Some Avenues in Graph Codes

by

K. B. Zeven

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday August 27, 2025 at 2:00 PM.

Student number:	5375444	
Project duration:	January 24, 2025 – August 27, 2025	
Thesis committee:	Dr. A. Bishnoi,	TU Delft, supervisor
	Prof. Dr. D.C. Gijswijt,	TU Delft, chair
	Dr. R.J. Fokkink,	TU Delft

TU Delft

Abstract

For a family of graphs $\mathcal{H}$, the maximum size of a collection of graphs $\mathcal{F}$ for which the symmetric difference $\oplus$ of two distinct graphs G_1, G_2 has the property $G_1 \oplus G_2 \notin \mathcal{H}$ (or $\in \mathcal{H}$) is denoted by $D_{\mathcal{H}}(n)$ (or $M_{\mathcal{H}}(n)$ respectively). We also denote the independence ratio by $d_{\mathcal{H}}(n) = D_{\mathcal{H}}(n)/2^{\binom{n}{2}}$.

This thesis gives a new approach to get an upper bound of $d_{\mathcal{H}}(n)$ where $\mathcal{H}$ is the family of graphs that contain cycles of length $2k$, i.e. $\mathcal{H} = \{H : C_{2k} \subseteq H\}$. The proof of which is based on a different proof by Noga Alon concerning the upper bound of $d_{\mathcal{H}}(n)$ where $\mathcal{H}$ is the family of graphs that contain stars with $2k$ edges, i.e. $\mathcal{H} = \{H : K_{1,2k} \subseteq H\}$. A bound on the number of C_{2k}-free graphs proven by Morris and Saxton indirectly already solved this problem, but Alon's approach could be expanded upon for different types of graph families for which the upper bound is not yet known. In this thesis we also perform a spectral analysis of the Cayley graph corresponding to the families $\mathcal{H} = \{H : L \subseteq H\}$ and $\mathcal{H} = \{H : L \cong H\}$, in order to give upper bounds on $D_{\mathcal{H}}(n)$. These upper bounds are compared to known lower bounds.

Contents

1 Introduction

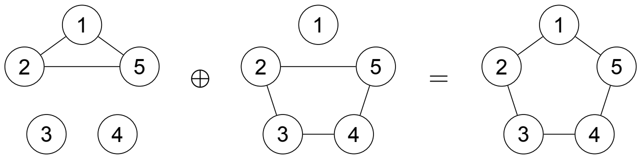
In this thesis, we will be analysing the maximum cardinality which a collection of graphs can be under the restriction that the symmetric difference of any two graphs within the collection does not contain certain subgraphs. The symmetric difference $\oplus$ can be seen as binary addition, or bit-wise XOR addition of the binary edge representations of the two graphs. In more technical terms, the *symmetric difference* of two graphs G_1, G_2 on the same n labeled vertices, is the graph $G = G_1 \oplus G_2$ with

$$V(G) = V(G_1) = V(G_2), \quad E(G) = (E(G_1) \setminus E(G_2)) \cup (E(G_2) \setminus E(G_1)).$$

That is, the edges of the symmetric difference are precisely the edges which are in exactly one of the two graphs. Equivalently, the edge set of the symmetric difference $G = G_1 \oplus G_2$ can also be denoted by taking the union of both edge sets, and removing the intersection of the edge sets, i.e.

$$E(G) = (E(G_1) \cup E(G_2)) \setminus (E(G_1) \cap E(G_2)).$$

For a visual, we refer to equation (1.1), where the symmetric difference of the 3-cycle 125 and the 4-cycle 2345 gives the 5-cycle 12345.



(1.1)

Indeed, edges 12 and 15 are only in the first graph, edges 23, 34 and 45 are only in the second graph, and finally the edge 25 is in both the first and the second graph. All other edges are in none of the two graphs, hence the symmetric difference is precisely the graph with the edges 12, 23, 34, 45 and 15, i.e. the 5-cycle 12345.

More specifically for this thesis we are interested in two objects:

Definition 1.1. Let $\mathcal{H}$ be a collection of graphs on n labeled vertices. We define $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ respectively as the maximum cardinality of a collection $\mathcal{F}$ of graphs on n labeled vertices with the condition that

- for any two distinct graphs $G_1, G_2 \in \mathcal{F}$, $G_1 \oplus G_2 \notin \mathcal{H}$, (this is the condition for $D_{\mathcal{H}}(n)$)
- for any two distinct graphs $G_1, G_2 \in \mathcal{F}$, $G_1 \oplus G_2 \in \mathcal{H}$. (this is the condition for $M_{\mathcal{H}}(n)$)

As a rule of thumb, you could remember which condition belongs to which function $D_{\mathcal{H}}(n)$ or $M_{\mathcal{H}}(n)$, by D for 'Do not' as in '$\notin \mathcal{H}$', and M for 'Must' as in '$\in \mathcal{H}$'. We note that the two objects $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ are closely related and they are dependent on each other, which will be shown later in Chapter 3. A collection $\mathcal{F}$ which satisfies the conditions for the $D_{\mathcal{H}}(n)$ problem, will be called an $\mathcal{H}$-*code*, due to it's resemblance to a problem within coding theory.

Also note that the conditions for $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ only apply to distinct graphs, hence the conditions are void and automatically met if there is only one graph in the collection $\mathcal{F}$. Hence $D_{\mathcal{H}}(n) \geq 1$

and $M_{\mathcal{H}}(n) \geq 1$ for any n. Furthermore, by definition, the objects $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ cannot exceed the total number of graphs on n vertices regardless of the collection $\mathcal{H}$. That is, $D_{\mathcal{H}}(n) \leq 2^{\binom{n}{2}}$ and $M_{\mathcal{H}}(n) \leq 2^{\binom{n}{2}}$. We give an easy example to get some intuition.

Example 1.2. Let $\mathcal{H} = \{H : K_2 \subseteq H\}$ and consider the problem of finding $D_{\mathcal{H}}(n)$. That is we are trying to find a maximum cardinality collection $\mathcal{F}$ of graphs on n vertices, such that for any two distinct graphs in $\mathcal{F}$, their symmetric difference does not contain an edge. Note that any two distinct graphs G_1 and G_2 must differ in some edge, otherwise they would be the same graph. Therefore, the symmetric difference of two distinct graphs will always contain an edge. Hence $D_{\mathcal{H}}(n) \leq 1$, and since $D_{\mathcal{H}'}(n) \geq 1$ for any $\mathcal{H}'$ we have equality: $D_{\mathcal{H}}(n) = 1$.

Similarly, by the exact same argument, the collection of all graphs on n vertices is a valid collection for the problem of $M_{\mathcal{H}}(n)$. Therefore $M_{\mathcal{H}}(n) \geq 2^{\binom{n}{2}}$, and since $M_{\mathcal{H}'}(n) \leq 2^{\binom{n}{2}}$ for any $\mathcal{H}'$, we have equality: $M_{\mathcal{H}}(n) = 2^{\binom{n}{2}}$.

This problem is a variation on a well known problem within the field of coding theory. We will give a few definitions first, which can be found in an algebra book on the topic, for instance in [17]. A *code* C is a collection of *codewords* $\{c_i\}_i$. Codewords are typically binary strings, and usually the codewords in a code are of the same length, say n. The *Hamming distance* d^H between two codewords c_1, c_2 is a metric and is defined as the number of positions where the codewords do not match, i.e. $d^H(c_1, c_2) = |\{i \in [n] : (c_1)_i \neq (c_2)_i\}|$. We note that, by definition, the Hamming distance between two binary strings c_1, c_2 is equal to the number of 1's in the bit-wise XOR of the two strings.

A code C is e-error detecting if the Hamming distance between any two codewords in C is greater than e, and so if the distance between a fixed codeword c' and any arbitrary codeword c is less or equal to e, c' is not a codeword, and must have an error. Lastly, a code C is e-error correcting if for any binary string w of length n, there exists at least one codeword $c \in C$ such that $d^H(w, c) \leq e$. That is, we can cover the space of binary strings of length n with Hamming spheres of radius e around each codeword c. This also implies that a code is e-error correcting if the minimum Hamming distance of tha

The well known problem in coding theory, which was referred to earlier, is the following: for a constant d and fixed codeword length n, what is the largest cardinality of a code C such that the Hamming distance between any two codewords of C is at least d. In terms of error correction, a code with minimal Hamming distance d is $(d-1)$-error detecting and $\lfloor (d-1)/2 \rfloor$-error correcting, which is optimal in efficiency.

A variation of this problem is, instead of restricting the codewords by minimum distance, restricting the codewords by avoiding certain substructures in the difference between codewords within the code. These substructures are more easily explainable in a graph theoretic sense. Noting that a graph on n vertices has $\binom{n}{2}$ possible edges, if we have a codeword of length $\binom{n}{2}$ for some positive integer n, we can associate that codeword with a graph on n vertices, by mapping each bit in the codeword to an edge in the graph. That is, codewords of length $\binom{n}{2}$ are characteristic vectors of edge-sets of graphs on n, and we can map them bijectively. Furthermore, as we have defined it this way, the difference of two codewords of length $\binom{n}{2}$ is the same as the symmetric difference of the graphs on n vertices belong to the two codewords by viewing the codewords as characteristic vectors of edge-sets.

