that contain or do not contain a spanning star. Additional families studied are all graphs that contain an induced or non-induced copy of a fixed graph T, or all graphs that do not contain such a subgraph. See also [6] for an extension.

2 Single graphs and complete graphs

The case that $\mathcal{H}$ consists of a single graph H and the case that $\mathcal{H}$ is the family of all cliques are of particular interest. Note that trivially, if every member of $\mathcal{H}$ has an odd number of edges then $d_{\mathcal{H}}(n) \geq \frac{1}{2}$ as the family of all graphs on $[n]$ with an even number of edges forms an $\mathcal{H}$-code. (In fact it is easy to see that $d_{\mathcal{H}}(n) = \frac{1}{2}$ in this case).

This suggests the following intriguing question.

Question 2.1. *Let $\mathcal{H}$ be a finite family of graphs closed under isomorphism. Is it true that $d_{\mathcal{H}}(n)$ tends to 0 as n tends to infinity if and only if $\mathcal{H}$ contains a graph with an even number of edges ? Equivalently: is it true that for any fixed graph H with an even number of edges, $d_H(n)$ tends to 0 as n tends to infinity ?*

In [2] it is shown that $d_H(n)$ does indeed tend to 0 if H is any graph obtained by gluing two copies of a subgraph H' along an independent set. Such a graph is called a *square*. These are the only graphs H (with an even number of edges) for which the result is known. In [14] it is proved that this is also the case for any *positive* H (a graph is positive iff the homomorphism density of it in any weighted graph is non-negative). The (wide open) positive graph conjecture (see [10]) asserts that a graph H is positive if and only if it is a square. Therefore, if the positive graph conjecture is true, then the result of [14] does not add any additional examples of graphs H for which $d_H(n)$ is known to tend to 0 as n grows to infinity. It is thus possible that the only graphs H for which $d_H(n)$ tends to 0 are the squares. The other extreme possibility is that this is the case for every graph H with an even number of edges, and the exact characterization can also be somewhere between these two extremes.

As is the case with standard Error-Correcting Codes, the linear variant of the problems, where the $\mathcal{H}$-codes considered are restricted to linear subspaces, that is, to families of graphs on $[n]$ closed under symmetric difference, is also interesting. Let $L_{\mathcal{H}}(n)$ denote the maximum possible cardinality of a linear $\mathcal{H}$-code of graphs on $[n]$. It is clear that $L_{\mathcal{H}}(n) \leq D_{\mathcal{H}}(n)$ for every family $\mathcal{H}$. The exact value of $L_{\mathcal{K}}(n)$ for the family of all cliques $\mathcal{K}$ is determined in [2]:

Theorem 2.2.

$$L_{\mathcal{K}}(n) = 2^{\binom{n}{2} - \lfloor n/2 \rfloor}.$$

For the non-linear case, in contrast, it is not even known whether or not $d_{\mathcal{K}}(n) = o(1)$.

Question 2.3. *Does $d_{\mathcal{K}}(n)$ converge to 0 as n tends to infinity ?*

The assertion of Theorem 2.2 is equivalent to the fact that the minimum possible co-dimension of a linear space of graphs on $[n]$ containing no nontrivial clique is $[n/2]$. Let $r = r(n)$ denote this minimum possible co-dimension, and let $G_1, \ldots, G_r$ be graphs on $[n]$ that form a basis of the dual space of such an extremal linear code. The proof that $r(n) \leq [n/2]$ is by an explicit construction described in [2]. The lower bound is proved using the Chevalley-Warning theorem-a classical result in Number Theory (see, e.g., [8] or [35]). It suffices to prove the bound for even n (as it implies the bound for $n+1$). Assume n is even and let $G_1, \ldots G_{n/2-1}$ be a family of $n/2 - 1$ graphs on $[n]$. We have to show that there is a clique on at least 2 vertices containing an even number of edges of each G_i. We show that in fact there is such a clique on an even (positive) number of vertices. Associate each vertex i with a variable x_i over Z_2 and consider the following homogeneous system of polynomial equations over Z_2. For each graph G_s in our family,

$$\sum_{ij \in E(G_s)} x_i x_j = 0.$$

In addition, add the linear equation $\sum_{i=1}^{n} x_i = 0$.

The sum of the degrees of the polynomials here is $2(n/2 - 1) + 1 = n - 1$, which is smaller than the number of variables. Since the system is homogeneous it admits the trivial solution $x_i = 0$ for all i. By the Chevalley-Warning Theorem there is an additional solution. Any other solution gives a clique on the set of vertices $\{i : x_i = 1\}$ which is nonempty, of even cardinality, and contains an even number of edges (possibly zero) of each G_i. This establishes the lower bound.

3 Graph-codes as independent sets in Cayley graphs

Any graph-code is an independent set in a Cayley graph of an elementary abelian 2-group. Indeed, for a given family of graphs $\mathcal{H}$ on the set of vertices $[n]$, we can associate a Cayley graph of the group $Z_2^{\binom{n}{2}}$. The vertices of this graph represent all possible labelled graphs on $[n]$. The generating set of the graph is $\mathcal{H}$, that is, two vertices G_1, G_2 are connected iff $G_1 \oplus G_2 \in \mathcal{H}$. A collection of graphs $\mathcal{F}$ on $[n]$ is an $\mathcal{H}$ graph-code if and only if it forms an independent set in this Cayley graph. Therefore, the maximum possible size of an $\mathcal{H}$-code is the independence number of this Cayley graph. This independence number can be bounded by spectral techniques using the known fact that the eigenvalues of Cayley graphs of abelian groups have a simple expression as character sums. In addition, since any Cayley graph is vertex transitive, its independence number is equal to the ratio between its number of vertices and its fractional chromatic number. This implies that one can upper bound this independence number by exhibiting any set (or even weighted set) of vertices with a small independence ratio.

