SIAM J. DISCRETE MATH.
Vol. 37, No. 1, pp. 379–403
© 2023 Society for Industrial and Applied Mathematics

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

STRUCTURED CODES OF GRAPHS*

NOGA ALON†, ANNA GUJGICZER‡, JÁNOS KÖRNER§, ALEKSA MILOJEVIĆ¶, AND GÁBOR SIMONYI‖

Abstract. We investigate the maximum size of graph families on a common vertex set of cardinality n such that the symmetric difference of the edge sets of any two members of the family satisfies some prescribed condition. We solve the problem completely for infinitely many values of n when the prescribed condition is connectivity or 2-connectivity, Hamiltonicity, or the containment of a spanning star. We also investigate local conditions that can be certified by looking at only a subset of the vertex set. In these cases a capacity-type asymptotic invariant is defined and when the condition is to contain a certain subgraph this invariant is shown to be a simple function of the chromatic number of this required subgraph. This is proven using classical results from extremal graph theory. Several variants are considered and the paper ends with a collection of open problems.

Key words. extremal problems, perfect 1-factorization, induced subgraphs, the regularity lemma

MSC codes. 05C35, 05C51, 05C70, 94B25

DOI. 10.1137/22M1487989

1. Introduction.

Celebrated problems of extremal combinatorics may get an exciting new flavor when the presence of some special structure is imposed in the condition. A prominent example is the famous Simonovits–Sós conjecture [38] proven by Ellis, Filmus, and Friedgut [17], which determines the maximum possible cardinality of a family of graphs on n labeled vertices in which the intersection of any two members contains a triangle. (The result of [17] shows, along with several far reaching generalizations, that the best is to take all graphs containing a given triangle, just as it was conjectured in [38]. This is clearly reminiscent of the Erdős–Ko–Rado theorem [20].) As another example we can also mention the Ramsey type problem investigated in [26] that was also initiated by a question of Sós and can be considered as a graph version of the first unsolved case of the so-called perfect hashing problem. (For details we refer to [26]).

* Received by the editors March 31, 2022; accepted for publication September 29, 2022; published electronically March 3, 2023.
https://doi.org/10.1137/22M1487989
Funding: The first author's research was supported in part by NSF grant DMS-1855464 and BSF grant 2018267. The second author's research was partially supported by the National Research, Development and Innovation Office (NKFIH) grant K-120706 of NKFIH Hungary. The fifth author's research was partially supported by the National Research, Development and Innovation Office (NKFIH) grants K-120706, K-132696, and SNN-135643 of NKFIH Hungary.

† Department of Mathematics, Princeton University, Princeton, NJ 08544, USA, and Schools of Mathematics and Computer Science, Tel Aviv University, Tel Aviv 69978, Israel (nalon@math.princeton.edu).

‡ Department of Computer Science and Information Theory, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics and MTA-BME Lendület Arithmetic Combinatorics Research Group, ELKH, Budapest, H-1117, Hungary (gujgicza@cs.bme.hu).

§ Sapienza University of Rome, Rome, 00198, Italy (korner@di.uniroma1.it).

¶ Department of Mathematics, Princeton University, Princeton, NJ 08544, USA (aleksam@princeton.edu).

‖ Alfréd Rényi Institute of Mathematics, Budapest, H-1053, Hungary, and Department of Computer Science and Information Theory, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, H-1111, Hungary (simonyi@renyi.hu).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

In this paper we study several problems we arrive at if the basic code distance problem (how many binary sequences of a given length can be given at most if any two differ in at least a given number of coordinates) is modified so that we do not prescribe the minimum distance of any two codewords but require that they differ in some specific structure. In particular, just as in the Simonovits–Sós problem, we seek the largest family of (not necessarily induced) subgraphs of a complete graph such that the symmetric difference of the edge sets of any two graphs in the family has some required property. We will consider properties like connectedness, Hamiltonicity, containment of a triangle, and some more. Formally, all of these can be described by saying that the graph defined by the symmetric difference of the edge sets of any two of our graphs belongs to a prescribed family of graphs (namely those that are connected, contain a Hamiltonian cycle, or contain a triangle, etc.).

Let $\mathcal{F}$ be a fixed class of graphs. A graph family $\mathcal{G}$ on n labeled vertices is called $\mathcal{F}$-good if for any pair of distinct $G, G' \in \mathcal{G}$ the graph $G \oplus G'$ defined by

$$V(G \oplus G') = V(G) = V(G') = [n],$$

where $[n] = \{1, \dots, n\}$ and

$$E(G \oplus G') = \{e : e \in (E(G) \setminus E(G')) \cup (E(G') \setminus E(G))\}$$

belongs to $\mathcal{F}$.

Let $M_{\mathcal{F}}(n)$ denote the maximum possible size of an $\mathcal{F}$-good family on n vertices. We are interested in the value of $M_{\mathcal{F}}(n)$ for various classes $\mathcal{F}$. We will give exact answers or both lower and upper bounds in several cases.

We mention that codes where the codewords are described by graphs already appear in the literature. In [40], for example, Tonchev looked at the usual code distance problem restricted to codes whose codewords are characteristic vectors of edge sets of graphs. Gray codes on graphs are also considered (see [32]), where the graphs representing the codewords should have some similarity properties if they are consecutive in a certain listing. Problems analogous to the present ones, though restricted to special graph classes, were also considered in [27,14]. A very interesting result along these lines is the one in [29].

This paper is organized as follows. In section 2 we give a general upper bound that will turn out to be sharp in several of the cases we consider. In section 3 we consider classes $\mathcal{F}$ defined by some global criterion as connectivity or 2-connectivity, Hamiltonicity, or containing a full star, that is, a vertex of degree $n-1$. We determine $M_{\mathcal{F}}(n)$ for infinitely many values of n and for all n in the first and the last case. In most of the cases when we give sharp bounds it is via also solving the problem we call dual: we give the largest possible size of a graph family for which the symmetric difference of no two of its members satisfies the original requirement. The case of the full star is an exception in this sense, nevertheless we also solve the dual problem in that case for all even n by using a celebrated lemma of Shearer. In section 4 we consider classes $\mathcal{F}$ defined by local conditions. This means that for certifying the condition it is enough to see just a special part of the graph pair at hand. A capacity-type asymptotic invariant is natural to define in these cases. It turns out that when the requirement is that the pairwise symmetric differences contain a certain subgraph, then this asymptotic invariant depends only on the chromatic number of the graph to be contained. We also discuss the case when the above containment is required in an induced manner, and obtain similar results in this case. The final section contains a collection of open problems.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

2. A general upper bound. To bound $M_{\mathcal{F}}(n)$ for various graph classes $\mathcal{F}$ it will often be useful to also consider the related problem of constructing large graph families in which no pair satisfies the condition prescribed by $\mathcal{F}$.

DEFINITION 1. *For a class of graphs $\mathcal{F}$ let $D_{\mathcal{F}}(n)$ denote the maximum possible size of a graph family on n labeled vertices (that is, each member of the family has $[n] = \{1, \ldots, n\}$ as vertex set), the symmetric difference of no two members of which belongs to $\mathcal{F}$. Determining $D_{\mathcal{F}}(n)$ will be referred to as the dual problem of determining $M_{\mathcal{F}}(n)$.*

Note that denoting by $\overline{\mathcal{F}}$ the class containing exactly those graphs that do not belong to $\mathcal{F}$ we actually have

$$D_{\mathcal{F}}(n) = M_{\overline{\mathcal{F}}}(n),$$

that is, the requirement of having no symmetric difference in $\mathcal{F}$ is clearly the same as saying that all symmetric differences belong to the complementary family $\overline{\mathcal{F}}$. Nevertheless, we will use the $D_{\mathcal{F}}(n)$ notation to emphasize the dual nature of the problem in those cases.

LEMMA 2.1. *For any graph class $\mathcal{F}$ we have*

$$M_{\mathcal{F}}(n) \cdot D_{\mathcal{F}}(n) \leq 2^{\binom{n}{2}}.$$

Proof. Let us define a graph $H_{\mathcal{F}}$ whose vertices are all the possible (simple) graphs on the vertex set $[n]$. Connect two such vertices if and only if the corresponding pair of graphs have their symmetric difference belonging to $\mathcal{F}$. Then by definition we have

$$M_{\mathcal{F}}(n) = \omega(H_{\mathcal{F}}) \quad \text{and} \quad D_{\mathcal{F}}(n) = \alpha(H_{\mathcal{F}}),$$

where $\omega(H)$ and $\alpha(H)$ denote the clique number and the independence number of the graph H, respectively. Observe that $H_{\mathcal{F}}$ is vertex-transitive, (in fact, it is a Cayley graph of the group $Z_2^{\binom{n}{2}}$). Indeed, if G_1 and G_2 are two graphs forming vertices of $H_{\mathcal{F}}$, then taking the symmetric difference of all n-vertex graphs forming vertices of $H_{\mathcal{F}}$ with the graph $G_1 \oplus G_2$ is an automorphism of $H_{\mathcal{F}}$ that maps G_1 to G_2. Since a vertex-transitive graph H always satisfies $\alpha(H)\omega(H) \leq |V(H)|$ (this can be seen by using that the fractional chromatic number $\chi_f(H)$ always satisfies $\omega(H) \leq \chi_f(H)$, while if H is a vertex-transitive graph, we also have $\chi_f(H) = \frac{|V(H)|}{\alpha(H)}$; cf. [36]), the statement follows. ∎

The above lemma makes it possible to bound $M_{\mathcal{F}}(n)$ from above by bounding $D_{\mathcal{F}}(n)$ from below. In particular, whenever we construct two families of graphs $\mathcal{A}$ and $\mathcal{B}$ on $[n]$ such that $A, A' \in \mathcal{A}$ implies $A \oplus A' \in \mathcal{F}$ and $B, B' \in \mathcal{B}$ implies $B \oplus B' \notin \mathcal{F}$, while $|\mathcal{A}||\mathcal{B}| = 2^{\binom{n}{2}}$, then we know that $|\mathcal{A}|$ and $|\mathcal{B}|$ realize the optimal values $M_{\mathcal{F}}(n)$ and $D_{\mathcal{F}}(n)$ for such families. Below we will see several cases when this simple technique can indeed be used to obtain these optimal values. An exception to this phenomenon is also presented by Theorems 6 and 7.

Remark 2.2. It is worth noting that Lemma 2.1 can be proven in a different way, with no reference to the fractional chromatic number. Indeed, if $G_1, \ldots, G_k$ is an $\mathcal{F}$-good family, while $T_1, \ldots, T_m$ is a family satisfying the conditions of the dual problem, then all the symmetric differences of the form $G_i \oplus T_j$ are different, implying $km \leq 2^{\binom{n}{2}}$. This is true because if $G_i \oplus T_j$ and $G_r \oplus T_s$ would be the same for

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

some $\{i,j\} \neq \{r,s\}$, then $(G_i \oplus T_j) \oplus (G_r \oplus T_s)$ would be the empty graph that could also be written (by commutativity and associativity of the symmetric difference) as $(G_i \oplus G_r) \oplus (T_j \oplus T_s)$. This would mean that $G_i \oplus G_r$ and $T_j \oplus T_s$ are two identical graphs. But if one of them is the empty graph (that is, $G_i = G_r$ or $T_j = T_s$), then the other cannot be empty, and if both are nonempty, then one of them belongs to $\mathcal{F}$ while the other one does not, so this is impossible. □

3. Global conditions.

3.1. Connectivity. When we speak about the class of connected graphs in the following theorem, we mean graphs with a single connected component, and hence without isolated vertices.