The collection $\mathcal{H}$ for the problem of finding $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ can be chosen arbitrarily. However, this paper will mainly be focused on a few specific local graph classes, as defined in Noga Alon's paper "Structured Codes on Graphs" [2]. The main class $\mathcal{H}$ we are interested in, is the class of graphs which contain a certain graph L as a subgraph, i.e. $\mathcal{H} = \{H : L \subseteq H\}$. Another graph class which is considered, while not local, is the class of graphs which are isomorphic to a certain graph L, i.e. $\mathcal{H} = \{H : L \cong H\}$.

Chapter 2 goes into the known results regarding our main problem, some variations of the main problem, and other results regarding graph codes. Chapter 3 starts with central theorems also described in "Structured Codes on Graphs" [2] regarding the objects $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ in Section 3.1. Afterwards, we study the so called independence ratio and give a new result for the independence ratio for cycles of even length in Section 3.2. Chapter 4 considers the eigenvalues of the Cayley graph corresponding with the family $\mathcal{H}$, in order to bound $D_{\mathcal{H}}(n)$, starting with giving the upper bounds we will use in Section 4.1. Afterwards we give the theory on how to use character sums to find the spectrum of the Cayley graph in Section 4.2. Finally we compare the upper bounds of $D_{\mathcal{H}}(n)$ against known lower bounds for a plethora of families $\mathcal{H}$ in Section 4.3. Lastly, Chapter 5 will give topics of further research.

2

Known results from the literature

This chapter will go into some results pertaining the main problem of finding the values $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ as defined in the introduction. Some of the results will be further explored in Chapter 3, with most attention going to the independence ratio. Other results regarding variations of this main problem are also be discussed here.

2.1. Results on the main problem

Most work on the specific problem of $\mathcal{H}$-codes so far has been done by Noga Alon, and his work is cited heavily in this paper. In "Structured Codes on Graphs" (2022) [2], Alon et al. proved exact values of $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ for several different families $\mathcal{H}$.

The list of families for which exact values were proven include the family of connected graphs and the family of 2-connected graphs - but exact values were only proven for an even number of vertices n. Also included are the family of graphs containing a spanning star, the family of graphs containing a Hamiltonian path, and the family of graphs containing a Hamiltonian cycle - for which the values were only proven for even values of n for which the perfect-1-factorization conjecture, which was conjectured by Kotzig [13], holds.

Some other results concern the values of exact values of $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ for the family of graphs $\mathcal{H}$ which contain a triangle K_3 - but only for the values $n=3,4,5,6$, and the family of graph which contain an odd cycle - but only for the values $n=3,4,5,6,7$. The case of the family containing a triangle will also be briefly mentioned at the end of Section 3.1.

In a followup paper, Alon [1] considered the so called independence ratio, for which the definition is given here.

Definition 2.1 (Independence ratio)**.** For a given family of graphs on n vertices $\mathcal{H}$, we define the *independence ratio* as

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

For this paper, Alon considered the families closed under isomorphism, such as the family $\mathcal{H}=\{H : L \cong H\}$ where L is some graph on n vertices. One can quickly prove for this family $\mathcal{H}$ that if L is a graph with an odd number of edges, then $D_{\mathcal{H}}(n)$ would be exactly $2^{\binom{n}{2}-1}$, and therefore $d_{\mathcal{H}}(n) = \frac{1}{2}$, see Example 3.6 for a proof. This gives rise to the main question Alon [1] poses in his paper:

"If $\mathcal{H}$ is a family closed under isomorphism, and there is a graph $H \in \mathcal{H}$ with an even number of edges, is it true that then $d_{\mathcal{H}}(n) \to 0$ as $n \to \infty$?".

In that same paper, Alon proved the following noteworthy results which answer the question posed for a selection of graphs with an even number of edges. We use the following notations: $K_{1,p}$ as a star with p edges, M_q as a matching of q edges, i.e. a graph with q edges where no two edges share an endpoint. We denote $\mathcal{K}(r)$ as the family of all cliques on at most r vertices, and $\mathcal{K}$ as the family of all cliques. Alon also uses the shorthand writing of $d_L(n)$ instead of $d_{\mathcal{H}}(n)$ when $\mathcal{H}=\{H : L \cong H\}$.

Theorem ([1]). *For all $k \in \mathbb{Z}_{\geq 1}$*

$$d_{K_{1,2k}}(n) = \Theta_k\left(\frac{1}{n^k}\right), \quad d_{M_{2k}}(n) = \Theta_k\left(\frac{1}{n^k}\right), \quad d_{\mathcal{K}(4k+3)}(n) = \Omega_k\left(\frac{1}{n^k}\right),$$

and also

$$d_{\mathcal{K}}(n) \geq \frac{1}{2^{\lfloor n/2 \rfloor}}.$$

There is another result concerning the family $\mathcal{H}$ of graphs which are isomorphic to a graph which is *made from two copies of H' on an independent set I*, see Definition 3.12, and the result is proven again in this thesis in Lemma 3.13. In particular, the bounds of $d_{K_{1,2k}}(n)$ and its proofs will also be further explored in Chapter 3.

Following up Alon's paper regarding the independence ratio, Versteegen [20] considered the problem of finding $D_{\mathcal{H}}(n)$ where the solution has to be a linear $\mathcal{H}$-code, i.e. a solution to $D_{\mathcal{H}}(n)$ which is closed under taking the symmetric difference. This solution is denoted by $D^{\text{lin}}_{\mathcal{H}}(n)$ and the corresponding independence ratio is denoted by $d^{\text{lin}}_{\mathcal{H}}(n)$.

Versteegen proved the following result:

Theorem ([20]). *There exists a constant $c > 0$ such that if $\mathcal{H}$ is a family of graphs closed under isomorphism and there is an $H \in \mathcal{H}$ with an even number of edges, then $d^{\text{lin}}_{\mathcal{H}}(n) \leq (c \cdot \log n)^{-1}$.*

This answers the question posed by Noga Alon for the specific $\mathcal{H}$ that have linear $\mathcal{H}$-codes. Since the problem $D^{\text{lin}}_{\mathcal{H}}(n)$ is more restricted than the nonlinear counterpart, we have $D^{\text{lin}}_{\mathcal{H}}(n) \leq D_{\mathcal{H}}(n)$, and therefore $d^{\text{lin}}_{\mathcal{H}}(n) \leq d_{\mathcal{H}}(n)$. Therefore the previous theorem's upper bound for $d^{\text{lin}}_{\mathcal{H}}(n)$ cannot be applied for $d_{\mathcal{H}}(n)$. These linear $\mathcal{H}$-codes will not be considered for this thesis, but are noteworthy nonetheless.

2.2. Variations of the main problem

Other variations of the $\mathcal{H}$-code problem have been studied as well. As an example, in their 2010 paper, Ellis, Friedgut and Filmus [6] considered so called *triangle-intersecting* families $\mathcal{F}$; a family of graphs $\mathcal{F}$ on n vertices is *triangle-intersecting* if for any two graphs $G_1, G_2 \in \mathcal{F}$ we have that $G_1 \cap G_2$, where $E(G_1 \cap G_2) = E(G_1) \cap E(G_2)$, contains a triangle. This problem stemmed from earlier work on intersecting families. Some of the work done is e.g. the results from Erdös, Ko and Rado [7] on the maximum size of k-uniform families where any two sets in the family share at least one element, and Katona's contribution [11] to determine the maximum size of families $\mathcal{F}$ where any two members of $\mathcal{F}$ have an intersection of size at least j for any j.

Ellis, Friedgut and Filmus [6] were able to prove a conjecture posed by Simonovits and Sós made in 1976, which can be found in [4], regarding the maximum possible size of such a triangle-intersecting family $\mathcal{F}$ on n vertices.

What was already known is that for any $n \geq 3$ we have $|\mathcal{F}| \geq 2^{\binom{n}{2}-3}$, and Simonovits and Sós conjectured that this was tight. The lower bound follows from the following proof:

Pick three vertices, $1, 2$ and 3 for example, and consider the family $\mathcal{F}$ of all graphs containing the triangle 123. It is clear from the definition that two graphs within $\mathcal{F}$ intersect at least in the triangle 123, and thus $\mathcal{F}$ is triangle-intersecting. The cardinality of $\mathcal{F}$ is the number of graphs which contain the edges 12, 13 and 23. One can count every graph by coin-flipping for each edge, which is not $12, 13$ or 23, whether to put it in the graph or not. Therefore we get $|\mathcal{F}| \geq 2^{\binom{n}{2}-3}$. Ellis, Friedgut and Filmus [6] were able to ascertain the upper bound $|\mathcal{F}| \leq 2^{\binom{n}{2}-3}$, using Fourier analytic methods. The upper bound was also proven to be tight if and only if the graphs in $\mathcal{F}$ contained a fixed triangle.