The relevance of graph eigenvalues to independent sets in graphs is well known and can be traced back to the old result that the independence number of any regular graph on N vertices in which the eigenvalues of the adjacency matrix are $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ is at most $\frac{-N\lambda_N}{\lambda_1-\lambda_N}$. This can be deduced from old results of Hoffman, a proof can be found in [27], where it is also shown that the above quantity bounds even the Shannon capacity of the graph (which is at least its independence number). It is simple and well known (see, e.g., [28]), that the eigenvalues of a Cayley graph of an abelian group with respect to a generating set S are all the quantities

$$\sum_{s \in S} \chi(s)$$

where χ ranges over all characters of the group.

For an $\mathcal{H}$ graph-code on $[n]$ this implies that the eigenvalues of the relevant Cayley graph C are the following. For any fixed labelled graph F on $[n]$ define a weight function w_F on the edges of the complete graph on $[n]$, where the weight of every edge is -1 if it belongs to F and 1 otherwise. The corresponding eigenvalue is the number of members of $\mathcal{H}$ with characteristic vector that has a positive inner product with w_F, minus the number of those with negative inner product. If $\mathcal{H}$ consists of a single graph H (and all its isomorphic copies) and this graph is positive, this implies that the eigenvalues of C satisfy $|\lambda_N|/\lambda_1 \leq c(H)/n$, where $N = 2^{\binom{n}{2}}$ is the number of its vertices and $c(H)$ is a constant (depending on H). Indeed, for each fixed graph F on $[n]$ the total weight (with respect to w_F) of all homomorphic images of H is non-negative, and only a fraction of $c(H)/n$ of them are not bijective. Thus the relevant eigenvalue cannot be too negative. This gives the result of [14] that for positive graphs H, $d_H(n)$ tends to 0 as n tends to infinity, which is proved in [14] in a related though somewhat different way, described without the spectral interpretation.

4 Global Properties

Several results providing exact values for the maximum possible cardinality of $\mathcal{H}$ graph-codes for families $\mathcal{H}$ defined by a global property of graphs appear in [4]. In this section we briefly describe some of them. To avoid confusion note that for any family $\mathcal{H}$, $D_{\mathcal{H}}(n)$ is the maximum possible cardinality of a collection of graphs on $[n]$ so that no symmetric difference of two distinct members belongs to $\mathcal{H}$. If $\overline{\mathcal{H}}$ denotes the family of all graphs not in $\mathcal{H}$, then $D_{\overline{\mathcal{H}}}(n)$ is the maximum possible cardinality of a collection of graphs on $[n]$ in which the symmetric difference of every two distinct members belongs to $\mathcal{H}$.

Proposition 4.1. *Let $\mathcal{H}$ be the family of all disconnected graphs. Then $D_{\mathcal{H}}(n) = 2^{n-1}$.*

The upper bound here is simple: if $\mathcal{F}$ is an $\mathcal{H}$-code on $[n]$ then no two distinct members G_1, G_2 of $\mathcal{F}$ can have exactly the same set of the neighbors of the vertex 1 (say). Indeed, otherwise 1 would be isolated in $G_1 \oplus G_2$, contradicting the requirement that it cannot be disconnected. The lower bound is given, for example, by the collection of all complete bipartite graphs (including the empty one).

A similar reasoning and construction works for the 2-connected case (or its complement), at least when n is even.

Proposition 4.2. *Let $\mathcal{H}$ be the family of all graphs that are not 2-connected. Then for even n, $D_{\mathcal{H}}(n) = 2^{n-2}$.*

For the 3-connected case an exact result is proved in [4] for linear graph-codes and for infinitely many value of n, using Hamming codes. Here is a construction showing that for any $n = 2^k - 1$, where $k \geq 2$ is an integer, there is a linear space of 2^{n-k-1} graphs on $[n]$ so that the symmetric difference of every two distinct members of the collection is 3-connected. Let C be the Hamming code of length $n = 2^k - 1$. This is a linear code of binary vectors of length n, its minimum distance is 3, and its dimension is $n-k$, that is, it contains 2^{n-k} vectors. Since the all 1 vector is in the code, for any vector $v = (v_1, \ldots, v_n)$ it contains, it also contains its complement $\overline{v} = (1-v_1, 1-v_2, \ldots, 1-v_n)$. For each pair of vectors $v, \overline{v} \in C$, let $H(v)$ denote the complete bipartite graph on $[n]$ with vertex classes $\{i : v_i = 0\}$ and $\{j : v_j = 1\}$. Let the graph-code $\mathcal{F}$ be the set of all graphs $H(v)$ as above. It is not difficult to check that $\mathcal{F}$ is a linear space, that is, it is closed under symmetric difference. In particular it contains the edgeless graph, which is $H(v)$ for the pair of vectors consisting of the all 0 and the all 1 vectors. The cardinality of $\mathcal{F}$ is clearly half the number of vectors in C, that is $|\mathcal{F}| = 2^{n-k-1}$. In addition, every member of $\mathcal{F}$ besides the edgeless graph is 3-connected, since the minimum distance of

C is 3, implying that each nonzero vector $v \in C$ has at least 3 coordinates which are 1 (and at least 3 which are zero, as the Hamming weight of $\overline{v}$ is also at least 3).

It is not difficult to see that for every n, any collection of more than $\frac{2^{n-1}}{n}$ graphs on $[n]$ must contain two distinct members so that the degree of vertex number 1 (say) in their symmetric difference is at most 2. Therefore any graph-code on $[n]$ in which the symmetric difference of any pair of graphs is 3-connected cannot be of size larger than $\frac{2^{n-1}}{n}$. For $n = 2^k - 1$ this, together with the fact that the cardinality of any linear graph-code has to be a power of 2, show that the construction above is tight for the linear case.

Similar ideas can be used to prove the following.

Proposition 4.3. *Let $\mathcal{H}$ be the family of all graphs with minimum degree strictly smaller than d. Then for every even n, $D_{\mathcal{H}}(n)$ is exactly the maximum cardinality of a binary error-correcting code of length $n-1$ and minimum distance at least d.*

The proof of the upper bound here proceeds by observing that the projections of the graphs in an optimal $\mathcal{H}$-code on the set of all edges incident with vertex number 1 are distinct and must form an error-correcting code of length $n-1$ and minimum distance at least d. The lower bound is obtained from a code C of length $n-1$ and minimum distance at least d as follows. Fix a proper edge-coloring of K_n by $n-1$ colors, and for each codeword $c = (c_1, c_2, \ldots, c_{n-1}) \in C$ take the graph consisting of all edges colored by all the colors j so that $c_j = 1$.