THEOREM 1. *Let $\mathcal{F}_c$ denote the class of connected graphs. Then*

$$M_{\mathcal{F}_c}(n) = 2^{n-1}.$$

Proof. First, we give a very simple dual family $\mathcal{B}_c$. Let it consist of all graphs on $[n]$ in which the vertex labeled n is isolated. Clearly, $|\mathcal{B}_c| = 2^{\binom{n-1}{2}}$ (that is the number of all graphs on $[n-1]$) and n is also isolated in the symmetric difference of any two of them, so no such symmetric difference is connected, This gives $D_{\mathcal{F}_c}(n) \geq 2^{\binom{n-1}{2}}$ and thus by Lemma 2.1 we have

$$M_{\mathcal{F}_c}(n) \leq 2^{\binom{n}{2}-\binom{n-1}{2}} = 2^{n-1}.$$

Now we show that this upper bound can be attained. Let the family $\mathcal{A}_c$ consist of all those graphs on $[n]$ that are the vertex-disjoint union of two complete graphs (where each vertex belongs to one of them) including the case when one of the two is on the empty set. Clearly, the number of these graphs is just half the number of subsets of $[n]$, that is exactly 2^{n-1}. All we have to show is that the symmetric difference of any two of these graphs is connected. Choose two arbitrary graphs G and G' from our family. Let G be the union of complete graphs on the complementary vertex sets K and L, while G' be the same on K' and L'. Let $A = K \cap L', B = L' \cap L, C = L \cap K'$, and $D = K' \cap K$. It is possible that one, but (if neither of K, L, K‘, L‘ is empty) only one, of A, B, C, D is empty. The edges of $G \oplus G'$ are all the edges of the complete bipartite graph with partite classes $A \cup C$ and $B \cup D$, so it must be connected. □

With just a little more consideration one can also treat the case of 2-connectedness at least for even n.

THEOREM 2. *Let $\mathcal{F}_{2c}$ denote the class of 2-connected graphs. Then if n is even, we have*

$$M_{\mathcal{F}_{2c}}(n) = 2^{n-2}.$$

Proof. The proof is a modification of the previous one. Therefore, we use the notation introduced there. The construction given there may result in symmetric differences that are not 2-connected only if $A \cup C$ or $B \cup D$ contains only one element. For even n this can be avoided if we consider only such graphs in our construction where the bipartition of $[n]$ defining the individual graphs has an even number of elements in both partite classes K and L. This proves the lower bound.

For the upper bound we consider all graphs in which the vertex n is either isolated or it has one fixed neighbor, say $n-1$. The symmetric difference of any two such graphs is not 2-connected, since n has at most one neighbor in it. The number of

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

such graphs is just twice the number of graphs in which n is an isolated point, that is, $2^{\binom{n-1}{2}+1}$ proving the matching upper bound by Lemma 2.1. ∎

Remark 3.1. The upper bound proven in Theorem 2 clearly holds also for odd n but we have not found a matching construction in general. For $n=3$ a triangle and the empty graph would do, still achieving the upper bound. But for larger odd n the best we could do is to take only those graphs from our construction for which in the corresponding bipartition the smaller partition class has an odd number of elements if $n \equiv 1 \pmod 4$ and it has an even number of elements if $n \equiv 3 \pmod 4$. The number of graphs obtained this way is $2^{n-2} - \binom{(n-2)}{(n-3)/2}$. □

Remark 3.2. Changing the graphs to their complements in the proofs of Theorems 1 and 2 makes these graph families vector spaces over the 2-element field, while they still satisfy the conditions as the symmetric differences do not change by complementation (or by taking the symmetric difference of all elements with any fixed graph which is the complete graph in case of complementation). □

It does not sound surprising that if we step further on to k-connectedness for $k>2$, then the problem becomes rather more complicated. Nevertheless, if we insist on linear codes, that is graph families closed under the symmetric difference operation, then for $k=3$ we can still determine the largest possible cardinality for infinitely many values on n using Hamming codes.

THEOREM 3. *Let $\mathcal{F}_{3c}$ be the class of 3-connected graphs, and let $M^{(\ell)}_{\mathcal{F}_{3c}}(n)$ denote the size of a largest possible linear graph family on vertex set $[n]$ any two members of which have a 3-connected symmetric difference. If $n=2^k-1$ for some integer $k\geq 2$, then*

$$M^{(\ell)}_{\mathcal{F}_{3c}}(n) = 2^{n-k-1}.$$

Proof. First, we prove that $D_{\mathcal{F}_{3c}}(n) \geq n2^{\binom{n-1}{2}}$ holds in general. Consider the family of all graphs on vertex set $[n]$ in which the degree of vertex n is at most 1. There are exactly $n2^{\binom{n-1}{2}}$ such graphs. The symmetric difference of any two of these graphs is at most 2-connected, since the vertex n has degree at most 2 in all these symmetric differences. This proves the claimed inequality, and by Lemma 2.1 this implies $M_{\mathcal{F}_{3c}}(n) \leq 2^{n-1}/n$.

It is well known that if a family of subsets of a finite set contains the empty set and is closed under the symmetric difference operation, then the cardinality of this set must be a power of 2. This follows immediately from linear algebra and the fact that such a family forms a vector space over $GF(2)$; cf., e.g., Lemma 3.1 in Kozlov's book [31] where a simple combinatorial proof of this fact is also presented. Since a linear graph family code on $[n]$ can be viewed as a collection of subsets of $E(K_n)$, this implies that $M^{(\ell)}_{\mathcal{F}_{3c}}(n)$ is a power of 2. Since we obviously have $M^{(\ell)}_{\mathcal{F}_{3c}}(n) \leq M_{\mathcal{F}_{3c}}(n)$, the upper bound proved above implies $M^{(\ell)}_{\mathcal{F}_{3c}}(n) \leq 2^d$ with $d = \lfloor \log_2 \frac{1}{n} 2^{n-1} \rfloor$ giving

$$M^{(\ell)}_{\mathcal{F}_{3c}}(n) \leq 2^{n-k-1}$$

for $n=2^k-1, k\geq 2$, which proves the required upper bound.

For the lower bound consider the Hamming code $\mathcal{C}_H(n)$ with length $n=2^k-1$ that exists for every $k\geq 2$. (For a nice quick account on Hamming codes, see, e.g., [10].) It is a linear code with minimum distance three that consists of 2^{n-k} binary codewords having the property that if $\boldsymbol{c}=(c_1,\ldots,c_n)$ belongs to the code, then so does $\bar{\mathbf{c}}=(\bar{c}_1,\ldots,\bar{c}_n)$ also, where $\bar{c}_i = 1-c_i$. For each codeword $\boldsymbol{c}\in\mathcal{C}_H(n)$ consider the bipartition of $[n]$ into the subsets $K_{\boldsymbol{c}}, L_{\boldsymbol{c}}$, where $K_{\boldsymbol{c}} = \{i : c_i = 0\}, L_{\boldsymbol{c}} = \{i : c_i = 1\}$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

and the complete bipartite graph $G_{K_{\boldsymbol{c}},L_{\boldsymbol{c}}}$ with partite classes $K_{\boldsymbol{c}}, L_{\boldsymbol{c}}$. Note that by the above mentioned property of Hamming codes we have $\boldsymbol{c} \in \mathcal{C}_H(n)$ if and only if $\bar{\boldsymbol{c}} \in \mathcal{C}_H(n)$ and thus since $G_{K_{\boldsymbol{c}},L_{\boldsymbol{c}}} = G_{K_{\bar{\boldsymbol{c}}},L_{\bar{\boldsymbol{c}}}}$, we get $\frac{1}{2}|\mathcal{C}_H(n)| = 2^{n-k-1}$ different complete bipartite graphs this way. All we have to prove is that the symmetric difference of any two of our graphs is 3-connected. This is equivalent to show that if $\mathbf{c}' \neq \mathbf{c}, \bar{\mathbf{c}}$, then the cardinality of both partite classes of $G_{K_{\boldsymbol{c}},L_{\boldsymbol{c}}} \oplus G_{K_{\boldsymbol{c}'},L_{\boldsymbol{c}'}}$, that is, of $(K_{\boldsymbol{c}} \cap K_{\boldsymbol{c}'}) \cup (L_{\boldsymbol{c}} \cap L_{\boldsymbol{c}'})$ and $(K_{\boldsymbol{c}} \cap L_{\boldsymbol{c}'}) \cup (K_{\boldsymbol{c}'} \cap L_{\boldsymbol{c}})$ is at least three. However, this immediately follows from the fact that the codeword $\mathbf{c}'$ must differ from both $\mathbf{c}$ and $\bar{\mathbf{c}}$ in at least three coordinates. This completes the proof. ∎

3.2. Hamiltonicity. A graph is connected if and only if it contains a spanning tree. Next, we consider what happens if we require the containment of specific spanning trees: a path in this subsection and a star in the next one.

THEOREM 4. *Let $\mathcal{F}_{H\mathrm{p}}$ denote the class of graphs containing a Hamiltonian path. Then for infinitely many values of n we have*

$$M_{\mathcal{F}_{H\mathrm{p}}}(n) = 2^{n-1}.$$

In particular, this holds whenever $n = p$ or $n = 2p-1$ for some odd prime p.

To prove the above theorem we will refer to the following old conjecture that is known to be true in several special cases. To state it we need the notion of perfect 1-factorization. It means the partition of the edge set of a graph into perfect matchings such that the union of any two of them is a Hamiltonian cycle.

PERFECT 1-FACTORIZATION CONJECTURE (P1FC) (Kotzig [28]). *The complete graph K_n has a perfect 1-factorization for all even $n > 2$.*

This conjecture is still open, in general; however, it is known to hold in several special cases, for example, whenever $n = p + 1$ (Kotzig [28]) or $n = 2p$ for some odd prime p (Anderson [8] and Nakamura [33]; cf. also Kobayashi [24]). For a recent survey, see Rosa [34], according to which the smallest open case of the conjecture is $n = 64$.

Proof of Theorem 4. Since Hamiltonian paths are connected, it follows from the proof of Theorem 1 that 2^{n-1} is again an upper bound. Now we show that it is also a lower bound whenever the P1FC holds for $n+1$. (Note that if the conjecture is true, then this means that our statement holds for all odd numbers at least 3, while for 1 it is void.)

Let n be an odd number for which K_{n+1} has a perfect 1-factorization $\mathcal{M}$ and v a fixed vertex of K_{n+1}. Note that deleting the edge incident to v from all matchings belonging to $\mathcal{M}$ we obtain n matchings of K_n such that the union of any two of them is a Hamiltonian path in $K_n := K_{n+1} \setminus \{v\}$. Now consider all those subgraphs of K_n that can be obtained as the union of an even number of these n matchings. Clearly, the symmetric difference of any two of them is also the union of at least two of these matchings and thus contains a Hamiltonian path. The number of graphs obtained this way is 2^{n-1}, matching the upper bound. ∎

The case of Hamiltonian cycles can be treated essentially the same way.

THEOREM 5. *Let $\mathcal{F}_{H\mathrm{c}}$ denote the class of graphs containing a Hamiltonian cycle. For all even values of n for which the P1FC holds, we have*

$$M_{\mathcal{F}_{H\mathrm{c}}}(n) = 2^{n-2}.$$

In particular, this is the case if $n = p + 1$ or $n = 2p$ for some odd prime p.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Proof. Since Hamiltonian cycles are 2-connected, it follows from the proof of Theorem 2 that 2^{n-2} is again an upper bound.