This paper sparked a conjecture that their results could be applied to cross-intersecting families and K_t-intersecting families. Berger and Zhao [3] proved this conjecture for K_4-intersecting families. That is, they proved that if $\mathcal{F}_1$ and $\mathcal{F}_2$ are families of graphs on n vertices and for any $G_1 \in \mathcal{F}_1$ and any $G_2 \in \mathcal{F}_2$ we have that $G_1 \cap G_2$ has K_4 as a subgraph, then $|\mathcal{F}_1||\mathcal{F}_2| \leq 4^{\binom{n}{2}-6}$, with equality if and only if $\mathcal{F}_1 = \mathcal{F}_2$ are the family of graphs on n vertices containing a fixed K_4.

2.3. Other results

Some noteworthy examples of the concept of constructing codes from graphs can be found in e.g. Tonchev's 2002 paper [18]. Denoting A as the adjacency matrix of some graph H on n vertices, Tonchev considered two generator matrices, (a) $G = [I_n, A]$ and (b) $G = A$. In both cases (a) and (b), one can define a linear code by taking the row space of the generator matrix. The linear code for type (a) has codewords of length $2n$, dimension n and a minimum hamming distance $d \leq \delta(H)+1$ where $\delta(H)$ is the minimum degree of H. The linear code for type (b) has codewords of length n, dimension equal to $\operatorname{rank}(A)$ over the $\mathbb{F}_2$, and minimum Hamming distance $d \leq \delta(H)$.

Tonchev then proves the following results:

Theorem ([18]). *The class $\mathcal{A}$ of binary linear codes of length $2n$ and dimension n defined by generator matrices of the form $[I_n, A]$ where A is a symmetric matrix, contains codes with minimum Hamming distance $d \geq 0.22n$.*

A *strongly regular graph* with parameters (n,k,a,c) is a graph on n vertices which is k-regular, each pair of adjacent vertices has a common neighbors, and each pair of non-adjacent vertices has c common neighbors. For this class of graphs, Tonchev gives the following result:

Theorem ([18]). *The dual code of a code of types (a) or (b) defined by a strongly regular graph with parameters (n,k,a,c) can correct up to*

$$\frac{k+\max(a,c)-1}{2\max(a,c)}$$

errors.

Kopparty et al. [12] defined graph codes using a slightly different distance metric. Here $G[S]$ denotes the induced subgraph of G on the vertices of $S \subseteq V(G)$.

Definition 2.2. Given two graphs G and H on the vertex set $[n]$, the graph distance $d_{\text{graph}}(G,H)$ is the size of the smallest set $S \subseteq [n]$ such that $G[[n] \setminus S] = H[[n] \setminus S]$.

They then define the rate and distance for a graph code C with n vertices where each edge xy is given a number $\alpha \in \mathbb{F}_q$, corresponding the number of 'edge types' there are on xy. If $q = 2$, every edge is given either a 0 or a 1. This corresponds to a simple graph, with an edge getting 0 meaning it is not in the graph, and an edge getting 1 meaning it is in the graph. We continue with $q = 2$ since this thesis pertains only simple graphs. The rate and distance are defined as follows:

- Rate: $R = \log_2(|C|)/\binom{n}{2}$
- Distance: The distance of a code is the largest d such that $d_{\text{graph}}(G,H)$ for all distinct $G,H \in C$. The relative distance δ is d/n.

Kopparty et al.[12] then show, among other results, that there exist binary graph codes achieving a rate of $R = (1-\delta)^2 - o(1)$ for any constant $\delta \in (0,1)$. We wont go further into the details and other results, and one should check out the rest of the results of this paper for themselves.

3

Central theorems, and new results on $\mathcal{H}$-codes

For this thesis, we are investigating, for a collection of graphs $\mathcal{H}$, the largest family of graphs $\mathcal{F}$ on n vertices such that the symmetric difference of two graphs G_1, G_2 within $\mathcal{F}$ share, or do not share, certain subgraphs $H \in \mathcal{H}$. We give the definition for the main objects which was given in the introduction again here for convenience.

Definition 3.1. Let $\mathcal{H}$ be a collection of graphs on n labeled vertices. We define $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ respectively as the maximum cardinality of a collection $\mathcal{F}$ of graphs on n labeled vertices with the condition that

- for any two distinct graphs $G_1, G_2 \in \mathcal{F}$, $G_1 \oplus G_2 \notin \mathcal{H}$, (this is the condition for $D_{\mathcal{H}}(n)$)
- for any two distinct graphs $G_1, G_2 \in \mathcal{F}$, $G_1 \oplus G_2 \in \mathcal{H}$. (this is the condition for $M_{\mathcal{H}}(n)$)

We start off in Section 3.1 with a few general results for the quantities $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ for any collection $\mathcal{H}$. In section 3.2 we further investigate the quantity $D_{\mathcal{H}}(n)$ and the independence ratio. In this section we will also prove a new result regarding the independence ratio when $\mathcal{H}$ is the collection of graphs containing cycles of length $2k$, for a fixed k.

3.1. Main theorems

Before we start with giving results regarding the objects $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$, we first give a result on *vertex-transitive* graphs. A graph G is *vertex-transitive* if for any two distinct vertices $v_1, v_2 \in V(G)$ we can find a graph automorphism $\varphi : V(G) \to V(G)$ such that $\varphi(v_1) = v_2$. That is, in a vertex-transitive graph, we can always swap the labels of any two vertices, and relabel other vertices accordingly to preserve edge-relations, and end up with the same graph. Therefore, in a vertex-transitive graph, every vertex has the same properties as any other vertex in the graph, e.g. every vertex has the same degree, belongs to the same number of maximum cliques or number of maximum independent sets etc.

We first give a result on vertex-transitive graphs regarding its independence number and its clique number, which will be useful in the future. This result is folklore and can be found in a book on the topic, for instance in [10], but we prove it here as well.

Theorem 3.2 (Clique-Coclique Bound [10]). *For a vertex-transitive graph G, we have*

$$\alpha(G)\omega(G) \leq |V(G)|.$$

Proof. Since G is vertex-transitive, every vertex is contained in at least one and the same number of maximum cliques. Let $\mathcal{C}$ be the set of all maximum size cliques in G, and let k be the number of maximum size cliques which contain a fixed vertex from G. Since G is vertex-transitive, this number k is the same ragrdless of fixed vertex.

Now let I be an independent set of arbitrary size. Then I intersects every maximum clique in $\mathcal{C}$ at most one vertex, any more and two independent vertices from I would belong to the same clique, which is contradictory. Therefore, since every vertex in G is in exactly k cliques from $\mathcal{C}$, $|\mathcal{C}|$ is at least

$$|\mathcal{C}| \geq \sum_{v \in I} \#\{\text{max. size clique containing } v\} = k|I|. \tag{3.1}$$

Also, since every vertex in G belongs to exactly k cliques from $\mathcal{C}$, the sum of the number of vertices on a clique of $\mathcal{C}$ over all cliques of $\mathcal{C}$ is precisely k times the number of vertices, i.e.

$$|\mathcal{C}|\omega(G) = k|V(G)|. \tag{3.2}$$

Therefore, combining Equations (3.1) and (3.2), we get

$$\frac{k|V(G)|}{\omega(G)} \geq k|I| \implies |V(G)| \geq |I|\omega(G).$$

Since this is true for any independent set I, it also holds for an indepedent set of maximum size, and therefore $\alpha(G)\omega(G) \leq |V(G)|$, showing Theorem 3.2. □

The following lemma was shown by Alon et al., 2022 [2], but will be proven here again with a more extended formulation.

Lemma 3.3 ([2]). *For any collection of graphs $\mathcal{H}$ we have*

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

Proof. Let $G_{\mathcal{H}}$ be the graph whose vertices are all graphs on n vertices, and two graphs G_1, G_2 on n vertices are joined by an edge if and only if $G_1 \oplus G_2 \in \mathcal{H}$. We will show that the graph $G_{\mathcal{H}}$ is vertex transitive.

We denote the set of all graphs on n vertices by $G[n]$. Now in $G_{\mathcal{H}}$ we can map any vertex G_1 to another vertex G_2 via the function $\varphi : G[n] \to G[n]$ defined by $\varphi(G) = G \oplus (G_1 \oplus G_2)$. Now we have

$$\varphi(G_1) = G_1 \oplus (G_1 \oplus G_2) = (G_1 \oplus G_1) \oplus G_2 = \overline{K_n} \oplus G_2 = G_2,$$

and we have

$$\begin{aligned} G_iG_j \in E(G_{\mathcal{H}}) &\iff \mathcal{H} \ni G_i \oplus G_j \\ &\iff \mathcal{H} \ni (G_i \oplus G_j) \oplus ((G_1 \oplus G_2) \oplus (G_1 \oplus G_2)) \\ &\iff \mathcal{H} \ni (G_i \oplus (G_1 \oplus G_2)) \oplus (G_j \oplus (G_1 \oplus G_2)) \\ &\iff \mathcal{H} \ni \varphi(G_i) \oplus \varphi(G_j) \\ &\iff \varphi(G_i)\varphi(G_j) \in E(G_{\mathcal{H}}). \end{aligned}$$

Thus φ is an automorphism of $G_{\mathcal{H}}$ mapping G_1 to G_2, and hence $G_{\mathcal{H}}$ is vertex transitive.