Another global property for which a precise result is known only for infinitely many values of n is that of not containing a Hamilton cycle. The proof here applies the known results about the so-called Perfect 1-factorization conjecture (P1FC) of Kotzig [26]. This conjecture asserts that for every even $n > 2$ the edges of the complete graph K_n can be partitioned into perfect matchings so that the union of any two of them forms a Hamilton cycle. This is still open in general, but is known in several special cases including the cases $n = p+1$ or $n = 2p$ for any odd prime p, see [34] and the references therein.

Proposition 4.4. *Let $\mathcal{H}$ be the family of all graphs not containing a Hamilton cycle. For all even values of n for which the P1FC holds, $D_{\mathcal{H}}(n) = 2^{n-2}$. In particular, this holds if $n = p+1$ or $n = 2p$ for some odd prime p.*

The results described in Propositions 4.1, 4.2, 4.4 admit linear extremal families. We proceed with a natural example in which the extremal families are not linear.

Let $\mathcal{H}$ denote the family of all graphs not containing a spanning star, that is, a star spanning all vertices. The following result deals with the maximum possible cardinality of $\mathcal{H}$ and of $\overline{\mathcal{H}}$ codes.

Proposition 4.5. *Let $\mathcal{H}$ denote the family of all graphs not containing a spanning star, that is, a star spanning all vertices. Then:*

- *$D_{\mathcal{H}}(n) = n + 1$ for all odd n and $D_{\mathcal{H}}(n) = n$ for all even n.*
- *For every even n, $D_{\overline{\mathcal{H}}}(n) = 2^{\binom{n}{2} - \frac{n}{2}}$.*

Note that the extremal families for the result in the first part of the proposition cannot be linear in general, as the extremal cardinality is not a power of 2 for most values of n.

The proof of the first part of the last proposition is combinatorial, the details can be found in [4]. The lower bound in the second part is simple, an extremal code (which is linear) is the collection of all graphs that do not contain any edge of some fixed perfect matching of K_n. The upper bound is established using Shearer's Lemma, which is the following basic result proved using the properties of the entropy function ([11], see also [5], Corollary 15.7.7).

Lemma 4.6 ([11])**.** *Let S be a finite set and let $A_1, \ldots, A_m$ be subsets of S such that every element of S is contained in at least k of the sets $A_1, \ldots, A_m$. Let $\mathcal{M}$ be a collection of subsets of S and let $\mathcal{M}_i = \{T \cap A_i : T \in \mathcal{M}\}$ for $1 \le i \le m$. Then*

$$|\mathcal{M}|^k \le \prod_{i=1}^{m} |\mathcal{M}_i|.$$

Let $\mathcal{M}$ be a collection of graphs on $[n]$ so that no symmetric difference of two members of $\mathcal{M}$ contains a spanning star (that is, a vertex of degree $n-1$). For each $i = 1, \ldots, n$ let S_i be the set of the $n-1$ edges of K_n incident with vertex i. Then for any $T, T' \in \mathcal{M}$ we cannot have $E(T') \cap S_i = S_i \setminus (E(T) \cap S_i)$, that is, $E(T)$ and $E(T')$ cannot be complementary on any S_i. Therefore, if $\mathcal{M}_i$ denotes the family of graphs obtained by taking the projection of all graphs from $\mathcal{M}$ to the edge set S_i, then $|\mathcal{M}_i| \le 2^{n-2}$. Since each edge of K_n appears in exactly two of the sets S_i, one can apply Shearer's Lemma (Lemma 4.6) to these sets with $k = 2$. This implies

$$|\mathcal{M}|^2 \le \prod_{i=1}^{n} |\mathcal{M}_i| \le 2^{n(n-2)}.$$

The desired upper bound follows by taking square roots.

In most of the results discussed in this section, there are linear extremal examples of graph-codes. One may wonder whether or not the gap between the maximum possible cardinality of an $\mathcal{H}$-graph-code and that of a linear $\mathcal{H}$ code can be bounded by a moderately growing function of n. This is of particular interest in view of the assertion of Theorem 2.2 and Question 2.3 that address these extremal questions for the family $\mathcal{H} = \mathcal{K}$ of all cliques.

It turns out that there are families $\mathcal{H}$ for which the ratio between $D_{\mathcal{H}}(n)$ and $L_{\mathcal{H}}(n)$ can be very large, indeed it can be exponential in n^2. A simple example with such a gap is the family $\mathcal{H}$ of all graphs on $[n]$ with more than $2g$ edges, where $g = \Theta(n^2)$. For this family, an $\mathcal{H}$-graph-code on $[n]$ is a collection of labelled graphs on $[n]$ in which the Hamming distance between (the characteristic vectors of the edge sets of) any two members of the collection is at most $2g$. The diameter of this set of vectors is at most $2g$, and a well known result of Kleitman [25] implies that the maximum possible cardinality of such a collection is

$$D_{\mathcal{H}}(n) = \sum_{i=0}^{g} \binom{\binom{n}{2}}{i}.$$

An extremal example is the collection of all graphs with at most g edges.

For $g = \lfloor c\binom{n}{2} \rfloor$ where $0 < c < 1/2$ is fixed, the above quantity is $2^{(H(c)+o(1))n^2/2}$, where

$$H(c) = -c\log_2 c - (1-c)\log_2(1-c)$$

is the binary entropy function and the $o(1)$-term tends to 0 as n tends to infinity.

On the other hand, any linear $\mathcal{H}$-graph-code cannot have dimension exceeding $2g$, since any vector space of dimension d contains a vector with Hamming weight at least d. Therefore $L_{\mathcal{H}}(n) \leq 2^{2g}$ (in fact equality holds, as shown by the family of all subgraphs of any fixed graph with $2g$ edges). It is easy to check that the ratio between $2^{(H(c)+o(1))n^2/2}$ and 2^{cn^2} is $2^{\Omega(n^2)}$ for any fixed $c \in (0, 1/2)$, providing an example with a large gap between $D_{\mathcal{H}}(n)$ and $L_{\mathcal{H}}(n)$.