Let n be an even number for which the P1FC holds, and let $\mathcal{M}$ be a perfect 1-factorization of K_n. Note that $\mathcal{M}$ contains $n-1$ matchings (indeed the edge-chromatic number of K_n for even n is $n-1$). Now consider the 2^{n-2} graphs we can obtain as the union of an even number of matchings from $\mathcal{M}$. Clearly, the symmetric difference of any two of them contains a Hamiltonian cycle. ∎

Remark 3.3. Since Hamiltonian cycles are 2-connected graphs the proof of Theorem 5 obviously gives an alternative proof of Theorem 2 for those values of n for which the P1FC is known to hold. (The situation is similar for Theorems 4 versus 1.) On the other hand, the construction in the proof of Theorem 2 utterly fails to give a good lower bound for the value of $M_{\mathcal{F}_{Hc}}(n)$ investigated in Theorem 5. Indeed, the symmetric difference of two graphs in the construction given in the proof of Theorem 2 contains a Hamiltonian cycle if and only if the sets denoted by $A \cup C$ and $B \cup D$ in that proof both have cardinality $\frac{n}{2}$ which happens exactly when the partition classes of the partitions (K, L) and (K', L') are orthogonal in the sense that by representing these bipartitions by characteristic vectors consisting of $+1$ and -1 coordinates in the obvious way, we get a collection of vectors that are pairwise orthogonal. So their number cannot be more than just n and we can give exactly n such vectors if and only if an $n \times n$ Hadamard matrix exists. □

3.3. Containing a spanning star. We have seen in the previous subsection that if we want every symmetric difference to contain a spanning tree which is a path, then for infinitely many values of n our family can be just as large as if we did not want more than just the connectedness of these symmetric differences. In this subsection we show that if the required spanning tree is a star, then the largest possible family is drastically smaller.

THEOREM 6. *Let $\mathcal{F}_S$ denote the class of graphs containing a spanning star, that is a vertex connected to all other vertices in the graph. Then we have*

$$M_{\mathcal{F}_S}(n) = \begin{cases} n+1 & \text{if } n \text{ is odd,} \\ n & \text{if } n \text{ is even.} \end{cases}$$

Proof. First, we prove the upper bound. Let $G_1, \ldots, G_m$ be an $\mathcal{F}_S$-good family on the vertex set $[n]$. Consider the complete graph K_m whose vertices are labeled with the graphs $G_1, \ldots, G_m$. For each edge $\{G_i, G_j\}$ of this graph assign an element $h \in [n]$ for which h is adjacent to all other elements of $[n]$ in the graph $G_i \oplus G_j$. By the definition of $\mathcal{F}_S$-goodness such an h exists for every pair of our graphs. Now observe that if an element $a \in [n]$ is assigned to two distinct edges e and f of our graph K_m, then e and f must be independent edges. Indeed, if that was not the case then we would have $e = \{G_i, G_j\}, f = \{G_i, G_k\}$ for some $i, j, k \in [n]$ and a would be a full-degree vertex (that is, one connected to all other vertices) in both of the graphs $G_i \oplus G_j$ and $G_i \oplus G_k$. But since $G_j \oplus G_k = (G_i \oplus G_j) \oplus (G_i \oplus G_k)$, that would mean that a is an isolated vertex in $G_j \oplus G_k$, so no vertex of this latter graph can have full degree contradicting the $\mathcal{F}_S$-goodness of our family. Thus our assignment of vertices from $[n]$ to the edges of our K_m partitions the edge set of K_m into sets of independent edges (every partition class consisting of the edges with the same assigned label). In other words, it defines a proper edge-coloring of K_m. This means that the number of possible labels, which is n, should be at least as large as the edge-chromatic number $\chi_e(K_m)$ of K_m. Since the latter is $m-1$ for even m and m for odd m, turning it

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

around we obtain that for odd n we must have $m \leq n+1$ and for even n we must have $m \leq n$.

Now we show that the upper bound we proved is sharp. First, assume that n is odd and consider a complete graph K_{n+1} on the vertices $v_1, \ldots, v_{n+1}$ along with an optimal edge-coloring $c\colon E(K_{n+1}) \to [n]$ of this graph. This edge-coloring partitions $E(K_{n+1})$ into n disjoint matchings $M_1, \ldots, M_n$, where M_j consists of the edges colored j for every $j \in [n]$. Now we construct the graphs $G_1, \ldots, G_{n+1}$ by telling for each potential edge ij of the complete graph on $[n]$ which G_k's will contain it and which ones will not. Consider the edge ij and the union of the matchings M_i and M_j (note that these matchings are in the "other" complete graph on $n+1$ vertices). This union is a bipartite graph on the vertex set $\{v_1, \ldots, v_{n+1}\}$ with two equal size partite classes A and B. Let ij be an edge of the graph G_k if and only if $v_k \in A$. (So ij will be an edge of exactly half of our graphs $G_1, \ldots, G_{n+1}$.) Do this similarly for all edges of K_n, the complete graph on vertex set $[n]$. This way we defined our $n+1$ graphs. We have to show that they form an $\mathcal{F}_S$-good family.

To this end consider two of our graphs, say G_h and G_k. The edge $\{v_h, v_k\}$ has got some color in our coloring c, call this color j. This means that $\{v_h, v_k\}$ belongs to the matching M_j. We claim this means that $j \in [n]$ is a full-degree vertex of $G_h \oplus G_k$. The latter is equivalent to the statement that every edge ji incident to the point j appears in exactly one of the graphs G_h and G_k. But this follows from the way we constructed our graphs: when we decided about the edge ji we considered the matchings M_i and M_j and the bipartite graph their union defines. Since $\{v_h, v_k\} \in M_j$, the points v_h and v_k are always in different partite classes of this bipartite graph, so whichever was called A, exactly one of v_h and v_k belonged to it. Thus the edge ij was declared to be an edge of exactly one of G_h and G_k. Since this is so for every $i \neq j$, j is indeed a full-degree vertex in $G_h \oplus G_k$.

Assume now that n is even. Then $n-1$ is odd and we can construct graphs $G_1, \ldots, G_n$ on vertex set $[n-1] = \{1, \ldots, n-1\}$ as given in the previous paragraph. These are not yet good, however, since we have an nth vertex that does not appear yet in any of the graphs. Note that we have $n-1$ matchings $M_1, \ldots, M_{n-1}$ involved in the construction so far whose indices are just the first $n-1$ vertices of our graphs. Think about the additional vertex n as the index of an additional "matching" M_n that has no edges at all. We decide about the involvement of the edges ni $(i < n)$ in our graphs analogously as we did for the earlier edges: Consider the bipartite graph $M_i \cup M_n$, which consists of just the edges of M_i, so it is a perfect matching on the vertex set $\{v_1, \ldots, v_n\}$. Let the two partite classes defined by this perfect matching be A and B and add the edge ni to the graph G_h if and only if v_h belongs to A. Now we can prove analogously to the odd case that the symmetric difference of any two of our graphs contains a vertex of degree $n-1$. Consider G_h and G_k. The edge between v_h and v_k in the auxiliary complete graph belongs to exactly one of the matchings M_j and every edge ij is in exactly one of the graphs G_h and G_k if $i \in \{1, \ldots, j-1, j+1, \ldots, n\}$. This completes the proof. ∎

The following remark is due to Gábor Tardos [39].

Remark 3.4. The statement and proof of the above theorem can also be presented in a more compact form as follows. There exists m graphs on vertex set $[n]$ forming an $\mathcal{F}_S$-good family if and only if $\chi_e(K_m) \leq n$. The proof is essentially the same as what was shown above but in the second part we do not have to distinguish between odd and even n, rather just say that $M_1, \ldots, M_n$ are the color classes of a proper

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

edge-coloring of K_m (some of which may be empty) and then define the graphs $G_1, \ldots, G_m$ the same way as above. □

The only graph family code proven to be optimal and nonlinear (or not the coset of a linear code) in this paper is the one appearing in the above Theorem 6. This is also the first case so far when the upper bound is proven without the use of Lemma 2.1. This suggests the question of what could be said about the dual problem in this case. The next theorem solves this dual problem for even values of n also showing that Lemma 2.1 would not give a sharp upper bound for $M_{\mathcal{F}_S}(n)$.

THEOREM 7. *If n is even, then*

$$D_{\mathcal{F}_S}(n) = 2^{\binom{n}{2}-\frac{n}{2}}.$$

When n is odd, then we have

$$2^{\binom{n}{2}-\frac{n+1}{2}} \le D_{\mathcal{F}_S}(n) \le 2^{\binom{n}{2}-\frac{n}{2}}.$$

For the proof we will need the following celebrated result from [13] (see also Corollary 15.7.7 in [7]).

Shearer's Lemma (see [13]). *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*

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

Proof of Theorem 7. We will prove

$$2^{\binom{n}{2}-\lceil \frac{n}{2} \rceil} \le D_{\mathcal{F}_S}(n) \le 2^{\binom{n}{2}-\frac{n}{2}}$$

that implies both the even and the odd case. For the lower bound fix a subgraph T of K_n with the minimum number $\lceil \frac{n}{2} \rceil$ of edges such that no vertex is isolated and take all possible subgraphs of K_n that contain none of the edges of T. The number of such subgraphs is $2^{\binom{n}{2}-\lceil \frac{n}{2} \rceil}$ and no two of them has a symmetric difference that contains all edges incident to any fixed vertex. This proves the lower bound.

For the upper bound consider a graph family $\mathcal{M}$ that satisfies the condition that no two of its elements have a symmetric difference with a vertex of degree $n-1$. For $i = 1, \ldots, n$ let S_i be the set of $n-1$ edges (of K_n) incident to 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. So 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, we can apply Shearer's lemma to these sets with $k = 2$. This gives

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

Taking square roots we get the upper bound. ∎

Note that if we restrict our attention to linear graph families for the dual problem treated in Theorem 7, then using again that the cardinality of such a family should

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

be a power of 2 (cf. the similar argument in the proof of Theorem 3) we get that our lower bound is also sharp for odd values of n.

4. Local conditions. In the previous section we investigated $M_{\mathcal{F}}(n)$ in cases when the required symmetric differences contain specific spanning subgraphs; therefore, to check whether these conditions are satisfied we have to consider our graphs on the whole vertex set. Now we turn to families $\mathcal{F}$ defined by containing some fixed small finite graphs, so the nature of these conditions will be local.

4.1. General local conditions.

DEFINITION 2. *A graph class $\mathcal{L}$ defines a local condition if it has the property that whenever H_1 is an induced subgraph of H_2 and H_1 belongs to $\mathcal{L}$, then so does also H_2. In short, we will refer to such an $\mathcal{L}$ as a local graph class.*

Note that the above definition implies that whenever two graphs F and G are in the $\mathcal{L}$-good relation (that is, $F \oplus G \in \mathcal{L}$), then any F' with $F'[U] \cong F$ and G' with $G'[U] \cong G$ for some $U \subseteq V(F') = V(G')$ (that is, F' and G' induce subgraphs isomorphic to F and G, respectively, on the same subset U of their vertex set) are also in the $\mathcal{L}$-good relation. This means that if two graphs are in this relation, then there is always some local certificate for this.

Here are some examples of local graph classes.

1. $\mathcal{L} = \{H : L \subseteq H\}$ for some fixed finite simple graph L. That is, $\mathcal{L}$ contains all graphs that contain a (not necessarily induced) subgraph isomorphic to L. When $\mathcal{L}$ is such a family we will use the simplified notation $M_L(n)$ for $M_{\mathcal{L}}(n)$.
2. $\mathcal{L} = \{H : L \subseteq_{ind} H\}$ for some fixed finite simple graph L. That is, $\mathcal{L}$ contains all graphs that have an induced subgraph isomorphic to L. When $\mathcal{L}$ is such a family we will use the simplified notation $M_{L,\mathrm{ind}}(n)$ for $M_{\mathcal{L}}(n)$. (Note that although the above two examples give different notions, the word "induced" is indeed needed in Definition 2.)
3. $\mathcal{L} = \mathcal{C}_{\mathrm{odd}} := \{H : C_{2k+1} \subseteq H \text{ for some integer } 1 \leq k\}$, that is, $\mathcal{C}_{\mathrm{odd}}$ contains all graphs that contain an odd cycle.
4. For some fixed integers h and ℓ we can define $\mathcal{L}_{h,\ell} = \{H : \exists U \subseteq V(H), |U| = h, |E(H[U])| = \ell\}$, that is, $\mathcal{L}_{h,\ell}$ is the class of all graphs that have an induced subgraph on h vertices with exactly ℓ edges.