Since $G_{\mathcal{H}}$ is vertex transitive, we can apply Theorem 3.2, the Clique-Coclique Bound, and find $\alpha(G_{\mathcal{H}}) \cdot \omega(G_{\mathcal{H}}) \leq |V(G_{\mathcal{H}})|$. By definition we now have $\alpha(G_{\mathcal{H}}) = D_{\mathcal{H}}(n)$ and $\omega(G_{\mathcal{H}}) = M_{\mathcal{H}}(n)$, and finally there are $2^{\binom{n}{2}}$ graphs on n vertices, so $G_{\mathcal{H}}$ consists of $2^{\binom{n}{2}}$ vertices. The theorem immediately follows from substituting in these values in the inequality. □

Remark 3.4. As noted in the proof of Lemma 3.3, we showed that $D_{\mathcal{H}}(n)$ is the independence number of the graph $G_{\mathcal{H}}$ and $M_{\mathcal{H}}(n)$ is the clique number of $G_{\mathcal{H}}$. For this reason, we can refer to $D_{\mathcal{H}}(n)$ as the independence problem, and $M_{\mathcal{H}}(n)$ as the clique problem.

Using Lemma 3.3, we can determine the values of $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ for some families $\mathcal{H}$.

Example 3.5. Consider for n vertices, the problem of finding $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ where $\mathcal{H} = \{H : K_n \subseteq H\} = \{H : K_n \cong H\}$. Note that by definition of the symmetric difference, we have that two graphs G_1, G_2 on n vertices have the property that $G_1 \oplus G_2 = K_n$ if and only if G_2 is the complement of G_1, i.e. $G_2 = \overline{G_1}$.

Therefore the collection $\mathcal{F} = \{\overline{K_n}, K_n\}$ is obviously valid for the clique problem $M_{\mathcal{H}}(n)$, and hence $M_{\mathcal{H}}(n) \geq 2$. We now use Lemma 3.3 and find that

$$2^{\binom{n}{2}} \geq M_{\mathcal{H}}(n) D_{\mathcal{H}}(n) \geq 2 D_{\mathcal{H}}(n) \implies D_{\mathcal{H}}(n) \leq 2^{\binom{n}{2}-1}.$$

Next, if we let $\mathcal{G}$ be the maximum cardinality collection of graphs on n vertices such that for any graph $G \in \mathcal{G}$, $\overline{G} \notin \mathcal{G}$, then $\mathcal{G}$ is a valid collection for the independence problem $D_{\mathcal{H}}(n)$. The cardinality of $\mathcal{G}$ is precisely half the total number of graphs, hence

$$D_{\mathcal{H}}(n) \geq \frac{2^{\binom{n}{2}}}{2} = 2^{\binom{n}{2}-1}.$$

This proves equality for the independence problem: $D_{\mathcal{H}}(n) = 2^{\binom{n}{2}-1}$, and using Lemma 3.3 again, also proves equality for the clique problem: $M_{\mathcal{H}}(n) = 2$.

Example 3.6. Consider for n vertices, the family $\mathcal{H} = \{H : F \cong H\}$ where F is a graph with an odd number of edges. The collection $\mathcal{F} = \{\overline{K_n}, F\}$ is again clearly valid for the clique problem, giving $M_{\mathcal{H}}(n) \geq 2$, which gives $D_{\mathcal{H}}(n) \leq 2^{\binom{n}{2}-1}$ as in Example 3.5.

Next, consider the collection $\mathcal{G}$ of all graphs with an even number of edges. By definition of the symmetric difference, for two graphs G_1, G_2 on n vertices, the number of edges of the symmetric difference $G = G_1 \oplus G_2$ is equal to

$$\begin{aligned} |E(G_1) \cup E(G_2)| - |E(G_1) \cap E(G_2)| &= \Big(|E(G_1)| + |E(G_2)| - |E(G_1) \cap E(G_2)|\Big) - |E(G_1) \cap E(G_2)| \\ &= |E(G_1)| + |E(G_2)| - 2|E(G_1) \cap E(G_2)|. \end{aligned} \tag{3.3}$$

Hence, for any two graphs H_1, H_2 in $\mathcal{G}$, we have that $|E(H_1)|$ and $|E(H_2)|$ are both even, and by equation (3.3), $H_1 \oplus H_2$ has an even number of edges. Therefore, for all graphs H_1, H_2 in $\mathcal{G}$ we have that $H_1 \oplus H_2 \ncong F$, and $\mathcal{G}$ is a valid collection for the independence problem.

Now we determine the size of $\mathcal{G}$ to give a lower bound for $D_{\mathcal{H}}(n)$. Note that the total number of graphs with $2k$ edges for some $0 \leq k \leq \lfloor n/2 \rfloor$ is equal to the number of combinations of $2k$ edges from $\binom{n}{2}$ edges, i.e. $\binom{m}{2k}$ where $m = \binom{n}{2}$. Then we simply have by the binomial formula:

$$\begin{aligned} D_{\mathcal{H}}(n) \geq |\mathcal{G}| = \sum_{k=0}^{\lfloor m/2 \rfloor} \binom{m}{2k} &= \frac{1}{2}\left(\sum_{k=0}^{m} \binom{m}{k} + \sum_{k=0}^{m} \binom{m}{k}(-1)^k\right) \\ &= \frac{1}{2}\left(\sum_{k=0}^{m} \binom{m}{k} 1^k 1^{m-k} + \sum_{k=0}^{m} \binom{m}{k}(-1)^k 1^{m-k}\right) \\ &= \frac{1}{2}\left((1+1)^m + (1+(-1))^m\right) = \frac{1}{2}(2^m) = 2^{m-1} = 2^{\binom{n}{2}-1}. \end{aligned}$$

This shows equality for the independence problem: $D_{\mathcal{H}}(n) = 2^{\binom{n}{2}-1}$, and using Lemma 3.3 we also get equality for the clique problem: $M_{\mathcal{H}}(n)$.

Along with Remark 3.4, we give another definition that pertains to the graph $G_{\mathcal{H}}$ from the proof of Lemma 3.3, and which will become useful in the future.

Definition 3.7 (Cayley digraph and Cayley graph)**.** For a group Γ and a set $S \subseteq \Gamma$, the *Cayley digraph* $G = \mathit{Cay}(\Gamma, S)$ is the directed graph whose vertices are the group elements, i.e. $V(G) = \Gamma$, and for the edges: $E(G) = \{(a,b) : ba^{-1} \in S\}$. If S does not contain the group identity of Γ and S is closed under taking inverses, then G contains no loops and only has undirected edges, and is said to be a *Cayley graph*.

Remark 3.8. As noted in the introduction, by representing graphs by their characteristic edge vector, we have a bijective mapping between graphs on n vertices and binary strings of length $m = \binom{n}{2}$. To this effect, let $\Gamma = \mathbb{Z}_2^m$, and let $S \subset \Gamma$ be the set of binary strings representations of the graphs $H \in \mathcal{H}$.

We find that the graph $G_{\mathcal{H}}$ in the proof of Lemma 3.3 is precisely the Cayley graph $Cay(\Gamma, S)$. This is evident from the following equivalence: Any two vertices (graphs) G_1, G_2 from the graph $G_{\mathcal{H}}$ have an edge iff $G_1 \oplus G_2 \in \mathcal{H}$, if and only if the binary edge representations of the two graphs $b_{G_1}, b_{G_2} \in \mathbb{Z}_2^m$ have the property that

$$b_{G_1} \overset{\text{XOR}}{\oplus} b_{G_2}^{-1} = b_{G_1} \overset{\text{XOR}}{\oplus} b_{G_2} \in S.$$

Note that this is a Cayley graph and not a digraph, as the operation of bitwise XOR is commutative, and the empty graph on n vertices, which is the graph corresponding to the edge vector $\mathbf{0}$, is never in $\mathcal{H}$.

In extremal graph theory, for a collection of graphs $\mathcal{G}$, the extremal number $\mathrm{ex}(n, \mathcal{G})$ denotes the maximum amount of edges a graph on n vertices can have without having a subgraph isomorphic to a G in $\mathcal{G}$. The following lemma gives some insight in the possible values both $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ could obtain for local graph classes, as defined in [2].

Lemma 3.9 ([2])**.** *For any local graph class $\mathcal{H}$ we have*

$$D_{\mathcal{H}}(n) \geq 2^{\mathrm{ex}(n,\mathcal{H})}, \tag{3.4}$$

and therefore

$$M_{\mathcal{H}}(n) \leq 2^{\binom{n}{2} - \mathrm{ex}(n,\mathcal{H})}. \tag{3.5}$$

Proof. We only have to show $D_{\mathcal{H}}(n) \geq 2^{\mathrm{ex}(n,\mathcal{H})}$ as the second inequality then follows from Lemma 3.3. To prove the first inequality we give a construction.

Specifically, let G be an n-vertex extremal graph for $\mathcal{H}$, i.e. $|E(G)| = \mathrm{ex}(n, \mathcal{H})$ and no subgraph of G is isomorphic to any H in $\mathcal{H}$. Let $\mathcal{G}$ be the collection of all subgraphs of G. Then the symmetric difference of any two graphs in $\mathcal{G}$ is itself a subgraph of G, and therefore cannot have a subgraph isomorphic to any H in $\mathcal{H}$. Therefore $\mathcal{G}$ satisfies the requirements for $D_{\mathcal{H}}(n)$ and we obtain that $D_{\mathcal{H}}(n) \geq |\mathcal{G}| = 2^{\mathrm{ex}(n,\mathcal{H})}$. □

Lemma 3.9 is useful for giving lower bounds, such as in Example 3.10.