Another example exhibiting a large gap between the linear and the non-linear cases deals with cliques. For $s < n$, let $\mathcal{K}(s)$ denote the family of all complete subgraphs of K_n with at least s vertices. The argument sketched in the proof of Theorem 2.2 can be easily modified to show that if $n - s(n)$ tends to infinity as n tends to infinity, then the maximum possible density of a linear $\mathcal{K}(s)$ graph-code tends to 0. Indeed, the minimum possible co-dimension of such a linear code is larger than $(n-s)/2$. To see that this is the case assume this is false, suppose $t \leq (n-s)/2$ and let G_i, $1 \leq i \leq t$ be a family of graphs on $[n]$ which forms a basis for the dual code. As in the proof of Theorem 2.2, associate each vertex i with a variable x_i over Z_2 and consider the homogeneous system of t polynomial equations over Z_2 consisting of the t quadratic equations

$$\sum_{ij \in E(G_s)} x_i x_j = 0 \text{ for all } 1 \leq s \leq t.$$

This has the trivial solution $x_i = 0$ for all i. Fix a solution x'_i with the maximum possible Hamming weight. Put $I = \{i : x'_i = 0\}$. If $|I| > 2t$, then fixing all the nonzero $x_i = x'_i = 1$ for $i \notin I$ and viewing the same system of equations above as equations in the variables $x_i, i \in I$, the number of variables exceeds the sum of the degrees. As $x_i = 0$ for $i \in I$ is a solution (with the fixed values of $x_i = x'_i$ for $i \notin I$), there is another solution, contradicting the maximality. This shows that there is a solution of Hamming weight at least $n - 2t \geq s$, showing that any linear graph-code of codimension at most $(n-s)/2$ contains a clique of size at least s.

In particular, if, for example, $s = n - n^{0.49}$, this implies that the maximum possible density of a linear graph-code for the family of all cliques of size at least s is at most $2^{-n^{0.49}/2} = o(1)$. On the other hand, there is a non-linear graph-code of density $1/2 - o(1)$ in which the symmetric difference of any two distinct members is not such a large clique (and in fact is not even a graph that contains such a large clique). Here is a construction. Fix a matching M of $2\lfloor n/4 \rfloor$ edges in K_n and consider all graphs that contain less than $|M|/2 - n^{0.49}/2$ edges of M. It is easy to check that this is a set of nearly half the graphs on $[n]$. On the other hand, the symmetric difference of any two graphs in this collection misses more than $n^{0.49}$ edges of M, and hence cannot contain a clique of size at least $n - n^{0.49}$.

5 Local Properties

The study of $\mathcal{H}$ graph-codes for families $\mathcal{H}$ defined by a local property of graphs also leads to several intriguing problems. An example mentioned already in the introduction is that of determining the maximum possible cardinality of a collection of graphs on $[n]$ in which the symmetric difference of every two members of the collection has independence number at least k. This is a local property, as it can be demonstrated by the induced subgraph on only k vertices. Let $\mathcal{H}(k)$ denote the family of all graphs with independence number smaller than k. The collection of all graphs in which some fixed set of k vertices is an independent set shows that

$$d_{\mathcal{H}(k)}(n) \geq 2^{-\binom{k}{2}}.$$

As mentioned in the introduction, this is tight for $k = 3$ as proved by Ellis, Filmus and Friedgut [15], settling a problem of Simonovits and Sós. It is also tight for $k = 4$, as proved by Berger and Zhao [9] and for $k = 5$ as shown by Mani and Zhang [31]. The following conjecture is natural.

Conjecture 5.1 (Simonovits and Sós, see also [15], [9], [31])**.** *For every integer $k \geq 3$ and $n \geq k$,*

$$d_{\mathcal{H}(k)}(n) = 2^{-\binom{k}{2}}.$$

A natural local graph property is that of containing a fixed graph L, either as a subgraph or as an induced subgraph. The relevant families of graphs here are $\mathcal{H}(L)$ and $\mathcal{H}^*(L)$. $\mathcal{H}(L)$ is the family of all graphs containing no copy of L as a subgraph, and $\mathcal{H}^*(L)$ is the family of all graphs containing no copy of L as an induced subgraph. Thus $D_{\mathcal{H}(L)}(n)$ is the maximum possible cardinality of a collection of graphs on $[n]$ in which the symmetric difference of any two members contains a copy of L. The quantity $D_{\mathcal{H}^*(L)}(n)$ is defined similarly, with the copy of L replaced by an induced copy. Note that $D_{\mathcal{H}^*(L)}(n) \leq D_{\mathcal{H}(L)}(n)$ for every L and n.

The problem of determining these functions precisely is open, even for relatively simple graphs L like a triangle. But the asymptotic behavior of them is well understood, as stated in the following result.

Theorem 5.2. *For every fixed graph L of chromatic number ℓ*

$$2^{(\frac{1}{\ell-1}-o(1))\frac{n^2}{2}} \leq D_{\mathcal{H}^*(L)}(n) \leq D_{\mathcal{H}(L)}(n) \leq 2^{\frac{1}{\ell-1}\frac{n^2}{2}},$$

where the $o(1)$-term tends to 0 as n tends to infinity.

The upper bound is easy. Fix a partition of $[n]$ into $\ell - 1$ nearly equal classes $V_1, \ldots, V_{\ell-1}$. Let $\mathcal{G}$ be a collection of graphs on $[n]$. If the cardinality of $\mathcal{G}$ exceeds $2^{\frac{1}{\ell-1}n^2/2}$ then by the pigeonhole principle it contains two distinct members G_1, G_2 that have exactly the same induced subgraphs on each V_i. Thus $G_1 \oplus G_2$ is $(\ell-1)$-colorable and hence cannot contain a copy of L. Establishing the lower bound for $D_{\mathcal{H}(L)}(n)$ (but not for $D_{\mathcal{H}^*(L)}(n)$) is also not very difficult. By a result of Erdős, Frankl and Rödl [17] proved using the regularity lemma of Szemerédi [36], the number of L-free subgraphs on n labelled vertices is

$$2^{(1-\frac{1}{\ell-1}+o(1))\frac{n^2}{2}}.$$

Consider the Cayley graph of $Z_2^{\binom{n}{2}}$ with generating set consisting of all L-free graphs on $[n]$. Recall that $D_{\mathcal{H}(L)}(n)$ is exactly the independence number of this graph. It has $N = 2^{\binom{n}{2}}$ vertices and maximum degree $\Delta = 2^{(1-\frac{1}{\ell-1}+o(1))n^2/2}$. Therefore its independence number is at least $N/(\Delta+1)$ providing the required lower bound.