In the following we prove some general results related to $M_{\mathcal{L}}(n)$ for local graph classes $\mathcal{L}$ and will further investigate the special case belonging to our first example above in the next subsection. In subsection 4.3 we will focus on $M_{K_3}(n)$ and $M_{\mathcal{C}_{\mathrm{odd}}}(n)$. In the final subsection we discuss the behavior of the functions $M_{L,\mathrm{ind}}(n)$ mentioned in the second example above.

The next proposition gives a straightforward upper bound on the value of $M_{\mathcal{L}}(n)$. It is in terms of $ex(n, \mathcal{L})$ that, as is usual in extremal graph theory, denotes the maximum number of edges a graph on n vertices can have without containing any $L \in \mathcal{L}$ as a subgraph.

PROPOSITION 8. *For any local graph class $\mathcal{L}$*

$$M_{\mathcal{L}}(n) \leq 2^{\binom{n}{2} - ex(n,\mathcal{L})}.$$

Proof. Consider an n-vertex graph H satisfying $|E(H)| = ex(n, \mathcal{L})$ and containing no subgraph isomorphic to any $L \in \mathcal{L}$. The family of all subgraphs of H clearly satisfies the requirements of the dual problem of $M_{\mathcal{L}}(n)$. This is because no subgraph

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

of H can have a subgraph belonging to $\mathcal{L}$ and the symmetric difference of two such subgraphs is also a subgraph of H so such a symmetric difference can also not contain any $L \in \mathcal{L}$. This family has size $2^{ex(n,\mathcal{L})}$, thus the claimed upper bound follows from Lemma 2.1. ∎

Proposition 8 and our following results will justify the relevance of the following notion in our current setting.

DEFINITION 3. *The rate $R_{\mathcal{L}}(n)$ of an optimal graph family code on n vertices satisfying the requirement prescribed by the local graph class $\mathcal{L}$ is defined as*

$$R_{\mathcal{L}}(n) := \frac{2}{n(n-1)} \log_2 M_{\mathcal{L}}(n).$$

We will soon see that the value $\limsup_{n\to\infty} R_{\mathcal{L}}(n)$ is strictly positive for any $\mathcal{L}$ belonging to this section. We will use the following theorem due to Wilson to show that the limit actually exists for all local graph classes.

Wilson's Theorem (see [42]). *For every finite simple graph T there exists a threshold $n_0(T)$ such that if $n > n_0(T)$ and the following two conditions hold, then the edge set of the complete graph K_n can be partitioned into subgraphs each of which is isomorphic to T. The two conditions are as follows:*

1. $\binom{n}{2}$ *is divisible by* $|E(T)|$;
2. $n-1$ *is divisible by the greatest common divisor of the degrees of vertices in* T.

Note that the two conditions in the above theorem are obviously necessary. The decomposition of K_n in the conclusion of the theorem is called a T-design when it exists; cf. [1].

THEOREM 9. *Let $\mathcal{L}$ be an arbitrary fixed local graph class. Then the value $\lim_{n\to\infty} R_{\mathcal{L}}(n)$ exists and is bounded from below by $R_{\mathcal{L}}(n)$ for every n.*

Proof. Let n be an arbitrary natural number, and let $\mathcal{G} = \{G_1, \dots, G_m\}$ be an optimal graph family code for $\mathcal{L}$ with $V(G_i) = [n], i \in \{1, \dots, m\}$, that is, one with $m = M_{\mathcal{L}}(n)$. By Wilson's theorem a K_n-design exists for K_N, whenever N is large enough and both $n-1$ divides $N-1$ and $\binom{n}{2}$ divides $\binom{N}{2}$. Take such an N and consider the K_n-design on K_N consisting of the subgraphs $K^{(1)}, \dots, K^{(r)}$, where $r = \frac{N(N-1)}{n(n-1)}$ and each $K^{(i)}$ is isomorphic to K_n. Now let $\mathcal{G}_j := \{G_1^{(j)}, \dots, G_m^{(j)}\}$ be an optimal graph family code for $\mathcal{L}$ on $V(K^{(j)})$ for every $j \in \{1, \dots, r\}$. (Obviously, we can choose each $\mathcal{G}_j$ to be isomorphic to $\mathcal{G}$.) Now define a graph family code on K_N for $\mathcal{L}$ as the collection of graphs that can be written in the form of $G_{\boldsymbol{a}} := \cup_{j=1}^{r} G_{a_j}^{(j)}$, where $\boldsymbol{a} = (a_1, \dots, a_r)$ runs through all possible sequences satisfying $a_i \in \{1, \dots, m\}$ for every i. Since there are m^r such sequences $\boldsymbol{a}$, this way we have m^r different graphs in our family. They indeed form a graph family code for $\mathcal{L}$ since for any two of them, $G_{\boldsymbol{a}}$ and $G_{\boldsymbol{b}}$ there is some j for which $a_j \neq b_j$ and thus $G_{\boldsymbol{a}} \oplus G_{\boldsymbol{b}} \supseteq_{ind} G_{a_j} \oplus G_{b_j} \supseteq_{ind} L$ for some $L \in \mathcal{L}$. This implies $M_{\mathcal{L}}(N) \geq m^r$ and thus

$$R_{\mathcal{L}}(N) \geq \frac{2}{N(N-1)} \log_2 m^r = \frac{2}{n(n-1)} \log_2 M_{\mathcal{L}}(n) = R_{\mathcal{L}}(n).$$

The requirements for N are satisfied if $N = kn(n-1)+1$ and k is large enough. (Also for $N = kn(n-1)+n$ and large enough k but considering the former is enough for our argument.) Since $M_{\mathcal{L}}(n)$ is clearly monotone nondecreasing in n (as we can always ignore some vertices and consider a graph family code only on the rest), we

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

can write that for any $kn(n-1)+1 \le i \le (k+1)n(n-1)$ we have $M_{\mathcal{L}}(i) \ge m^r$ for $r = \frac{\binom{kn(n-1)+1}{2}}{\binom{n}{2}}$. Introducing the sequence $b_i := m^r$ for $r = \frac{\binom{kn(n-1)+1}{2}}{\binom{n}{2}}$ whenever $kn(n-1)+1 \le i \le (k+1)n(n-1)$ we can write

$$\begin{aligned}\liminf_{i\to\infty} \frac{2}{i(i-1)} \log_2 M_{\mathcal{L}}(i) &\ge \liminf_{i\to\infty} \frac{2}{i(i-1)} \log_2 b_i \\ &\ge \liminf_{k\to\infty} \frac{1}{\binom{(k+1)n(n-1)}{2}} \log_2 m^{\frac{\binom{kn(n-1)+1}{2}}{\binom{n}{2}}} \\ &= \liminf_{k\to\infty} \frac{\binom{kn(n-1)+1}{2}}{\binom{(k+1)n(n-1)}{2}} \frac{2}{n(n-1)} \log_2 m = R_{\mathcal{L}}(n).\end{aligned}$$

This proves that $\lim_{n\to\infty} R_{\mathcal{L}}(n)$ exists and is equal to $\sup_n R_{\mathcal{L}}(n)$. ∎

Remark 4.1. The above proof is similar to proving that the limit defining Shannon capacity of graphs exists which is usually done using Fekete's lemma. Here, however, there are some technical subtleties (because of the divisibility requirements for N) that made it simpler to present a full proof than to refer simply to Fekete's lemma. □

In view of Theorem 9 the following definition is meaningful.

DEFINITION 4. *The distance capacity (or distancity for short) of a local graph class $\mathcal{L}$ is defined as*

$$DC(\mathcal{L}) := \lim_{n\to\infty} R_{\mathcal{L}}(n).$$

Based on Turán's celebrated theorem [41] (cf. also, e.g., in [16]) and the famous theorem of Erdős and Stone [22], Erdős and Simonovits [21] proved that if $\mathcal{L}$ is an arbitrary family of graphs, then

$$\lim_{n\to\infty} \frac{ex(n,\mathcal{L})}{\binom{n}{2}} = 1 - \frac{1}{\chi_{\min}(\mathcal{L})-1}, \tag{4.1}$$

where $\chi_{\min}(\mathcal{L}) = \min_{L\in\mathcal{L}} \chi(L)$ and $\chi(G)$ denotes the chromatic number of graph G. (We assume that $\chi_{\min}(\mathcal{L}) \ge 2$. For the case when $\mathcal{L}$ contains some edgeless graph, see Remark 4.10.)

Note that Proposition 8 and the above result determining the order of magnitude of $ex(n,\mathcal{L})$ has the following immediate consequence for the distancity.

COROLLARY 10. *For any local graph class $\mathcal{L}$ with $\chi_{\min}(\mathcal{L}) \ge 2$ we have*

$$DC(\mathcal{L}) \le \frac{1}{\chi_{\min}(\mathcal{L})-1}.$$

4.2. Containing a prescribed subgraph. Now we focus on local graph classes mentioned in our first example after Definition 2: we have some fixed finite simple graph L and consider $\mathcal{L} = \{H : L \subseteq H\}$. As said above in this case we will use the notation $M_L(n)$ for $M_{\mathcal{L}}(n)$ and, similarly, we will also denote $R_{\mathcal{L}}(n)$ and $DC(\mathcal{L})$ by $R_L(n)$ and $DC(L)$, respectively. We prove that in this case the upper bound of Corollary 10 is always sharp.

THEOREM 11. *For any fixed graph L we have*

$$DC(L) = \frac{1}{\chi(L)-1}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

For the proof we will use a result by Erdős, Frankl, and Rödl [18] about the number $F_n(L)$ of graphs on n labeled vertices containing no subgraph isomorphic to L.

Erdős–Frankl–Rödl Theorem (see [18]). *Suppose* $\chi(L) = r \geq 3$. *Then*

$$F_n(L) = 2^{ex(n,K_r)(1+o(1))}.$$

Note that this gives

$$F_n(L) = 2^{\binom{n}{2}\left(1-\frac{1}{\chi(L)-1}+o(1)\right)}$$

by (4.1) (in fact, already directly by Turán's theorem).

While the proof of the Erdős–Frankl–Rödl theorem is based on Szemerédi's Regularity lemma, a similar result for bipartite L easily follows from (4.1) (or from the Kővári–Sós–Turán theorem [30]). Indeed, it implies that if L is bipartite, then $F_n(L) < \binom{\binom{n}{2}}{\varepsilon\binom{n}{2}}$ for any $\varepsilon > 0$ provided $n > n_0(\varepsilon)$, and that implies the claimed statement. (To see the latter, one can use the well-known fact (cf., e.g., Lemma 2.3 in [15]) that

$$\binom{t}{\alpha t} = 2^{t(h(\alpha)+o(1))},$$

where $h(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function and $0 \leq \alpha \leq 1$ is meant to be such that αt is an integer. Applying this for $t := \binom{n}{2}$ and $\alpha = \varepsilon$ we obtain that for any $0 < \varepsilon < 1$ the number $\binom{\binom{n}{2}}{\varepsilon\binom{n}{2}}$ is more than $2^{\delta\binom{n}{2}}$ for some positive δ.)

Proof of Theorem 11. It follows immediately from Corollary 10 that the right-hand side is an upper bound on the left-hand side so we only have to prove the reverse inequality.