Example 3.10. Consider $\mathcal{H} = \{H : P_3 \subseteq H\}$; P_3 is the path graph on 3 vertices. By Lemma 3.9 we know $D_{\mathcal{H}}(n) \geq 2^{\mathrm{ex}(n,\mathcal{H})}$. Note that $\mathrm{ex}(n, \mathcal{H})$ is the maximum number of edges that a graph G on n vertices can have such that G does not contain a P_3 as a subgraph. Since P_3 is the same as two edges linked by a single endpoint, $\mathrm{ex}(n, \mathcal{H})$ is the maximum number of edges that a graph G on n vertices can have such that no two edges of G share an endpoint. This is precisely the definition of a matching, hence $\mathrm{ex}(n, \mathcal{H}) = \mu(K_n) = \lfloor n/2 \rfloor$. This gives $D_{\mathcal{H}}(n) \geq 2^{\lfloor n/2 \rfloor}$.

We note that the problem of finding $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ is an open problem for many families $\mathcal{H}$, even families that are relatively simple. For instance, for $\mathcal{H} = \{H : K_3 \subseteq H\}$, $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ are open for $n \geq 7$. In this case, we know that $\mathrm{ex}(n, \{K_3\}) = \lfloor n^2/4 \rfloor$, which gives a lower bound on $D_{\mathcal{H}}(n)$ of $2^{\lfloor n^2/4 \rfloor}$ via Lemma 3.9. To prove the lower bound is tight for any n, we have two options: either

- Give a general proof that having more than $2^{\lfloor n^2/4 \rfloor}$ graphs inside a collection gives rise to two graphs whose symmetric difference contains a triangle, thereby immediately proving the upper bound of inequality (3.4), or
- Give a construction for a collection of $2^{\binom{n}{2} - \lfloor n^2/4 \rfloor}$ graphs such that every two graphs in that collection have their symmetric difference contain a triangle, for any n, thereby proving the tightness of inequality (3.5), implying the tightness of inequality (3.4).

Alon et al. [2] gave constructions for the clique problem $M_{\{K_3\}}(n)$ for $n = 3, 4, 5, 6$. These constructions give the exact value as the upper bound in (3.5), thereby giving the exact lower bound in (3.4).

3.2. Bounds on the independence ratio $d_{\mathcal{H}}(n)$ for some families $\mathcal{H}$

We introduce the independence ratio as defined in Alon's paper "Graph Codes" [1]:

Definition 3.11 (Independence ratio)**.** For the maximum cardinality of an $\mathcal{H}$-free code on n vertices $D_{\mathcal{H}}(n)$ we define the **independence ratio** as

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

By Lemma 3.3 and Remark 3.8, this is the ratio of the independence number of the Cayley graph corresponding to $\mathcal{H}$, compared to the total number of graphs on n vertices, hence the name. This can be a helpful definition, since the order of $D_{\mathcal{H}}(n)$ is often around the same as that of the total number of graphs. Alon [1] only considered the families $\mathcal{H}'$ that are isomorphism classes $\mathcal{H}' = \{H : L \cong H\}$, for some graphs L. For this section, the new proofs consider only the families $\mathcal{H} = \{H : L \subseteq H\}$, and therefore (most of) the new proofs will not work for determining an upper bound of $d_{\mathcal{H}'}(n)$ for the isomorphism classes.

Due to the difference in definition, Alon's [1] approach for giving a lower bound on the independence ratio with $\mathcal{H}' = \{H : L \cong H\}$ does not work for determining a lower bound for the classes $\mathcal{H} = \{H : L \subseteq H\}$. This discrepancy will be quickly discussed in Section 3.2.3.

We will now start with some theorems proven by Alon [1] in Section 3.2.1, and build on these theorems to get a result regarding the independence ratio for cycles of even length: $\mathcal{H} = \{H : C_{2k} \subseteq H\}$, in Section 3.2.2.

3.2.1. Upper bounds for the independence ratio

The next definition concerns a type of graph with an even number of edges, which will be helpful for Lemma 3.13.

Definition 3.12 (Copies of graphs on an independent set)**.** Let H' be a graph and let I be the vertices of an independent set of H'. A graph H is *made from two copies of H' on I*, say H_1 and H_2 (so $H_1 \cong H_2 \cong H'$), if

- $V(H) = V(H_1) \cup V(H_2)$ and $E(H) = E(H_1) \cup E(H_2)$,
- $V(H_1) \cap V(H_2) = I$ and $E(H_1) \cap E(H_2) = \varnothing$.

Particularly, $H \cong H_1 \oplus H_2$.

An example of a graph made from two copies of another graph can be seen in Figure 3.1. Definition 3.12 can be extended for any number of copies on the same independent set; the only vertices each of the copies share with each other is the same independent set I.

The following theorem gives an upper bound on $d_{\mathcal{H}}(n)$ regarding the class $\mathcal{H}$ of graphs containing a subgraph H, where H of the form as in Definition 3.12:

Lemma 3.13 ([1])**.** *For a graph H' with an independent set I, let H be the graph made from two copies of H' on I. Then for $\mathcal{H} = \{G : H \cong G\}$ we have*

$$d_{\mathcal{H}}(n) \leq \frac{1}{\left\lfloor \frac{n-|I|}{|V(H')|-|I|} \right\rfloor} = O\left(n^{-1}\right).$$

Proof. Write down the number of vertices of H' as $a+b$ where $a = |I|$ and $b = |V(H')| - |I|$. The number of vertices of m copies of H' on I, is equal to

$$\left|\bigcup_{i=1}^{m} V(H_i')\right| = \sum_{i=1}^{m} |V(H_i')| - (m-1)|I| = m(a+b) - (m-1)a = a + mb,$$

since we are counting the vertices in the independent set m times in the sum. Consider now the vertex set of H: $[n] = \{1, 2, \dots, n\}$. The maximum number of copies of H' on I we can fit in $[n]$ is hence equal to $m = \lfloor (n-a)/b \rfloor$.

To this effect, fit m copies of H' on I in the index set $[n]$ and let $\mathcal{F}$ be the set consisting of these m copies. WLOG, each copy H_i' shares the independent set I on the last a vertices $\{n-a+1, n-a+2, \dots, n\}$, and each copy H_i' for $1 \le i \le m$ has its non-shared vertices on the vertex set $\{(i-1)b+1, (i-1)b+2, \dots, ib\}$.

Since for any two i, j with $i \ne j$ we have that the copies H_i', H_j' have the property that $H \cong H_i' \oplus H_j'$, the set $\mathcal{F}$ forms a clique in the Cayley graph associated with the $\mathcal{H}$-code. Therefore by Lemma 3.3, we get

$$M_{\{H\}}(n) \ge m \implies D_{\{H\}}(n) \le \frac{2^{\binom{n}{2}}}{m} \implies d_{\{H\}}(n) \le \frac{1}{m} = \frac{1}{\lfloor \frac{n-a}{b} \rfloor} = \frac{1}{\lfloor \frac{n-|I|}{|V(H')|-|I|} \rfloor}.$$

□

Note that this lemma is only applicable when $n \ge |V(H)|$. As a consequence of Lemma 3.13, we can give an upper bound for cycles C_{2k} of even length, where $k \in \mathbb{Z}_{\ge 2}$. Any even cycle of length $2k$ can be seen as being made of two copies of the path graph on $(k+1)$ vertices P' and P'', on the independent set being the two degree 1 vertices of the path, as shown in Figure 3.1. Please note that in this figure, the vertices $k+2$ up to $2k$ that are drawn in P'' do not appear in P'. This is for readability; these vertices are in P' and should be thought of as isolated vertices in the graph. The same goes for the vertices 2 up to k which are in P' but not in P''.

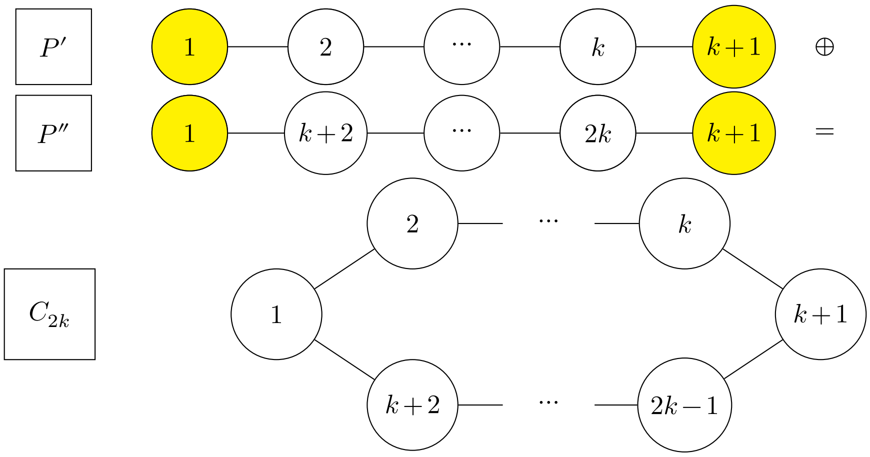


Figure 3.1: Two paths P' and P'' of length $k+1$, with the independent set being the two outer vertices, labeled in red. ”Glueing” the two paths on the red vertices gives the $2k$ cycle. All graphs here are graphs on the same n vertices, isolated vertices are omitted.