The proof of the same asymptotic lower bound for the quantity $D_{\mathcal{H}^*(L)}(n)$ is more complicated. The basic approach is to choose a large integer k, and fix a partition of $[n]$ into k pairwise disjoint sets U_i of nearly equal size. Consider the family $\mathcal{F}$ of all graphs on $[n]$ in which each U_i is an independent set. The main technical part of the proof is showing that the number of these graphs that do not contain an induced copy of L is at most $\Delta = 2^{(1-\frac{1}{\ell-1})n^2/2+o(n^2)}$. This can be established using the regularity lemma and Turán's Theorem. It follows that the set $\mathcal{F}$ contains a subset of at least $|\mathcal{F}|/(\Delta+1)$ members so that the symmetric difference of every pair contains an induced copy of L. This provides a lower bound of

$$2^{(\frac{1}{\ell-1}-\frac{1}{k})n^2/2-o(n^2)}$$

for $D_{\mathcal{H}^*(L)}(n)$, and as k can be chosen to be arbitrarily large the desired lower bound follows. The details can be found in [4]. See also [6] for a strengthening of the last theorem dealing with graph-codes in which the symmetric difference of every two members contains at least a prescribed number of pairwise vertex disjoint copies of L.

As described in Section 2, the only graphs H for which it is known that $d_H(n)$ tends to 0 as n tends to infinity are the ones obtained by gluing two copies of a subgraph along an independent set. For these graphs $d_H(n) \leq c(H)/n$. There are only a few examples of such graphs H for which $d_H(n) = o(1/n)$ and the asymptotic behaviour of this quantity is known up to a constant factor. Two such examples are stars and matchings with an even number of edges.

Proposition 5.3. *For any fixed positive integer k, let $K_{1,2k}$ denote the star with $2k$ leaves and let M_{2k} denote a matching with $2k$ edges. Then*

$$d_{K_{1,2k}}(n) = \Theta_k(1/n^k) \quad \textit{and} \quad d_{M_{2k}}(n) = \Theta_k(1/n^k).$$

The proofs for stars and matchings are similar, we start with a sketch for the case of stars. The argument for the upper bound applies the following result of Frankl and Füredi.

Lemma 5.4 ([19]). *For every fixed positive integers $\ell > \ell_1 + \ell_2$ there exist $n_0 = n_0(\ell)$ and $d_\ell > 0$ so that for all $n > n_0$, if $\mathcal{F}$ is a family of ℓ-subsets of $[n]$ in which the intersection of each pair of distinct members is of cardinality either at least $\ell - \ell_1$ or strictly smaller than ℓ_2, then*

$$|\mathcal{F}| \leq d_\ell \cdot n^{\max\{\ell_1,\ell_2\}}.$$

Let $C(n, K_{1,2k})$ be the Cayley graph on the set of all graphs on $[n]$ where two are adjacent iff their symmetric difference is a copy of $K_{1,2k}$. Our objective is to upper bound the independence number of this graph. Let $\mathcal{G}$ be the family of all stars $K_{1,2k-1}$ with center 1 and $2k-1$ leaves among the vertices $\{2, 3, \ldots, n\}$. Thus $|\mathcal{G}| = \binom{n-1}{2k-1}$. If two such stars share exactly $k-1$ common leaves then their symmetric difference is a copy of $K_{1,2k}$. Therefore, by Lemma 5.4 above with $\ell = 2k-1, \ell_1 = \ell_2 = k-1$, the maximum cardinality of a subset of $\mathcal{G}$ which is independent in the Cayley graph $C(n, K_{1,2k})$ is at most some $c_k(n-1)^{k-1}$ for all sufficiently large n. As discussed in Section 3 the fact that the Cayley graph is vertex transitive supplies now the required upper bound

$$\frac{c_k(n-1)^k}{|\mathcal{G}|} \leq O_k(\frac{1}{n^k}),$$

for $d_{K_{1,2k}}(n)$. The proof for matchings is similar, starting with the family of all subsets of cardinality $2k-1$ of a fixed matching of cardinality $\lfloor n/2 \rfloor$. The symmetric difference of any two matchings that share exactly $k-1$ common edges is a copy of M_{2k}. Thus the proof can proceed as in the case of stars.

To prove the lower bound for stars, it suffices to show that the chromatic number of the Cayley graph $C = C(n, K_{1,2k})$ is at most $O(n^k)$. Let s be the smallest integer so that $2^s - 1 \geq n$. As shown by the columns of the parity check matrix of a BCH-code with designed distance $2k+1$ (see [29] or [5], Section 16.2), there is a collection S of $2^s - 1$ binary vectors of length $r = ks$ so that no sum of at most $2k$ of them (in Z_2^{ks}) is the zero vector. Fix a proper edge coloring c of K_n by n colors. For each edge e let v_e be the vector number $c(e)$ in S and color each graph F by the sum (in Z_2^{ks}) of the labels $v(e)$ for all edges e of F. This is clearly a proper coloring of $C(n, K_{1,2k})$, since no sum of labels of all edges of a copy of $K_{1,2k}$ is the zero vector. The desired lower bound for stars follows. For matchings one can use a similar construction, starting with a (non-proper) edge coloring of K_n by at most n colors in which each color class forms a star.