To this end let G_L denote the graph whose vertices are all possible graphs on n labeled vertices and two are connected if and only if their symmetric difference does not contain L as a subgraph. (Note that this is just the complementary graph of $H_{\mathcal{F}}$ used in the proof of Lemma 2.1 when $\mathcal{F}$ is set to be the local graph class $\mathcal{L}$ belonging to our problem.) Then $M_L(n)$ is equal to the independence number $\alpha(G_L)$ of G_L. Clearly, G_L is vertex-transitive (cf. the argument in the proof of Lemma 2.1 for $H_{\mathcal{F}}$), in particular, it is regular. Since the degree of its vertex representing the edgeless graph is just $F_n(L)$, we get (denoting the maximum degree of a graph G by $\Delta(G)$) that

$$\begin{aligned} M_L(n) = \alpha(G_L) &\geq \frac{|V(G_L)|}{\Delta(G_L)+1} = \frac{|V(G_L)|}{F_n(L)+1} \\ &= \frac{2^{\binom{n}{2}}}{2^{\binom{n}{2}\left(1-\frac{1}{\chi(L)-1}+o(1)\right)}} = 2^{\binom{n}{2}\left(\frac{1}{\chi(L)-1}+o(1)\right)} \end{aligned}$$

by the Erdős–Frankl–Rödl theorem (and by the above discussion also for bipartite graphs). Putting this inequality into the definition of $DC(L)$ the required result follows. ∎

Corollary 12. *Let* $\mathcal{G}$ *be a set of graphs, each containing at least one edge, and let* $\mathcal{L}_{\mathcal{G}}$ *be the local graph class containing all graphs that contain at least one* $G \in \mathcal{G}$ *as a subgraph. Then*

$$DC(\mathcal{L}_{\mathcal{G}}) = \frac{1}{\chi_{\min}(\mathcal{L}_{\mathcal{G}})-1} = \frac{1}{\chi_{\min}(\mathcal{G})-1}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

In particular,

$$DC(\mathcal{C}_{\text{odd}}) = DC(K_3) = \frac{1}{2}.$$

Proof. The second statement is clearly a special case of the first one, so it is enough to prove the latter. It is a straightforward consequence of Corollary 10 that the left-hand side is bounded from above by the right-hand side. For the reverse inequality note the trivial fact that $DC(\mathcal{L}_{\mathcal{G}}) \geq DC(G)$ for any $G \in \mathcal{G}$. Applying this for some $G \in \mathcal{G}$ that satisfies $\chi(G) = \min_{G \in \mathcal{G}} \chi(G) = \chi_{\min}(\mathcal{L}_{\mathcal{G}})$ the statement follows from Theorem 11. ∎

Remark 4.2. It is straightforward from the foregoing that the above results also determine for any graph family $\mathcal{G}$ the asymptotic behavior of the value $D_{\mathcal{L}_{\mathcal{G}}}(n)$ belonging to the dual problem. Indeed, by Lemma 2.1 and Corollary 12 we have that $\lim_{n\to\infty} \frac{1}{\binom{n}{2}} \log D_{\mathcal{L}_{\mathcal{G}}}(n) \leq 1 - DC(\mathcal{L}_{\mathcal{G}}) = 1 - \frac{1}{\chi_{\min}(\mathcal{G})-1}$ while a matching lower bound follows from the argument in the proof of Proposition 8. Thus we have

$$\lim_{n\to\infty} \frac{2}{n(n-1)} \log D_{\mathcal{L}_{\mathcal{G}}}(n) = 1 - \frac{1}{\chi_{\min}(\mathcal{G}) - 1}$$

for any graph family $\mathcal{G}$. This means that by taking all subgraphs of a graph with the largest possible number of edges without containing a subgraph from $\mathcal{G}$ we obtain asymptotically a largest family of graphs, no two of which have any $G \in \mathcal{G}$ in their symmetric difference. □

4.3. Containing a triangle or an odd cycle. In this subsection we are investigating $M_L(n)$ for small values of n and the simplest 3-chromatic graph, which is the triangle K_3. We will also look at the analogous problem when K_3, the cycle of length 3 is replaced by the family of all odd cycles.

For $L = K_3$ the bound of Proposition 8 gives us $M_{K_3}(n) \leq 2^{\binom{n}{2} - \lceil \frac{n}{2} \rceil \lfloor \frac{n}{2} \rfloor}$. Below we show that this upper bound is tight whenever n is at most 6.

The first part of the following proposition is very simple and we present it only for the sake of completeness.

PROPOSITION 13. *We have* $M_{K_3}(3) = 2$ *and* $M_{K_3}(4) = 4$.

Proof. For $n = 3$ the statement is trivial: take the empty graph and a triangle on three vertices, this 2-element family already achieves the value of the upper bound which is 2 for $n = 3$.

For $n = 4$ we give the following four graphs on the vertex set $\{1, 2, 3, 4\}$ by their edge sets. Let

$$E(G_0) = \emptyset, \quad E(G_1) = \{12, 23, 13, 34\},$$
$$E(G_2) = \{23, 34, 24, 14\}, \quad E(G_3) = \{12, 13, 24, 14\}.$$

It takes just an easy checking to see that the symmetric difference of any two of these graphs contains a triangle. Since the upper bound in Proposition 8 is also 4 in this case, this proves that $M_{K_3}(4) = 4$. ∎

Remark 4.3. Note that both of the above simple constructions are closed under the symmetric difference operation, that is they form a linear space over $GF(2)$ when the graphs are represented by the characteristic vectors of their edge sets. In fact, the second construction could also be presented as the vector space generated in this sense by any two of the graphs G_1, G_2, G_3. □

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

PROPOSITION 14.

$$M_{K_3}(5) = 16.$$

Proof. The value of the upper bound in Proposition 8 gives 16 for $n = 5$, so we only have to prove that 16 is also a lower bound. To this end we will give a set of graphs forming a vector space in the sense of Remark 4.3. We will give this vector space by a set of generators, although in a somewhat redundant way. (Our reason to keep this redundancy is that the construction has more symmetry this way.)

Think about the vertices $\{1,2,3,4,5\}$ as if they were given on a circle at the vertices of a regular pentagon in their natural order. Consider the graph with edge set

$$E(G_1) := \{12, 23, 13, 35\}.$$

Let G_2, G_3, G_4, G_5 be the four graphs we obtain from G_1 by rotating it along the circle containing the vertices so that vertex 1 moves to 2, 2 to 3, etc. Thus we have

$$E(G_2) = \{23, 34, 24, 41\}, \quad E(G_3) = \{34, 45, 35, 52),$$
$$E(G_4) = \{45, 51, 41, 13\}, \quad E(G_5) = \{51, 12, 52, 24\}.$$

Now we consider the linear space the characteristic vectors of the edge sets of these five graphs $G_i, i \in \{1,2,3,4,5\}$ generate. These graphs can be defined as the elements of the family $\mathcal{G} = \{G_I : I \subseteq [5]\}$, where

$$G_I = \oplus_{i \in I} G_i,$$

meaning that $V(G_I) = [5]$ and $E(G_I)$ contains exactly those edges that appear in an odd number of the graphs G_i with $i \in I$.

Note that every edge of the underlying K_5 on $[5]$ appears in exactly two of the graphs $G_1, \ldots, G_5$; therefore, for $I = [5]$ we have that G_I is the empty graph just as $G_\emptyset$ is. This implies that for every $I \subseteq [5]$ and $\overline{I} := [5] \setminus I$ we have $G_I = G_{\overline{I}}$, thus every graph in our graph family has exactly two representations as G_I for some $I \subseteq [5]$. (The two representations are given by I and $\overline{I}$ as we have seen. It also follows that if $J \neq I, \overline{I}$, then $G_J \neq G_I$, otherwise we would have $G_{J \oplus I}$ be the empty graph for $J \oplus I \notin \{\emptyset, [5]\}$ contradicting that every edge appears exactly twice in the sets $E(G_i),\ i = 1, \ldots, 5.$) Thus we have indeed $\frac{1}{2}2^5 = 16$ graphs in our family matching our upper bound for $n = 4$.

We have to show that the symmetric difference of any two of our graphs contains a triangle. Since our construction is closed for the symmetric difference operation, this is equivalent to say that all graphs in our family except the empty graph contains a triangle. Since $G_I = G_{\overline{I}}$ it is enough to prove that G_I contains a triangle for all $1 \leq |I| \leq 2, I \subseteq [5]$. This is easy to see when $|I| = 1$. For subsets with $|I| = 2$ it is enough to check this for $I = \{1,2\}$ and $I = \{1,3\}$ by the rotational symmetry of our construction. But these two cases are easy to check: $G_{\{1,2\}}$ contains the triangles on the triples of vertices $1,2,4$ and $1,3,4$, while $G_{\{1,3\}}$ contains the triangle on vertices $1,2,3$. ∎

PROPOSITION 15.

$$M_{K_3}(6) = 64.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Proof. The value of the upper bound given by Proposition 8 is 2^6 for $L = K_3$ and $n = 6$, so we need only prove the lower bound.

To this end we give a construction of 64 graphs forming a graph family code on [6] for K_3. The construction will have several similarities to that in Proposition 14 though with somewhat less symmetry. But again our graphs will form a vector space in the sense of Remark 4.3 to be specified through a set of seven generators that altogether cover each one of the edges of the underlying K_6 exactly twice, so every member of our graph family will have exactly two representations by the generators just as in the proof of Proposition 14. Here are the details.

Think about the six vertices $1, \dots, 6$ as being on a circle in the vertices of a regular hexagon in their natural order as we go around the circle. Our first four generator graphs are the following four edge-disjoint triangles (plus three isolated points) given by their edge sets as follows.

$$E(G_1) = \{12, 23, 13\}, \quad E(G_2) = \{34, 45, 35\},$$
$$E(G_3) = \{56, 16, 15\}, \quad E(G_4) = \{24, 46, 26\}.$$

The other three graphs are three K_4's (plus two isolated vertices) that are rotations of each other, in particular,

$$E(G_5) = \{12, 24, 45, 15, 14, 25\}, \quad E(G_6) = \{23, 35, 56, 26, 25, 36\},$$
$$E(G_7) = \{34, 46, 16, 13, 36, 14\}.$$

It is easy to check that the above seven graphs cover each edge of the underlying K_6 exactly twice. Just as in the proof of Proposition 14 this implies that the generated family of graphs of the form

$$G_I = \oplus_{i \in I} G_i,$$

where I runs through all subsets of [7], contains exactly two representations of this form for each of its members, namely

$$G_I = G_J \text{ if and only if } J = [7] \setminus I.$$

Thus our family has $2^6 = 64$ members that matches our upper bound. Now we have to show that the symmetric difference of every pair of our graphs contains a triangle. Since the family is closed under symmetric difference this is equivalent to every G_I except $G_\emptyset = G_{[7]}$ containing a triangle. To show this we consider the representation of each of our graphs as G_I where I contains at most one of the three K_4 generators, that is $|I \cap \{5, 6, 7\}| \leq 1$. When $I \cap \{5, 6, 7\} = \emptyset$ but I itself is nonempty, then this is trivial as in such a case G_I is the union of some of the edge-disjoint graphs $G_1, \dots, G_4$ each of which is a triangle itself. In case $|I \cap \{5, 6, 7\}| = 1$, then by symmetry we may assume w.l.o.g. that $I \cap \{5, 6, 7\} = \{5\}$. Then if we also have $\{1, 2\} \subseteq I$, then the triangles on vertices $1, 3, 4$ and $2, 3, 5$ (and more) will be contained in G_I. So we may assume that at least one of G_1 and G_2 is not part of our representation of G_I and by symmetry, we may assume $2 \notin I$. But then to avoid the triangles on vertices $1, 4, 5$ and $2, 4, 5$ being in G_I we need both $3 \in I$ and $4 \in I$. In this case, however, we will have the triangle on vertices $4, 5, 6$ present in G_I. This completes the proof. ∎

Recall $\mathcal{C}_{\text{odd}}$ to be the class of all graphs containing an odd cycle. Since $ex(n, \mathcal{C}_{\text{odd}}) = ex(n, K_3)$ the upper bound of Proposition 8 is also $2^{\binom{n}{2} - \lceil \frac{n}{2} \rceil \lfloor \frac{n}{2} \rfloor}$ for $M_{\mathcal{C}_{\text{odd}}}(n)$. Since

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

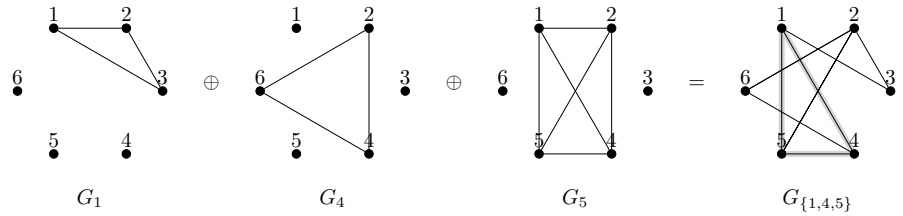

FIG. 1. *Graphs G_1, G_4, G_5, and their generated graph $G_{\{1,4,5\}}$ in the proof of Proposition* 15.