From this we see, by Lemma 3.13:

Corollary 3.14. *For all $k \in \mathbb{Z}_{\ge 2}$*

$$d_{\{H: C_{2k} \cong H\}}(n) \le \frac{1}{\lfloor \frac{n-2}{(k+1)-2} \rfloor} = \frac{1}{\lfloor \frac{n-2}{k-1} \rfloor} = O_k(n^{-1}).$$

For any prime k, we can give a slightly better upper bound for $d_{\{H: C_{2k} \subseteq H\}}(n)$ than $O(n^{-1})$. Before giving this bound, we will give context and go through a proof given in Alon's 2023 paper ”Graph Codes” [1]. The proof in question concerns an upper bound given to the independence ratio for when $\mathcal{H}$ is the collection of graphs isomorphic to stars with an even number of edges $K_{1,2k}$. A direct upper bound for this independence ratio was proven for any prime k using a modular version of the Frankl-Wilson Theorem, see Theorem 3.15, and for an arbitrary k an asymptotic bound was proven using a result from Frankl and Füredi, see Theorem 3.16.

Theorem 3.15 (Frankl-Wilson 1981 ([16]))**.** *Let p be a prime, and let L be a list of r distinct residue classes modulo p. Let $\mathcal{F}$ be a family of subsets of $[n]$ and suppose that $|F| \notin L \pmod{p}$ for all $F \in \mathcal{F}$ and that for every two distinct $F_1, F_2 \in \mathcal{F}$, $|F_1 \cap F_2| \in L \pmod{p}$. Then*

$$|\mathcal{F}| \le \sum_{i=0}^{r} \binom{n}{i}.$$

Theorem 3.16 (Frankl-Füredi 1985 ([9])). *Let ℓ be a fixed positive integer. For every fixed nonnegative integers ℓ_1, ℓ_2 such that $\ell > \ell_1 + \ell_2$ there exist constants $n_0 = n_0(\ell)$ and $d_\ell > 0$ so that for all $n > n_0$, if $\mathcal{F}$ is a family of ℓ-subsets of $[n]$ in which the intersection of each pair of distinct members is of cardinality either at least $\ell - \ell_1$ or strictly smaller than ℓ_2, then*

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

We will also need another theorem regarding the relationship between independence ratios of subgraphs within a vertex transitive graph. This result, Theorem 3.17, is well-known and part of the folklore when it comes to vertex transitive graphs.

Theorem 3.17. *Let G be a vertex transitive graph, and let $H \subseteq G$ be a subgraph of G, then*

$$\frac{\alpha(G)}{|V(G)|} \leq \frac{\alpha(H)}{|V(H)|}.$$

A proof of this theorem is explained in Appendix A, as the proof is lengthy and showing the proof here would distract from the logic of this section.

The theorem proven in Alon's paper, which is given with a more expanded proof, is the following:

Theorem 3.18 ([1]). *Let $\mathcal{H} = \{H : K_{1,2k} \cong H\}$. For prime k,*

$$d_{\mathcal{H}}(n) \leq \frac{\sum_{i=0}^{k-1} \binom{n-1}{i}}{\binom{n-1}{2k-1}}$$

and for all $k \in \mathbb{Z}_{\geq 2}$, for sufficiently large n, there exists a constant c_k such that

$$d_{\mathcal{H}}(n) \leq \frac{c_k \cdot (n-1)^{k-1}}{\binom{n-1}{2k-1}}, \quad \textit{i.e.} \quad d_{\mathcal{H}}(n) = O_k\left(n^{-k}\right).$$

Proof. In accordance to Theorem 3.17, we will bound the independence ratio of the Cayley graph associated with the $\mathcal{H}$-code by the independence ratio of some appropriate subgraph of the Cayley graph.

To this effect, consider, on the vertex set $[n]$, the collection $\mathcal{G}$ of all stars with its center at the vertex 1 and $(2k-1)$ leaves in the vertices $\{2, 3, \ldots, n\}$, with $k \in \mathbb{Z}_{\geq 2}$. Then $|\mathcal{G}| = \binom{n-1}{2k-1}$. Using Theorem 3.17 we get:

$$d_{\mathcal{H}}(n) \leq \frac{D_{\mathcal{H}}(n)}{2^{\binom{n}{2}}} \overset{\text{Theorem 3.17}}{\leq} \frac{\max_{\mathcal{F} \subseteq \mathcal{G} \text{ indep}} |\mathcal{F}|}{|\mathcal{G}|} \leq \frac{\max_{\mathcal{F} \subseteq \mathcal{G} \text{ indep}} |\mathcal{F}|}{\binom{n-1}{2k-1}}. \tag{3.6}$$

We now try to bound the size of an independent set $\mathcal{F} \subseteq \mathcal{G}$, using the Frankl-Wilson Theorem 3.15. If two stars from $\mathcal{G}$ intersect in exactly $k-1$ leaves, then the resulting symmetric difference of the two stars will form a star with its center at 1 and $2k$ leaves, see Figure 3.2.

Each star in $\mathcal{G}$ is represented by a $(2k-1)$-subset of $\{2, 3, \ldots, n\}$. For a collection $\mathcal{F} \subseteq \mathcal{G}$ of $(2k-1)$-subsets to be independent in the Cayley graph corresponding to $\mathcal{H}$, any two distinct $F_1, F_2 \in \mathcal{F}$ has an intersection of size unequal to $k-1$.

Hence each of the subsets of $\mathcal{F}$ is of cardinality $-1 \mod k$ and each intersection of distinct subsets of $\mathcal{F}$ is of cardinality unequal to $-1 \mod k$. For prime k we can now apply the Frankl-Wilson Theorem 3.15 to find $|\mathcal{F}| \leq \sum_{i=0}^{k-1} \binom{n-1}{i}$, and so by (3.6)

$$d_{\mathcal{H}}(n) \leq \frac{\sum_{i=0}^{k-1} \binom{n-1}{i}}{\binom{n-1}{2k-1}}. \tag{3.7}$$

For the non-prime k we can use the Frankl-Füredi Theorem 3.16. In this case $\ell = 2k-1$ and we take $\ell_1 = \ell_2 = k-1$. This choice is acceptable, since the cardinality of the intersection of two distinct $(2k-1)$-subsets in $\mathcal{F}$ is unequal to $k-1$, which will either be at least $k = 2k-1-(k-1) = \ell - \ell_1$, or strictly smaller than $k-1 = \ell_2$. The Frankl-Füredi Theorem therefore proves for sufficiently large n the

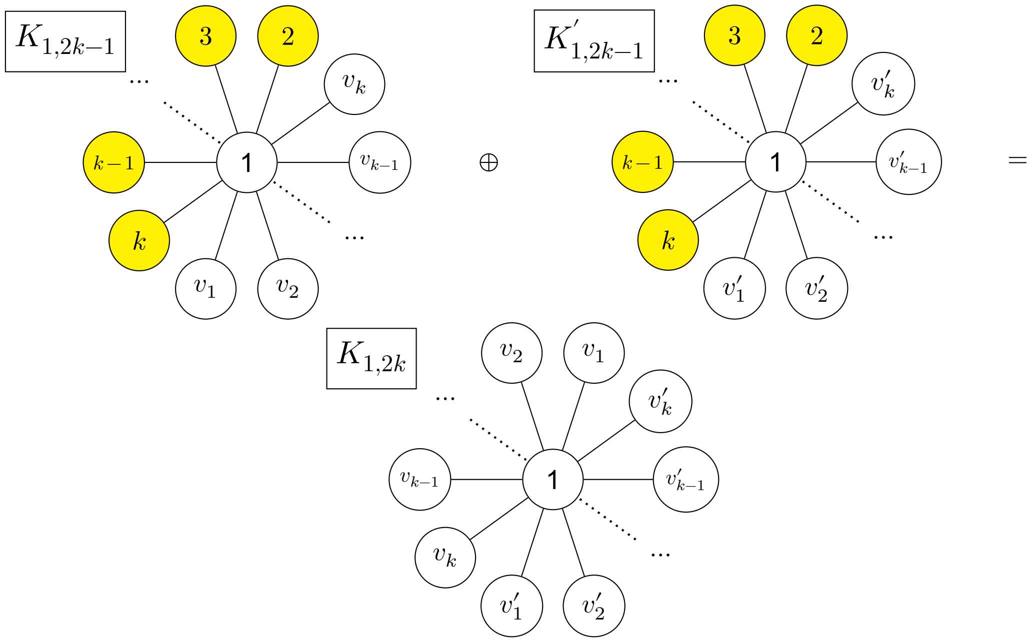


Figure 3.2: Two $(2k-1)$-stars with common center 1 intersecting in $k-1$ vertices, labeled in yellow, with their symmetric difference creating a $(2k)$-star. Similarly as with Figure 3.1, all stars here are graphs on n vertices, and the isolated vertices are omitted.