6 A Ramsey-type problem

It is not difficult to show, using Ramsey's Theorem, that for any fixed graph H with an even number of edges, $L_H(n) = o(2^{\binom{n}{2}})$. Indeed, showing that the co-dimension of every linear H graph-code is larger than t is equivalent to proving that for any collection of graphs G_i, $1 \leq i \leq t$ on $[n]$ there is a copy of H that intersects each G_i by an even number of edges. Assign to each edge e of K_n a binary vector of length t whose i-th coordinate is 1 iff e is an edge G_i. This defines an edge coloring of K_n by 2^t colors, so if n is sufficiently large as a function of t and H, there is a monochromatic copy of H, by Ramsey's Theorem. This copy contains an even number of edges of each G_i

As observed in [2], (see also [37]), this reasoning shows that $L_H(n)$ is closely related to a Ramsey-type problem similar to one first considered by Erdős and Gyárfás [18]. Call an edge-coloring of the complete graph on n vertices *even-H-free* if there is no copy of H in which every color appears an even number of times. Note that, in particular, if H has an even number of edges then there is no monochromatic copy of H in any such coloring. Let $r_H(n)$ denote the minimum possible number of colors in an even-H-free coloring of K_n. It can be shown (see [2], [37]) that $r_H(n) \leq n^{o(1)}$ if and only if $L_H(n) \geq 2^{\binom{n}{n}}/n^{o(1)}$. As mentioned in [2], this suggests the problem of studying $r_H(n)$. Recent papers considering this problem include [13], [7], [21], [37], [24], [38]. The constructions in all of them are based on appropriate extensions and modifications of the approach of [30], [12] for tackling the related problem of Erdős and Gyárfás about the minimum number of colors in an edge coloring of K_n in which every copy of K_r contains at least a prescribed number s of colors. The results in [13], [7], [38] suggest that for every fixed complete graph K_r, $r_{K_r}(n) \leq n^{o(1)}$, but even if this will be proved (at the moment it is known only for $r \leq 8$), there is still a very large quantitative gap between the known upper and lower bounds, even for $r = 4$. Note that the function $r_{K_r}(n)$ is interesting only when $\binom{r}{2}$ is even, hence $r = 4$ is the smallest nontrivial case. The best known lower bound for $r_{K_4}(n)$ is $\Omega(\log n)$, due to Fox and Sudakov [20]. This improves the simple $\Omega(\log n/ \log\log n)$ lower bound that follows from the known bounds for the multicolor Ramsey numbers of K_4. It is still far from the known upper bound which is $2^{O(\sqrt{\log n})}$.

An interesting problem is that of characterizing the graphs H for which $r_H(n) \geq n^{c_H}$ for some positive constant c_H. In [37] Versteegen describes a sufficient condition for the existence of such a c_H.

A graph $H = (V, E)$ has an even decomposition if there is an integer k and a sequence of sets $\emptyset = V_k \subset V_{k-1} \subset V_k \subset \ldots \subset V_0 = V$ so that for every $1 \leq i \leq k$, $V_{i-1} - V_i$ is an independent set and the number of edges between $V_{i-1} - V_i$ and V_i is even.

Proposition 6.1 ([37])**.** *If H has an even decomposition then there is a positive constant c_H so that $r_H(n) \geq n^{c_H}$.*

It is also proved in [37] that almost all graphs with an even number of edges have an even decomposition. A stronger quantitative estimate is established in [24], where the authors show that the fraction of graphs on h vertices with an even number of edges that do not admit an even decomposition is $e^{-\Omega(h^2)}$. An intriguing conjecture suggested in [37] is that for any H with an even number of edges $r_H(n)$ is at least n^{c_H} for some $c_H > 0$ if and only if H admits an even decomposition.

The extremal case dealing with graphs H with the largest possible growth of $r_H(n)$ is also interesting, suggesting the following question.

Problem 6.2. *Characterize all graphs H so that $r_H(n) \geq c(H)n^2$ for some constant $c(H) > 0$.*

Note that trivially $r_H(n) \leq \binom{n}{2} (< n^2/2)$ for any H, as shown by the coloring assigning different colors to all edges of K_n. Together with Narang we proved that for every star-forest H in which the number of edges in each connected component is even and is at most half the total number of edges of H, $r_H(n)$ grows quadratically as a function of n.

7 Connectivity graph-codes

For a fixed graph H call a collection $\mathcal{G}$ of spanning subgraphs of H a *connectivity (graph) code for H* if the symmetric difference of any two distinct subgraphs in $\mathcal{G}$ is a connected spanning subgraph of H. Note that a usual error-correcting code is good if its minimum distance is large, namely, the symmetric difference of any two distinct codewords has a large Hamming weight. When considering subgraphs, it is natural to interpret "large" as connected and spanning. Let $m(H)$ denote the maximum possible cardinality of a connectivity graph-code for H.

It is easy to see that the maximum possible cardinality of a connectivity code for H is at most $2^{k'(H)} \leq 2^{\delta(H)}$, where $k'(H)$ is the edge-connectivity of H and $\delta(H)$ is its minimum degree. In [4] it is shown that equality holds if H is the complete graph K_n, that is, $m(K_n) = 2^{n-1}$. In [6] it is proved that equality holds also for the 3 by 3 torus $C_3 \times C_3$. This is the product of two cycles of length 3 in which two vertices are adjacent iff they are equal in one coordinate and adjacent in the other. The edge connectivity (and degree of regularity) here is 4, and it is shown in [6] that $m(C_3 \times C_3) = 2^4 = 16$. The same authors also showed that $m(C_4 \times C_4) = 2^4 = 16$ and asked whether or not for all $m \geq n \geq 3$, $m(C_m \times C_n) = 2^4 = 16$. In [3] it is proved that this is not the case: for $n = 3$ and any $m \geq 37$, $m(C_m \times C_n) \leq 8$. On the other hand, it is proved in [3] that for any (mild) d-regular expander H, where $d \geq d_0$ for some (explicit) absolute constant d_0, $m(H) = 2^d$. The definition of a mild expander here is that any set W of at most half the vertices of H is incident with at least $9|W|\log d$ edges connecting it to its complement. Such expanders are known to exist for every $d \geq d'$ for some absolute constant d', see for example, [1].