$K_3 \cong C_3$ is an odd cycle, we obviously have $M_{K_3}(n) \leq M_{\mathcal{C}_{\text{odd}}}(n)$ and so by Propositions 13, 14, and 15 the previous upper bound is also sharp for $M_{\mathcal{C}_{\text{odd}}}(n)$ when $n \in \{3,4,5,6\}$. Although we could not prove that $M_{K_3}(7)$ is also equal to this upper bound, we can show this at least for $M_{\mathcal{C}_{\text{odd}}}(7)$.

PROPOSITION 16.

$$M_{\mathcal{C}_{\text{odd}}}(7) = 2^9.$$

Proof. The upper bound $2^{\binom{n}{2}-\lceil \frac{n}{2} \rceil \lfloor \frac{n}{2} \rfloor}$ is equal to 2^9, so it is enough to prove that this is also a lower bound. This we do similarly as in the proofs of Propositions 14 and 15.

Again, we think about the seven vertices forming the set $[7]$ as the vertices of a regular 7-gon around a cycle in their natural order. We define $7+3=10$ simple graphs $G_1, \ldots, G_7$ and $G_8, \ldots, G_{10}$ that will generate our family. Let G_1 be the triangle with edges $12, 24, 14$ and $G_2, \ldots, G_7$ be its six possible rotated versions, that is the triangles with edge sets $\{23,35,25\}, \{34,46,36\}, \ldots, \{17,13,37\}$, respectively. Note that these seven triangles cover all pairs of vertices exactly once, that is, they form a Steiner triple system. The three other graphs G_8, G_9, G_{10} are three edge-disjoint seven-cycles, namely those with edge sets

$$\{12,23,34,45,56,67,17\}, \{13,35,57,27,24,46,16\}, \{14,47,37,36,26,25,15\},$$

respectively. Note that these three graphs also cover all pairs of vertices exactly once and that the edge sets of a G_i for $i \in [7]$ and G_j with $j \in \{8,9,10\}$ intersect in exactly one element. Since our ten graphs cover the edges of the underlying K_7 exactly twice, just as in the proofs of Propositions 14 and 15, the generated family

$$\oplus_{i \in I} G_i$$

as I runs over all subsets of $\{1, \ldots, 10\}$ will have exactly 2^9 distinct members each of which is represented by two subsets of $\{1, \ldots, 10\}$, some I, and its complement. All we are left to show for proving $M_{\mathcal{C}_{odd}}(7) \geq 2^9$ is that each such G_I except $G_\emptyset = G_{[10]}$ contains an odd cycle. If $I \subseteq [7]$, this is obvious and also is if $I \subseteq \{8,9,10\}$. When both $I \cap [7]$ and $I \cap \{8,9,10\}$ are nonempty, then we consider that representation G_I which has $|I \cap [7]| \leq 3$. If we have $|I \cap \{8,9,10\}| = 1$, then whichever 7-cycle we have (that is, whichever of G_8, G_9, G_{10}) it will have two consecutive edges that do not appear in either of the at-most three triangles. If we take the first pair of such edges (as we go along our 7-cycle in an appropriate direction) for which the previous one is an edge of one of our triangles (since we take at least one triangle and each triangle intersects each 7-cycle, such an edge must exist), then the construction ensures that these two consecutive edges close up to a K_3 in our G_I. In case we have two 7-cycles in our G_I

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

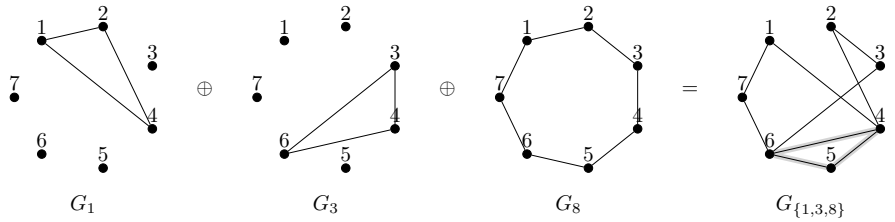

FIG. 2. *Graphs G_1, G_3, G_8, and their generated graph $G_{\{1,3,8\}}$ in the proof of Proposition* 16.

representation, then those create 7 distinct K_3's in their union. Each of our triangles intersect exactly three of those seven K_3's created, so if we have $|I \cap [7]| \leq 2$, then at least one of these seven K_3's remain untouched. Thus we are left with the case of two 7-cycles and exactly three triangles. For this case let us switch to the complementary representation with four triangles and one 7-cycle. By symmetry, we may assume that our 7-cycle is G_8. If the four triangles are such that two consecutive edges of G_8 do not appear in any of them, then we can finish the argument as before. If this is not the case, then the four triangles must leave three such edges of G_8 uncovered which form a matching. Because of symmetry we may assume that these are the edges $12, 34, 56$. This also tells us exactly which are the four triangles we have in the representation of G_I, namely those that contain the remaining four edges, that is, G_2, G_4, G_6, and G_7. In this case G_I contains the K_3, for example, on the vertices $2, 5, 6$. Finally, if we have all of the three 7-cycles in our representation, then the complementary representation has no 7-cycle at all and this case we have already covered. This completes the proof. ∎

4.4. Containing a prescribed induced subgraph. In this subsection we discuss local graph classes mentioned in the second example after Definition 2. Here we have a fixed finite simple graph L and consider the family $\mathcal{L}$ of all graphs containing L as an induced subgraph. Recall that in this case we let $M_{L,\text{ind}}(n)$ denote $M_{\mathcal{L}}(n)$, and similarly, we denote $R_{\mathcal{L}}(n)$ and $DC(\mathcal{L})$ by $R_{L,\text{ind}}(n)$ and $DC(L, \text{ind})$, respectively. In this section we prove that requiring the subgraphs to be induced does not change the answer from that of section 4.2. The upper bound, of course, trivially carries over from the noninduced case, while the lower bound strengthens the one in Theorem 11.

THEOREM 17. *For any fixed graph L we have*

$$DC(L, \text{ind}) = \frac{1}{\chi(L) - 1}.$$

Remark 4.4. Note that despite the apparent similarity, the proof of Theorem 11 does not carry over to show Theorem 17. Still, it is possible to describe the asymptotic number of induced L-free graphs for a fixed graph L by introducing the partition number $r(L)$. Define $r(L)$ as the largest integer r so that there is some integer s, $0 \leq s \leq r$ such that the vertices of L cannot be covered by s cliques and $r - s$ independent sets. Results obtained independently by Alekseev [2] and by Bollobás and Thomason [11, 12] imply that the number of induced-L-free graphs on n vertices is $2^{(1-1/r(L))n^2/2+o(n^2)}$. Note that $r(L) \geq \chi(L) - 1$, as the vertices of L cannot be covered by $\chi(L) - 1$ independent sets. There are cases when equality holds and thus the required result follows (an example for that is $L = C_5$, the 5-length cycle), but in general, $r(L)$ can be much larger than $\chi(L) - 1$, as shown, for example, by any

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

long (even or odd) cycle. Hence the proof of Theorem 11 does not give the required bound for an arbitrary L. ∎

The main tool in the proof of Theorem 17 is Lemma 4.5, which bounds the number of balanced k-partite induced-L-free graphs.

LEMMA 4.5. *For a fixed positive integer k and a fixed graph L, consider the set of balanced k-partite graphs G on k-classes $A_1, \ldots, A_k$, each having size n/k. Among those graphs, at most $2^{(1-\frac{1}{\chi(L)-1})n^2/2+o(n^2)}$ do not contain an induced subgraph isomorphic to L.*

In what follows we will assume some familiarity with Szemerédi's regularity lemma and the terminology related to it. A good introduction to these notions (with a full proof of the lemma itself) can be found, for example, in section 7.4 of Diestel's book [16].

Before presenting the proof of Lemma 4.5, whose proof relies on the regularity lemma, we need to establish the following auxiliary claim about finding induced subgraphs using regularity. Versions of this result have been used before, for completeness we include a simple proof.

LEMMA 4.6. *For every $0 < \delta < \frac{1}{2}$ and a graph L, there exist positive constants $\varepsilon = \varepsilon(\delta, L)$ and $n_0 = n_0(\delta, L)$ with the following property. Suppose G is a graph whose vertices are partitioned into $\chi(L)$ independent sets, $V_1, V_2, \ldots, V_{\chi(L)}$ of equal size which is at least n_0. If the pairs (V_i, V_j) are ε-regular and if their density satisfies $d(V_i, V_j) \in (\delta, 1-\delta)$ for all i, j, then G contains an induced copy of L.*

Proof. Let l be the number of vertices of L, and let $V(L) = U_1 \cup \cdots \cup U_{\chi(L)}$ be a partition of L into $\chi(L)$ independent sets. The idea is to find a copy of L in G such that the vertices of U_i come from V_i, for all i. Having this goal in mind, refine the partition $V_1, \ldots, V_{\chi(L)}$ by partitioning every V_i into $|U_i|$ smaller sets, $V_{i,1}, \ldots, V_{i,|U_i|}$, of nearly equal sizes. Subdividing V_i does not affect regularity of the new pairs significantly, and so the pairs $(V_{i,j}, V_{k,m})$ are still $l\varepsilon$-regular. Similarly, the density of the pairs $(V_{i,j}, V_{k,m})$ is in the interval $(\delta - \varepsilon, 1 - \delta + \varepsilon)$ if $i \neq k$, and zero otherwise. Now, we use the following standard lemma (see, e.g., Lemma 3.2 in [4]).

Lemma on regularity and induced subgraphs (see [4]). *For every $0 < \delta_0 < 1$ and $l \in \mathbb{Z}_{>0}$, there exist positive constants $\varepsilon_0 = \varepsilon_0(\delta_0, l)$ and $\mu_0 = \mu_0(\delta_0, l)$ with the following property. Suppose L is a graph on l vertices, $v_1, \ldots, v_l$ and $S_1, \ldots, S_l$ is an l-tuple of disjoint vertex sets of a large graph G such that every pair $S_i S_j$ is ε_0-regular with density at least δ_0 if $v_i v_j$ is an edge of L and at most $1 - \delta_0$ if $v_i v_j$ is not an edge of L. Then, G contains at least $\mu_0 \prod_{i=1}^{l} |S_i|$ tuples $(w_1, \ldots, w_l) \in S_1 \times \ldots \times S_l$ spanning an induced copy of L, where each w_i corresponds to v_i.*

Note that identifying the sets $V_{i,j}$ with S_i from this lemma satisfies the conditions, as the pairs $(V_{i,j}, V_{i,k})$ of density 0 occur only when the corresponding vertices of L belong to the same independent set U_i. Therefore, if we set $\delta_0 = \delta - \varepsilon, \varepsilon_0 = \varepsilon l$ and choose ε small enough, we conclude that for large enough n_0, G contains at least one induced copy of L, as needed. ∎

Proof of Lemma 4.5. The proof uses Szemerédi's regularity lemma to partition an arbitrary induced-L-free graph, after which a standard "cleaning" argument shows that the main contribution to the total number of such graphs comes from the number of possible bipartite graphs induced by regular pairs of density bounded away from 0 and 1. Then, Turán's theorem provides an upper bound for the number of such pairs, completing the proof. The details follow.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Begin by fixing a small $\delta > 0$ and the corresponding $\varepsilon = \varepsilon(\delta, L) > 0$ from Lemma 4.6. What we have in mind is a not-yet-fixed induced-L-free graph G whose vertex set will be partitioned and we will calculate the number of ways we can connect pairs of vertices so that our graph becomes indeed induced-L-free. By Szemerédi's regularity lemma, there is an ε-regular partition $V(G) = V_0 \cup V_1 \cup \ldots \cup V_T$, where T is a constant, that is, it can be bounded from above by a number $T_0(\varepsilon, k)$ which does not depend on n. We may also assume that this partition refines the original partition $V(G) = A_1 \cup \cdots \cup A_k$, i.e., that every V_i for $i \geq 1$ is a subset of some A_j. (This follows from the standard proof of the regularity lemma (cf., e.g., [16]) that works by iterating the refinement of some original partition which we can choose to be the one given by the partite classes A_i.)