existence of a constant c_k such that $|\mathcal{F}| \le c_k \cdot (n-1)^{\max\{k-1,k-1\}} = c_k \cdot (n-1)^{k-1}$. Hence, in the same way as (3.6), we acquire

$$d_{\mathcal{H}}(n) \le \frac{\max_{\mathcal{F}\subseteq\mathcal{G} \text{ indep}} |\mathcal{F}|}{|\mathcal{G}|} \le \frac{c_k \cdot (n-1)^{k-1}}{\binom{n-1}{2k-1}}, \quad \text{i.e.} \quad d_{\mathcal{H}}(n) = O_k\left(n^{-k}\right).$$

□

The proofs for the upper bounds of the independence number by Alon presented so far concern the isomorphism classes $\mathcal{H} = \{H : L \cong H\}$. These bounds do still hold for the independence number of the classes $\mathcal{H} = \{H : L \subseteq H\}$ via Remark 3.19:

Remark 3.19. If two graphs G_1, G_2 on the vertex set $[n]$ have that $G_1 \oplus G_2$ contains no L as a subgraph, then $G_1 \oplus G_2$ is surely not isomorphic to L. Therefore any collection $\mathcal{F}$ valid for $D_{\mathcal{H}}(n)$ where $\mathcal{H} = \{H : L \subseteq H\}$, is also valid for $D_{\mathcal{H}'}(n)$ where $\mathcal{H}' = \{H : L \cong H\}$. Hence $D_{\mathcal{H}'}(n) \ge D_{\mathcal{H}}(n)$.

Therefore any lower bound on $D_{\mathcal{H}}(n)$ is a lower bound on $D_{\mathcal{H}'}(n)$, and similarly, any upper bound on $D_{\mathcal{H}'}(n)$ is an upper bound on $D_{\mathcal{H}}(n)$.

3.2.2. Upper bound for the independence ratio, for even cycles

We will use the idea of the proof of Theorem 3.18, to prove an upper bound to the independence ratio for the family $\mathcal{H} = \{H : C_{2k} \subseteq H\}$ of graphs containing cycles of length $2k$. This is a separate result from the results Alon [1] found since the families Alon considers are the families $\mathcal{H} = \{H : L \cong H\}$. This new result however improves the bound we found from Corollary 3.14.

Theorem 3.20. *For an integer a, denote $\mathcal{H}_{2a} = \{H : C_{2a} \subseteq H\}$. For all $k \in \mathbb{Z}_{\geq 2}$, there exists a constant c_k such that for sufficiently large n:*

$$d_{\mathcal{H}_{2k}}(n) \leq \frac{c_k \cdot n^{\max\{k-2,2\}}}{\binom{n}{k+1}} = O_k\left(n^{\max\{-3,-k+1\}}\right).$$

Particularly,

$$d_{\mathcal{H}_6}(n) = O\left(n^{-2}\right), \quad d_{\mathcal{H}_{2k}}(n) = O_k\left(n^{-3}\right) \quad \forall k \geq 4$$

Note: the upper bound Theorem 3.20 does not give a better upper bound for $d_{\mathcal{H}_4}(n)$ than Corollary 3.14, so it is left out here.

Proof. Like Theorem 3.18, we use Theorem 3.17 and bound $d_{\mathcal{H}_{2k}}(n)$ the independence ratio of some appropriate subgraph of the Cayley graph corresponding to $\mathcal{H}_{2k}$.

Consider the collection $\mathcal{G}$ of $(k+1)$-cliques on the vertex set $[n]$. The cardinality of $\mathcal{G}$ is equal to the number of combinations of $k+1$ elements from n elements, i.e, $|\mathcal{G}| = \binom{n}{k+1}$.

Similarly to equation (3.6) we get:

$$d_{\mathcal{H}_{2k}}(n) \leq \frac{D_{\mathcal{H}_{2k}}(n)}{2^{\binom{n}{2}}} \overset{\text{Theorem 3.17}}{\leq} \frac{\max_{\mathcal{F}\subseteq\mathcal{G} \text{ indep}} |\mathcal{F}|}{|\mathcal{G}|} \leq \frac{\max_{\mathcal{F}\subseteq\mathcal{G} \text{ indep}} |\mathcal{F}|}{\binom{n}{k+1}}. \tag{3.8}$$

We now want to bound the size of an independent set $\mathcal{F} \subseteq \mathcal{G}$ in the Cayley graph corresponding to $\mathcal{H}$, using the Frankl-Füredi Theorem 3.16. To be able to use this theorem, we will show that if two $(k+1)$-cliques on the same n vertices intersect in exactly 2 vertices, omitting the isolated vertices, that its symmetric difference contains a $2k$-cycle. For no other number of intersections will the symmetric difference of the two $(k+1)$-cliques contain a $2k$-cycle. Therefore, we may use the Frankl-Füredi Theorem 3.16 to bound an independent set $\mathcal{F} \subseteq \mathcal{G}$ by forbidding vertex intersections of size 2.

For two $(k+1)$-cliques K' and K'' on n vertices, omitting the isolated vertices, if they intersect in:

- 0 vertices, the symmetric difference $K' \oplus K''$ will be two disjoint $(k+1)$-cliques, and cannot contain a $2k$-cycle,
- 1 vertex, the symmetric difference $K' \oplus K''$ does not contain a $2k$-cycle, since we cannot cross the same vertex in a cycle.
- ≥ 3 vertices, then the number of vertices in the symmetric difference $K' \oplus K''$ that have edges connected to them will be less or equal than $(k+1)+(k+1)-3 = 2k-1$, and therefore the symmetric difference cannot contain a $2k$-cycle.

If K' and K'' intersect in 2 vertices, say the vertices a and b, a $2k$-cycle can be made as shown in Figure 3.3. More explicitly, let P' be a Hamiltonian path from a to b in K' and let P'' be a Hamiltonian path from b to a in K''. Since the symmetric difference of K' and K'' in this case only removes the intersecting edge ab of K' and K'', and P' and P'' do not contain the edge ab, stitching the Hamiltonian paths P' and P'' together forms a $2k$-cycle.

Therefore we have that $\mathcal{F} \subseteq \mathcal{G}$ is an independent set in the Cayley graph corresponding to $\mathcal{H}$, if and only if for every two $(k+1)$-cliques $K', K'' \in \mathcal{F}$ we have that K' and K'' do not intersect in exactly 2 vertices, omitting the isolated vertices.

Since every $(k+1)$-clique can be represented by an unordered $(k+1)$-subset of $[n]$, we can use the Frankl-Füredi Theorem to bound the number of unordered $(k+1)$-subsets of $[n]$ with forbidden intersection 2, which in turn bounds the cardinality of $\mathcal{F}$. Since for two unordered $(k+1)$-subsets we

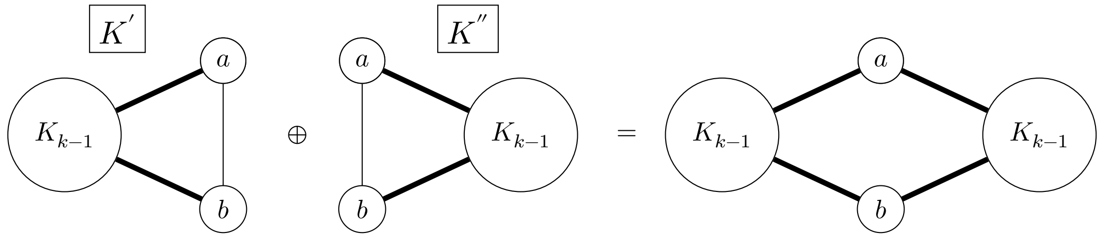


Figure 3.3: Two $(k+1)$-cliques intersecting only in the vertices a and b, with their symmetric difference creating a $(2k)$-cycle. The straight lines are edges, while the thick lines represent all edges from the vertices in K_{k-1} going to the vertices a and b. A $2k$-cycle in the symmetric difference is created by starting from a, then going to the left K_{k-1} and taking a path through every vertex, then going to b, then going to the right K_{k-1} and taking a path through every vertex there, and then returning to a. Similarly as with Figure 3.1, all graphs here are graphs on n vertices, and the isolated vertices are omitted.

are forbidding intersections of size 2, the size of an intersection is either at least $3 = \ell - (k-2)$, or strictly less than 2. Hence letting $\ell_1 = k-2$ and $\ell_2 = 2$ meets the requirements.

The Frankl-Füredi Theorem therefore proves for sufficiently large n the existence of a constant c_k such that $|\mathcal{F}'| \leq c_k \cdot n^{\max\{k-2,2\}}$. Hence we acquire using equation (3.8):

$$d_{\mathcal{H}_{2k}}(n) \leq \frac{\max_{\mathcal{F}\subseteq\mathcal{G} \text{ indep}} |\mathcal{F}|}{\binom{n}{k+1}} \leq \frac{c_k \cdot n^{\max\{k-2,2\}}}{\binom{n}{k+1}} = O_k\left(n^{\max\{-3,-k+1\}}\right).$$

For $k=3$ this will give an order of $O(n^{-2})$ and for any $k \geq 4$ we get an order of $O_k(n^{-3})$, which was to be demonstrated. □

One difference between the proof of Theorem 3.18 and the proof of Theorem 3.20, is that the Frankl-Wilson Theorem is not used. As for the reason the Frankl-Wilson Theorem was not used: The number of intersections for two $(k+1)$-cliques has to be exactly 2 for a $2k$-cycle to be in the symmetric difference. Therefore, an independent set $\mathcal{F}$ in $\mathcal{G}$ would be a collection of subsets of size $k+1 \equiv 1 \mod k$ where each intersection is unequal to $2 \mod k$. The conditions for the Frankl-Wilson Theorem are not satisfied this way.