The proof actually shows that any such expander admits a linear connectivity graph-code of size 2^d. This is established by first proving that a graph $H = (V, E)$ admits a linear connectivity code of size 2^d if and only if one can assign to each edge $e \in E$ a vector $v(e) \in Z_2^d$, so that for every cut $(S, V-S) = \{e \in E : e \cap S \neq \emptyset, e \cap (V-S) \neq \emptyset\}$ the set of vectors $\{v(e), e \in (S, V-S)\}$ spans Z_2^d.

In order to use this fact to get the required connectivity code it is required to show that for any mild d-regular expander H it is possible to assign each edge e a vector $v(e) \in Z_2^d$ so that the vectors assigned to the edges of each cut of H span Z_2^d. In particular, the vectors assigned to all edges incident with any single vertex must form a basis. The expansion properties of the graph ensure that the cuts which are not 1-vertex cuts have significantly more edges than the 1-vertex cuts, and hence intuition suggests that it will be simpler to ensure their vectors span the whole space. A natural approach to try and prove the existence of the required binary vectors is probabilistic, assigning vectors randomly. Since, however, the probability that d random vectors of Z_2^d form a basis is exactly $\prod_{i=1}^{d}(1-2^{-i})$, which is bounded away from 1 (indeed smaller than $1/2$), special care is needed to ensure that the vectors assigned to all edges in every 1-vertex cut form a basis. To do so, it is useful not to assign vectors randomly to all edges, but only to most of them. A simple yet useful observation here is the following.

Fact: Let E' be a subset of edges of a d-regular graph $H = (V, E)$, and let $v(e), e \in E'$ be an assignment of a vector in Z_2^d for any edge $e \in E'$. Suppose that for every vertex u the set of vectors $v(e)$ assigned to all edges in E' that are incident with u is linearly independent. Then it is possible to complete the given partial assignment by assigning a vector $v(e) \in Z_2^d$ to every edge $e \in E - E'$, so that for every vertex u, the set of vectors $v(e)$ assigned to all d edges incident with it forms a basis of Z_2^d.

Using this fact it is now possible to take a spanning subgraph (V, E') of the given d-regular expander $H = (V, E)$ in which all degrees are close to d, and assign random vectors from Z_2^d to all edges of E'. It is possible to apply the (asymmetric) Lovász Local Lemma and ensure that the following two properties hold:

- For each vertex v, the vectors assigned to the edges of E' incident with v are linearly independent.
- For all cuts that are not cuts of single vertices, the labels assigned to the edges

of E' in these cuts already span Z_2^d.

The fact above can then be used to assign binary vectors to the remaining edges, satisfying the desired properties.

The application of the local lemma here requires some technical computation, see [3] for the details.

Let $f(d)$ be the largest f so that there are infinitely many d-regular graphs H satisfying $m(H) = f$. By the above mentioned result about the connectivity codes of expanders there exists an absolute constant d_0 so that $f(d) = 2^d$ for every $d \geq d_0$. On the other hand it is clear that $f(d) < 2^d$ for $d = 2$ and using the Plotkin bound from coding theory [33], [29] it is not difficult to show that $f(3) < 8 = 2^3$.

Question 7.1. *Is $f(d) = 2^d$ for all $d \geq 4$?*

It seems interesting to study the computational problem of computing or estimating $m(H)$ for a given input graph H. As already mentioned, $m(H)$ is always at most $2^{k'(H)}$, where $k'(H)$ is the edge connectivity of H. On the other hand, $m(H)$ is always at least $2^{\lfloor k'(H)/2 \rfloor}$. This is because an immediate consequence of the known result of Nash-Williams about packing edge-disjoint trees in graphs [32] implies that H has at least $k = \lfloor k'(H)/2 \rfloor$ pairwise edge disjoint spanning trees T_i. The collection of all 2^k unions $\cup_{i \in I} E(T_i)$ of the edge sets of any subset I of these trees is a (linear) connectivity code for H. Since $k'(H)$ can be computed in polynomial time, this supplies an efficient algorithm for approximating the logarithm of $m(H)$ up to a factor of (roughly) 2.

Question 7.2. *Is there an efficient algorithm that approximates $m(H)$, for a given input graph H, up to a constant factor? Is the problem of computing $m(H)$ NP-hard? APX-hard?*

It can be shown that the following somewhat related computational problem is NP-hard.

Proposition 7.3. *The problem of computing the maximum possible cardinality of a collection of subgraphs of a given input graph, so that the minimum degree of the symmetric difference of any two of them is at least 2, is NP-hard.*

It turns out that this problem is NP-hard even for cubic graphs. For such graphs it is clear that the maximum possible size of such a collection of subgraphs cannot exceed 4. Indeed, fix a vertex v and observe that the sets of neighbors of v of the members of the collection must form a code of length 3 and minimum distance at least 2. It is clear that such a code cannot have more than 4 distinct elements.

We claim that a cubic graph $G = (V, E)$ contains 4 subgraphs satisfying the property in the proposition above if and only if it has chromatic index 3, that is, its set of edges can be decomposed into 3 perfect matchings. To see it note that if there are such matchings, then the empty subgraph together with the three unions of two of the matchings is a collection of 4 subgraphs with the desired property. If there are such 4 subgraphs, then without loss of generality one of them is the edgeless subgraph. It is not difficult to check that this implies that each of the other 3 subgraphs (V, E_i), $1 \leq i \leq 3$, must be 2-regular, and so is any symmetric difference of two of them. This implies that the three sets of edges $E - E_i$ are perfect matchings defining a proper 3-edge coloring of G, establshing the claim.

The desired NP-hardness now follows from the known result of Holyer [23] that the problem of deciding whether or not the chromatic index of a given cubic graph is 3, is NP-complete.

It is worth noting that the assertion of Proposition 4.3 can be extended to families of subgraphs of any given regular class 1 graph (and not only of K_n for even n as in that proposition).

Proposition 7.4. *Let H be a k-regular graph class* 1 *graph, that is, a k-regular graph whose edges can be colored properly by k colors. Then the maximum cardinality of a family of subgraphs of G so that the symmetric difference of any two distinct ones has minimum degree at least d is exactly the maximum possible cardinality of a binary error-correcting code of length k and minimum distance at least d.*

8 A concluding remark

This short survey contains several open questions dealing with Graph-Codes. While some of these problems may be very difficult, some look doable. Let me finish by quoting a sentence from the paper of Paul Erdős [16] in which he describes his favorite open problems in Combinatorics:

“a subject needs challenging unsolved and not quite hopeless problems to keep alive.”