We split the pairs (V_i, V_j) into three classes: the irregular pairs, the ε-regular pairs with $d(V_i, V_j) \in [0, \delta] \cup [1-\delta, 1]$, and the ε-regular pairs with $d(V_i, V_j) \in (\delta, 1-\delta)$. We first show that the contribution of all pairs except those in the third class to the total number of graphs is negligible, and then we discuss the number of pairs in the third class.

There are at most $(T+1)^n = 2^{o(n^2)}$ ways to distribute the vertices into parts $V_0, \ldots, V_T$, and there are at most $2^{\varepsilon n^2}$ ways to choose the edges incident to V_0, as $|V_0| \leq \varepsilon n$. Note that there are no edges within the other parts V_i, so we have no additional choice here. The number of ways to associate pairs to classes is at most $3^{\binom{T}{2}}$, which is a constant (depending on ε).

Further, there are at most $2^{(\frac{n}{T})^2}$ ways to choose the edges between each of the irregular pairs, and there are at most εT^2 irregular pairs. Hence, the number of choices for the edges between all irregular pairs is at most $2^{\varepsilon n^2}$. Similarly, for the parts of density close to 0 or 1, we have at most $2\sum_{i=0}^{\delta(n/T)^2} \binom{(n/T)^2}{i} \leq \frac{\delta n^2}{T^2} \binom{(n/T)^2}{\delta(n/T)^2} \leq \frac{\delta n^2}{T^2} (e\delta^{-1})^{\delta(n/T)^2}$, and there are at most $\binom{T}{2}$ such pairs. Hence, the total number of choices for the edges in these pairs is at most $e^{\delta(1-\log\delta)\binom{n}{2}+o(n^2)}$.

Finally, we need to bound the contribution from the pairs (V_i, V_j) with $d(V_i, V_j) \in (\delta, 1-\delta)$. Define an auxiliary graph G' on the vertex set $\{V_1, \ldots, V_T\}$, in which V_iV_j is an edge if and only if (V_i, V_j) forms an ε-regular pair of density in $(\delta, 1-\delta)$. Lemma 4.6 shows that if G contains no induced copy of L, then G' cannot contain a subgraph isomorphic to $K_{\chi(L)}$. Applying Turán's theorem, we conclude that at most $(1 - \frac{1}{\chi(L)-1})T^2/2$ pairs among $V_1, \ldots, V_T$ can be ε-regular and have density bounded away from 0 and 1. Hence, there are at most $2^{(1-\frac{1}{\chi(L)-1})(T^2/2)\frac{n^2}{T^2}} \leq 2^{(1-\frac{1}{\chi(L)-1})n^2/2+o(n^2)}$ choices for the edges in this case.

To complete the argument, we let $\delta, \varepsilon \to 0$ and note that the contribution of all pairs except those in the third class are negligible. We conclude that the number of k-partite balanced induced-L-free graphs on n vertices is at most $2^{(1-\frac{1}{\chi(L)-1})n^2/2+o(n^2)}$, as stated. ∎

Using Lemma 4.5, the proof of Theorem 17 follows almost immediately.

Proof of Theorem 17. Analogously to what was done in the proof of Theorem 11, now consider the graph $G_{L,k}$ whose vertices are all balanced k-partite graphs on n vertices, in which two vertices are adjacent if and only if their symmetric difference contains no induced copy of L. As shown by Lemma 4.5, the maximum degree in this graph is at most $\Delta(G_{L,k}) \leq 2^{(1-\frac{1}{\chi(L)-1})n^2/2+o(n^2)}$. On the other hand, $G_{L,k}$ has $2^{\binom{k}{2}(\frac{n}{k})^2} = 2^{(1-\frac{1}{k})n^2/2}$ vertices. Hence, the size of its maximum independent set is at least

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$$\alpha(G_{L,k}) \geq \frac{|V(G_{L,k})|}{\Delta(G_{L,k})+1} \geq 2^{\left(\frac{1}{\chi(L)-1}-\frac{1}{k}\right)n^2/2+o(n^2)}.$$

As k can be arbitrarily large, we conclude that $DC(L, \text{ind}) = \frac{1}{\chi(L)-1}$, completing the proof. ∎

As a short digression from the main line of our discussion we mention that in the case of bipartite graphs, one can prove a result analogous to Lemma 4.5 with much better bounds, as shown in [3], and [5] with an improved error term.

LEMMA 4.7. *For every fixed bipartite graph L there is some $\varepsilon = \varepsilon(L) > 0$ so that the number of bipartite graphs on two classes of vertices A and B, both of size m, that do not contain an induced copy of L is smaller than $2^{m^{2-\varepsilon}}$.*

The statement of this lemma, with a nonoptimal value of ε, follows from the results in the above mentioned papers that estimate the number of bipartite graphs that do not contain an induced copy of the universal bipartite graph $U(k)$ with k vertices in one vertex class and 2^k in the other, connected in all possible ways to the vertices of the first class. Every bipartite graph L is an induced copy of $U(k)$ for all sufficiently large k. Although this is good enough for our purpose here, we present next a shorter new proof of the lemma, which gives a tight estimate up to a logarithmic factor in the exponent for many graphs L.

LEMMA 4.8. *Let L be a bipartite graph L with color classes of sizes s and $t \geq s$. Then the number of bipartite graphs on two classes of vertices A and B, each of size m, that do not contain an induced copy of L, is at most $2^{c(s,t)m^{2-1/s}\log m}$.*

The above is tight, up to the logarithmic term in the exponent, for every pair $t \geq s$ where t is sufficiently large as a function of s. Indeed, as shown by the (projective) norm-graphs of [25, 6], there is a bipartite graph with classes of vertices of size m each and with $\Omega(m^{2-1/s})$ edges that contains no copy of the complete bipartite graph $K = K_{s,t}$, provided $t > (s-1)!$ Every member of the collection of all subgraphs of this graph contains no (induced or noninduced) copy of K.

Proof. We apply the Sauer–Perles–Shelah lemma (see [35], [37]) which states that for any collection $\mathcal{C}$ of more than $\sum_{i=0}^{d}\binom{q}{i}$ functions from a set Q of size q to $\{0,1\}$ there is a subset $D \subset Q$ of cardinality $d+1$ *shattered* by $\mathcal{C}$. That is, for any function $g\colon D \mapsto \{0,1\}$ there is an $f \in \mathcal{C}$ such that $g(x) = f(x)$ for all $x \in D$. Let L be a bipartite graph with classes of vertices of sizes $t \geq s$, and let $\mathcal{G}$ be a collection of graphs on the two vertex classes A and B, where $|A| = |B| = m$. Let $d = ex(2m, K_{s,t})$ be the maximum possible number of edges in a graph on $2m$ vertices that contains no copy of the complete bipartite graph $K_{s,t}$. By the Kővári–Sós–Turán theorem [30], $d \leq b(s,t)m^{2-1/s}$. By the Sauer–Perles–Shelah lemma, if

$$|\mathcal{G}| \geq 1 + \sum_{i=0}^{d}\binom{m^2}{i},$$

then the collection of graphs $\mathcal{G}$, viewed as a collection of functions from the set of all edges of the complete bipartite graph on the classes of vertices A, B to $\{0,1\}$, shatters a set of $d+1$ edges. By the definition of d this set of edges contains a complete bipartite graph with vertex classes of sizes s and t. Any subgraph of this graph (and in particular L) is an induced subgraph of some member of $\mathcal{G}$. Since the right-hand side of the last inequality is smaller than

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$$m^{2d} \le m^{2b(s,t)m^{2-1/s}} = 2^{c(s,t)m^{2-1/s}\log m},$$

this completes the proof. ∎

The following is a counterpart of Corollary 12 for the induced case.

COROLLARY 18. *Let $\mathcal{G}$ be a set of graphs, each containing at least one edge, and let $\mathcal{L}_{(\mathcal{G},\mathrm{ind})}$ be the local graph class containing all graphs that contain at least one $G \in \mathcal{G}$ as an induced subgraph. Then*

$$DC(\mathcal{L}_{(\mathcal{G},\mathrm{ind})}) = \frac{1}{\chi_{\min}(\mathcal{G}) - 1}.$$

Proof. The upper bound is immediate from $DC(\mathcal{L}_{(\mathcal{G},\mathrm{ind})}) \le DC(\mathcal{L}_{\mathcal{G}})$ and Corollary 12. The lower bound follows, as in Corollary 12, from the fact that $DC(\mathcal{L}_{(\mathcal{G},\mathrm{ind})}) \ge DC(G,\mathrm{ind})$ for any $G \in \mathcal{G}$. In particular, picking a graph G with $\chi(G) = \chi_{\min}(\mathcal{G})$ suffices. ∎

Remark 4.9. It is straightforward that the above results imply a strengthening (as far as the upper bound is concerned) of the statement in Remark 4.2, namely that we have for the dual problem also in the induced case

$$\lim_{n\to\infty} \frac{2}{n(n-1)} \log D_{\mathcal{L}_{(\mathcal{G},\mathrm{ind})}}(n) = 1 - \frac{1}{\chi_{\min}(\mathcal{G}) - 1}.$$

This follows from Lemma 2.1 and Corollary 18 just as the statement in Remark 4.2 followed from Lemma 2.1 and Corollary 12. □

Remark 4.10. If L is the edgeless graph on $r \ge 2$ vertices, then it is clear that for every $n \ge r$ there is a family of $2^{\binom{n}{2}-\binom{r}{2}}$ graphs on $[n]$ so that the symmetric difference between any two contains an induced copy of L. Indeed we simply take all graphs which agree on the $\binom{r}{2}$ edges of a fixed r-clique. Note that for every fixed r this is a constant fraction of all graphs on n vertices, much larger than $M_L(n)$ or $M_{L,\mathrm{ind}}(n)$ for any graph L with at least two edges. This (including the constant) is clearly tight for $r = 2$ (the family cannot contain a graph and its complement), and by the main result of [17] it is tight also for $r = 3$ (the result in [17] holds for any family in which any two members agree on a triangle, not only if any two intersect in a common triangle–-the equivalence of these two statements is proved already in [13].) More recently, Berger and Zhao [9] proved the analogous result also for $r = 4$. We do not know if the above lower bound is tight for $r > 4$. Note that an equivalent formulation of the question here is the determination of $M_{\mathcal{F}}(n)$ for $\mathcal{F}$ which is the family of all graphs with independence number at least r. ∎

5. Open problems. In this final section we collect some related problems left open.

PROBLEM 1. *For what graph families $\mathcal{F}$ is it true that $M_{\mathcal{F}}(n)$ is achieved by a linear graph family code, that is one that is closed under the symmetric difference operation?*

Our results here include examples where this is the case as well as ones in which it is not. Indeed in Theorem 6 the precise answer is n or $n+1$, and if this is not a power of 2 there is no optimal linear solution. Another family of examples in which the optimal family cannot be achieved by a linear example is that in which the family $\mathcal{F}$ is the family of all graphs with at most $2r$ edges, where r is chosen so that the sum

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$$\sum_{i=0}^{r}\binom{\binom{n}{2}}{i}$$

is not a power of 2. Indeed, by a theorem of Kleitman [23] (for usual codes) the size of the optimum family here is the size of the family of all graphs with at most r edges.

The construction in the proof of Theorem 1 has the property that for any two of its graphs G and G' with an equal number of edges their two *asymmetric* differences

$$G \setminus G' = ([n], E(G) \setminus E(G')) \quad \text{and} \quad G' \setminus G = ([n], E(G') \setminus E(G))$$

are isomorphic. This suggests the following question.

PROBLEM 2. *What is the maximum possible size of a graph family $\mathcal{A}$ of graphs on n vertices satisfying that if $A, A' \in \mathcal{A}$, then $A \setminus A'$ and $A' \setminus A$ are isomorphic?*

Clearly, the conditions of the above problem imply that all graphs in a family satisfying the above criteria must have the same number of edges. An appropriate subfamily of the one constructed in the proof of Theorem 1 satisfies the requirements but a significantly larger family can be given by taking $\lfloor n/k \rfloor$ vertex-disjoint stars, each on k vertices. Such a family satisfies the requirements because any star of one graph in it can have a common edge with at most one star of another graph of the family and two stars on the same number of vertices obviously satisfy the criteria as they cancel the same number of edges from each other resulting in two smaller (possibly edgeless) stars that are also isomorphic. This gives a lower bound that is superexponential in n but we do not have any nontrivial upper bound.

Theorems 4 and 6 show a huge difference between requiring a spanning path or a spanning star in the symmetric differences. One may wonder what happens "in between." Note that if we formulate this "in betweenness" so that we want to have a spanning tree with diameter at most k, then while with $k = 2$ we are at Theorem 6 and with $k = n - 1$ at Theorem 4, already for $k = 3$ we get the same result as for $k = n - 1$ by the construction in the proof of Theorem 1. (This is simply because complete bipartite graphs contain spanning trees of diameter at most 3.) So it seems plausible to formulate questions in terms of more specific "natural" sequences of spanning trees $T_1, T_2, \ldots$. (In the problem below the notation $M_{T_n}(n)$ is meant to denote the largest possible cardinality of a family of graphs on vertex set $[n]$ such that the symmetric difference of any two of them contains T_n as a subgraph.)

PROBLEM 3. *For what "natural" sequences $T_1, T_2, \ldots, T_i, \ldots$ of trees (with T_i having exactly i vertices for every i) will the value of $M_{T_n}(n)$ grow only linearly in n? A similar question is valid if T_i is replaced by $\mathcal{T}_i$, some "natural" family of i-vertex trees.*

Propositions 13, 14, 15, and 16 showed that the upper bound of Proposition 8 can be sharp for small values of n for the requirement that a triangle or at least an odd cycle is contained in the symmetric differences. It would be interesting to know whether this can also happen for large values of n.

PROBLEM 4. *Is*

$$M_{K_3}(n) = 2^{\binom{n}{2} - \lceil \frac{n}{2} \rceil \lfloor \frac{n}{2} \rfloor}$$

true always or at least for infinitely many values of n? Even if this is not so, does the analogous equality hold for $M_{\mathcal{C}_{\text{odd}}}(n)$?

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Note that there are much better known estimates for the number of triangle-free graphs on n labeled vertices than the one we have used here, in fact, it is known that almost all of these graphs are bipartite [19]. While this improves the gap between the upper and lower bounds that follow from our proofs for $M_{K_3}(n)$, it is still far from determining its precise value.

The final problem we mention is related to the remark in the end of the last subsection.

PROBLEM 5. *Is it true that for any fixed $r > 4$ the maximum possible cardinality of a family of graphs on n labeled vertices in which the symmetric difference between any two members has independence number at least r, is exactly a $1/2^{\binom{r}{2}}$ fraction of the number of all graphs on these vertices?*

Acknowledgment. We thank Gábor Tardos for his remark presented as Remark 3.4.

REFERENCES

[1] P. ADAMS, D. BRYANT, AND M. BUCHANAN, *A survey on the existence of G-designs*, J. Combin. Des., 16 (2008), pp. 373–410.
[2] V. E. ALEKSEEV, *On the entropy values of hereditary classes of graphs*, Discrete Math. Appl., 3 (1993), pp. 191–199.
[3] N. ALON, J. BALOGH, B. BOLLOBÁS, AND R. MORRIS, *The structure of almost all graphs in a hereditary property*, J. Combin. Theory Ser. B, 101 (2011), pp. 85–110.
[4] N. ALON, E. FISCHER, M. KRIVELEVICH, AND M. SZEGEDY, *Efficient testing of large graphs*, Combinatorica, 20 (2000), pp. 451–476.
[5] N. ALON, S. MORAN, A. YEHUDAYOFF, AND S. RANK, *VC dimension and spectral gaps*, in Proceedings of 29th Conference on Learning Theory, COLT 2016, pp. 47–80; *Mat. Sb.,* 208 (2017), pp. 1724-1757.
[6] N. ALON, L. RÓNYAI, AND T. SZABÓ, *Norm-graphs: Variations and applications*, J. Combin. Theory Ser. B, 76 (1999), pp. 280–290.
[7] N. ALON AND J. SPENCER, *The Probabilistic Method*, 4th ed., John Wiley & Sons, Hoboken, NJ, 2016.
[8] B. A. ANDERSON, *Finite topologies and Hamiltonian paths*, J. Combin. Theory Ser. B, 14 (1973), pp. 87–93.
[9] A. BERGER AND Y. ZHAO, *K_4-Intersecting Families of Graphs*, preprint, https://arxiv.org/abs/2103.12671, 2021.
[10] E. BERLEKAMP, *Algebraic Coding Theory,* revised ed., World Scientific, Hackensack, NJ, 2015.
[11] B. BOLLOBÁS AND A. THOMASON, *Projections of bodies and hereditary properties of hypergraphs*, Bull. London Math. Soc., 27 (1995), pp. 417–424.
[12] B. BOLLOBÁS AND A. THOMASON, *Hereditary and monotone properties of graphs*, The Mathematics of Paul Erdős, II, R. L. Graham and J. Nešetřil, eds*., Algorithms Combin.* 14, Springer-Verlag, Berlin, 1997, pp. 70–78.
[13] F. R. K. CHUNG, P. FRANKL, R. L. GRAHAM, AND J. B. SHEARER, *Some intersection theorems for ordered sets and graphs*, J. Combin. Theory Ser. A, 43 (1986), pp. 23–37.
[14] G. COHEN, E. FACHINI, AND J. KÖRNER, *Connector families of graphs*, Graphs Combin., 30 (2014), pp. 1417–1425.
[15] I. CSISZÁR AND J. KÖRNER, *Information Theory. Coding Theorems for Discrete Memoryless Systems*, 2nd ed., Cambridge University Press, Cambridge, 2011.
[16] R. DIESTEL, *Graph Theory*, Grad. Texts in Math. 173, 5th ed., Springer, Berlin, 2017.
[17] D. ELLIS, Y. FILMUS, AND E. FRIEDGUT, *Triangle-intersecting families of graphs*, J. Eur. Math. Soc., 14 (2012), pp. 841–885.
[18] P. ERDŐS, P. FRANKL, AND V. RÖDL, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs Combin., 2 (1986), pp. 113–121.
[19] P. ERDŐS, D. J. KLEITMAN, AND B. L. ROTHSCHILD, *Asymptotic enumeration of K_n-free graphs,* Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II, Atti dei Convegni Lincei 17, Accad Naz. Lincei, Rome, 1976 pp. 19–27

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

[20] P. Erdős, C. Ko, and R. Rado, *Intersection theorems for systems of finite sets*, Quart. J. Math. Oxford Ser. (2), 12 (1961), pp. 313–320.
[21] P. Erdős and M. Simonovits, *A limit theorem in graph theory*, Studia Sci. Math. Hungar., 1 (1966), pp. 51–57.
[22] P. Erdős and A. H. Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc., 52 (1946), pp. 1087–1091.
[23] D. J. Kleitman, *On a combinatorial conjecture of Erdős*, J. Combin. Theory, 1 (1966), pp. 209–214.
[24] M. Kobayashi, *On perfect one-factorization of the complete graph*, Graphs Combin., 5 (1989), pp. 351–353.
[25] J. Kollár, L. Rónyai, and T. Szabó, *Norm-graphs and bipartite Turán numbers*, Combinatorica, 16 (1996), pp. 399–406.
[26] J. Körner and G. Simonyi, *Trifference*, Studia Sci. Math. Hungar., 30 (1995), pp. 95–103, also in: *Combinatorics and its Applications to the Regularity and Irregularity of Structures*, W. A. Deuber and V. T. Sós eds., Akadémiai Kiadó, Budapest.
[27] J. Körner, S. Messuti, and G. Simonyi, *Families of graph-different Hamilton paths*, SIAM J. Discrete Math., 26 (2012), pp. 321–329, https://doi.org/10.1137/110837814.
[28] A. Kotzig, *Hamilton graphs and Hamilton circuits*, in *Theory of Graphs and Its Applications*, Publ. House Czechoslovak Acad. Sci., Prague, 1964, pp. 63–82.
[29] I. Kovács and D. Soltész, *Triangle-different Hamiltonian paths*, J. Combin. Theory, Ser. B, 129 (2018), pp. 1–17.
[30] T. Kővári, V. T. Sós, and P. Turán, *On a problem of K. Zarankiewicz*, Colloq. Math., 3 (1954), pp. 50–57.
[31] D. Kozlov, *Combinatorial Algebraic Topology*, Springer-Verlag, Berlin, Heidelberg, 2008.
[32] T. Mütze, *Combinatorial Gray Codes-An Updated Survey*, preprint, https://arxiv.org/abs/2202.01280, 2022.
[33] G. Nakamura, *Dudeney's round table problem for the cases of $n = p + 1$ and $n = 2p$ (in Japanese)*, Sugaku Sem., 159 (1975), pp. 24–29.
[34] A. Rosa, *Perfect* 1*-factorizations*, Math. Slovaca, 69 (2019), pp. 479–496.
[35] N. Sauer, *On the density of families of sets*, J. Combin. Theory Ser. A, 13 (1972), pp. 145–147.
[36] E. R. Scheinerman and D. H. Ullman, *Fractional Graph Theory,* Wiley-Intersci. Ser. Discrete Math. Optim., John Wiley and Sons, Chichester, 1997.
[37] S. Shelah, *A combinatorial problem; stability and order for models and theories in infinitary languages*, Pacific J. Math., 41 (1972), pp. 247–261.
[38] M. Simonovits and V. T. Sós, *Intersection theorems for graphs*, in Proceedings of the Fifth Hungarian Colloquium Combinatorics and Graph Theory, 1976, Orsay, Paris, pp. 389–391.
[39] G. Tardos, *private communication*, 2022.
[40] V. D. Tonchev, *Error-correcting codes from graphs*, Discrete Math., 257 (2002), pp. 549–557.
[41] P. Turán, *Egy gráfelméleti szélsőértékfeladatról (On an extremal problem in graph theory, in Hungarian)*, Középiskolai Matematikai és Fizikai Lapok, 48 (1941), pp. 436–452.
[42] R. M. Wilson, *Decompositions of complete graphs into subgraphs isomorphic to a given graph*, in Proceedings of the Fifth British Combinatorial Conference (Aberdeen, 1975), Congr. Numer. 15, Utilitas Math., 1976, pp. 647–659.

Downloaded 04/04/24 to 193.224.79.242 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.