I believe that the study of the problems discussed here is likely to keep the subject active.

Acknowledgment: I thank an anonymous referee for helpful comments.

References

[1] Noga Alon, Explicit expanders of every degree and size, Combinatorica 41 (2021), no. 4, 447–463.

[2] Noga Alon, Graph-codes, European J. Combinatorics 116 (2024), Paper No. 103880, 7 pp.

[3] Noga Alon, Connectivity graph-codes, Random Structures and Algorithms 65 (2024), no. 3, 451–459.

[4] Noga Alon, Anna Gujgiczer, János Körner, Aleksa Milojević and Gábor Simonyi, Structured Codes of Graphs, SIAM J. Discrete Math. 37 (2023), no. 1, 379–403.

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

[6] Bo Bai, Yu Gao, Jie Ma and Yuze Wu, Phase transitions of structured codes of graphs, SIAM J. Discrete Math. 38 (2024), no. 2, 1902–1914.

[7] Patrick Bennett, Emily Heath and Shira Zerbib, Edge-coloring a graph G so that every copy of a graph H has an odd color class, arXiv:2307.01314, 2023.

[8] Zenon Ivanovich Borevich and Igor Rostislavovich Shafarevich, *Number Theory*, Academic Press, New York, 1966.

[9] Aaron Berger and Yufei Zhao, K_4-intersecting families of graphs, J. Combin. Theory Ser. B 163 (2023), 112–132.

[10] Omar Antolin Camarena, Endre Csóka, Tamás Hubai, Gábor Lippner and László Lovász, Positive graphs, European J. Combin. 52B (2016), 290–301.

[11] Fan R. K. Chung, Peter Frankl, Ronald L. Graham and James B. Shearer, Some intersection theorems for ordered sets and graphs, J. Combin. Theory, Ser. A 43 (1986), 23–37.

[12] David Conlon, Jacob Fox, Chungbum Lee and Benny Sudakov, The Erdős-Gyárfás problem on generalized Ramsey numbers, Proc. London Math. Soc. 110 (2015), 1–18.

[13] Alex Cameron and Emily Heath, New upper bounds for the Erdős-Gyárfás problem on generalized Ramsey numbers, Combinatorics, Probability and Computing 32(2):349–362, 2023.

[14] David Conlon, Joonkyung Lee and Leo Versteegen, Around the positive graph conjecture, arXiv:2404.17467, 2024.

[15] David Ellis, Yuval Filmus and Ehud Friedgut, Triangle-intersecting families of graphs, Journal of the European Mathematical Society 14 (2012), No. 3, 841–885.

[16] Paul Erdős, On the combinatorial problems which I would most like to see solved, Combinatorica 1 (1981), no. 1, 25–42.

[17] Paul Erdős, Peter Frankl and Vojtěch Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, Graphs Combin. 2 (1986), no. 2, 113–121.

[18] Paul Erdős and András Gyárfás, A variant of the classical Ramsey problem, Combinatorica 17 (1997), 459–467.

[19] Peter Frankl and Zoltan Füredi, Forbidding just one intersection, J. Combin. Theory Ser. A 39 (1985), no. 2, 160–176.

[20] Jacob Fox and Benny Sudakov, Ramsey-type problem for an almost monochromatic K_4, SIAM J. Discrete Math. 23(1) (2008), 155–162.

[21] Gennian Ge, Zixiang Xu and Yixuan Zhang, A new variant of the Erdős–Gyárfás problem on K_5, arXiv:2306.14682, 2023.

[22] William Timothy Gowers, The first unknown case of polynomial DHJ, Blogpost, 2009. https://gowers.wordpress.com/2009/11/14/the-first-unknown-case-of-polynomial-dhj/

[23] Ian Holyer, The NP-completeness of edge-coloring, SIAM J. Comput. 10 (1981), no. 4, 718–720.

[24] Oliver Janzer and Fredy Yip, The probability that a random graph is even-decomposable, Random Structures Algorithms 67 (2025), no. 3, Paper No. e70034.

[25] Daniel J. Kleitman, On a Combinatorial Conjecture of Erdős, Journal of Combinatorial Theory 1 (1966), 209–214.

[26] Anton Kotzig, Hamilton graphs and Hamilton circuits, in: Theory of Graphs and its Applications, Publ. House Czechoslovak Acad. Sci., Prague, 1964.

[27] László Lovász, On the Shannon capacity of a graph, IEEE Transactions on Information Theory IT-25, (1979), 1-7.

[28] László Lovász, *Combinatorial Problems and Exercises*, North Holland, Amsterdam, 1979, Problem 11.8.

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

[30] Dhruv Mubayi, Edge-coloring cliques with three colors on all 4-cliques, Combinatorica 18 (1998), no. 2, 293–296.

[31] Nitya Mani and Owen Jianwen Zhang, Tetrahedron-intersecting families of 3-uniform hypergraphs, manuscript, 2025.

[32] Crispin St. J. A. Nash-Williams, Edge-disjoint spanning trees of finite graphs, J. London Math. Soc. 36 (1961), 445–450.

[33] Morris Plotkin, Binary codes with specified minimum distance, IRE Transactions on Information Theory 6 (1960), 445–450.

[34] Alexander Rosa, Perfect 1-factorizations, Mathematica Slovaca, 69 (2019), no. 3, 479–496.

[35] Wolfgang M. Schmidt, *Equations over Finite Fields, an Elementary Approach*, Springer Verlag Lecture Notes in Math., 1976.

[36] Endre Szemerédi, Regular partitions of graphs, Colloques Internationaux C.N.R.S. No 260-Problémes Combinatoires et Théorie des Graphes, Orsay (1976).

[37] Leo Versteegen, Upper bounds for linear graph-codes, Random Structures Algorithms 66 (2025), no. 1, Paper No. e21263, 7 pp.

[38] Fredy Yip, A variant of the Erdős-Gyárfás problem for K_8, European J. Combin. 132 (2026), part B, Paper No. 104267.