One way to perhaps generally improve the bound is to consider, instead of cliques of size $k+1$, larger cliques of size $k+m$ for some $m > 1$. If two $(k+m)$-cliques on the vertex set $[n]$ were to intersect in, say, s vertices, omitting the isolated vertices, then the resulting symmetric difference is of the form described in Figure 3.4.

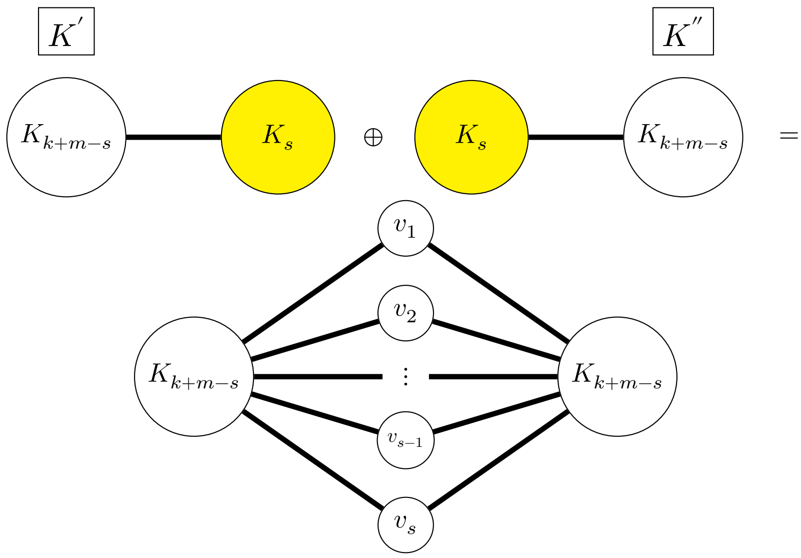


Figure 3.4: Two $(k+m)$-cliques intersecting in s vertices $v_1, v_2, \ldots, v_s$. Thick lines represent every edge going from a vertex in a clique to a single vertex, or to a vertex in another clique, depending on the graph K', K'' or $K' \oplus K''$ in context. Similarly as with Figure 3.1, all graphs here are graphs on n vertices, and the isolated vertices are omitted.

We give the following remark, which uses the same ideas as in the proof of Theorem 3.20:

Remark 3.21. If K' and K'' are two $(k+m)$-cliques as in Figure 3.4, then we have the following:

- If $s=0$, we have two disjoint $(k+m)$-cliques which cannot contain a $2k$-cycle in the symmetric difference if $k+m<2k$.
- If $s=1$, since a cycle cannot pass over the same vertex, this cannot contain a $2k$-cycle either if $k+m<2k$.
- If $s>2m$, then the total number of vertices in the symmetric difference which have edges connected to them is equal to $(k+m-s)+(k-m+s)+s=2k-2m-s<2k$.

Therefore only the values of s for which a $2k$-cycle could be contained in $K'\oplus K''$ if $k+m<2k$, are $2\le s\le 2m$.

This remark leads to the following general theorem.

Theorem 3.22. *For an integer a, denote $\mathcal{H}_{2a}=\{H : C_{2a}\subseteq H\}$. For any $m\ge 1$ and $k\in\mathbb{Z}_{\ge 2m}$ such that $k+m\le n$, there exists a constant c_k such that for sufficiently large n:*

$$d_{\mathcal{H}_{2k}}(n)\le \frac{c_k\cdot n^{\max\{k-m-1,2\}}}{\binom{n}{k+m}}=O_k\left(n^{\max\{-2m-1,-k-m+2\}}\right).$$

Proof. This proof follows the same general idea of Theorem 3.20. We bound the independence ratio $d_{\mathcal{H}_{2k}}(n)$ by considering only a subgraph of the Cayley graph corresponding to $\mathcal{H}_{2k}$ and using Theorem 3.17. To this effect, let $\mathcal{G}$ be the collection of $(k+m)$-cliques on the vertex set $[n]$. Then $|\mathcal{G}|=\binom{n}{k+m}$. We will show that the intersections of two $K',K''\in\mathcal{G}$ of size $2\le s\le 2m$ result in a $2k$-cycle contained in $K'\oplus K''$. Together with Remark 3.21 this allows for the Frankl-Füredi Theorem 3.16 to be applied to an independent set $\mathcal{F}\subseteq\mathcal{G}$ in the Cayley graph. To show this we split in two cases:

- s is even: Consider the symmetric difference from Figure 3.4. We can draw the following cycle given in Figure 3.5.

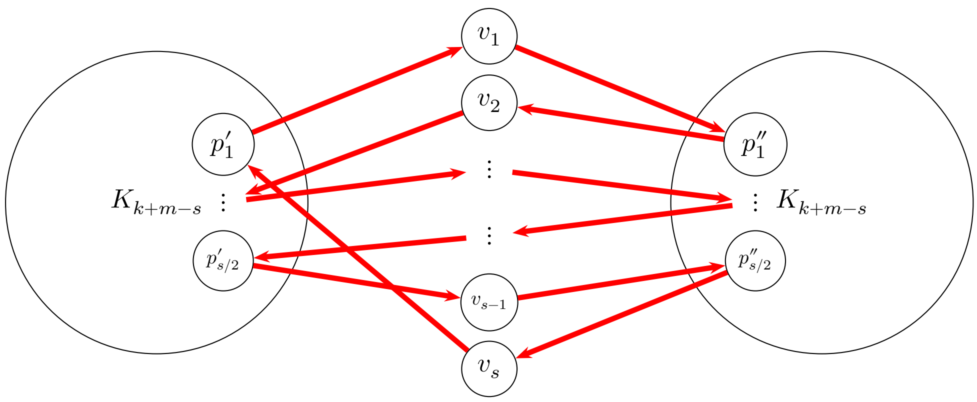


Figure 3.5: The symmetric difference from Figure 3.4 for an even number of intersections s. A $2s$-cycle is given in red arrows, which can be padded using the leftover vertices from both K_{k+m-s}. Other edges have not been drawn for readability

This cycle is of length $2s$, but we can always make the cycle bigger by adding the leftover vertices from both K_{k+m-s}. This gives a cycle of length $(k+m-s)+(k+m-s)+s=2k+2m-s$. We have $2k+2m-s\ge 2k \iff s\le 2m$, hence we can always construct a $2k$-cycle this way if $s\le 2m$.

- s is odd: Consider again the symmetric difference from Figure 3.4. We can draw the following cycle given in Figure 3.6, which is essentially Figure 3.5 only considering the vertices v_1 up until v_{s-1}.

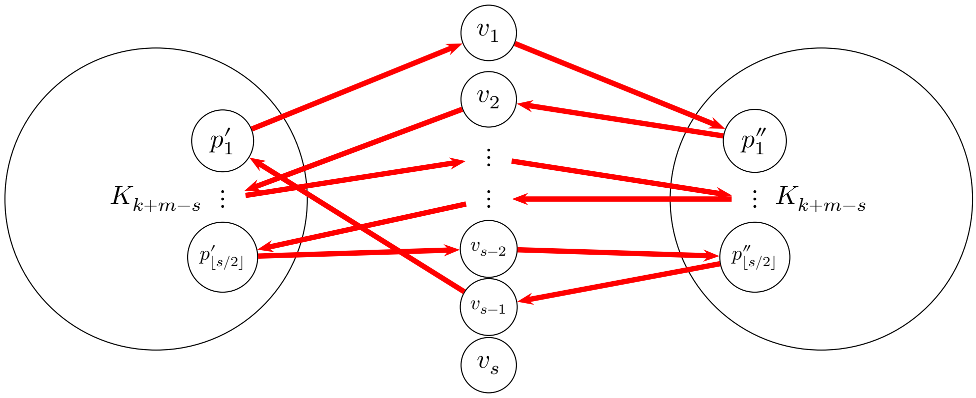


Figure 3.6: The symmetric difference from Figure 3.4 for an even number of intersections s. A $2(s-1)$-cycle is given in red arrows, which can be padded using the leftover vertices from both K_{k+m-s}. Other edges have not been drawn for readability

This cycle is of length $2(s-1)$, but similarly to the even case we can always make the cycle bigger by adding the leftover vertices from both K_{k+m-s}. This gives a cycle of length $(k+m-s)+(k+m-s)+s-1 = 2k+2m-s-1$. We have $2k+2m-s-1 \geq 2k \iff s \leq 2m-1$, hence we can always construct a $2k$-cycle this way if $s \leq 2m-1$.

This covers every single intersection $2 \leq s \leq 2m$. We note that both these back-and-forth constructions can only exist if there are enough vertices in the clique K_{k+m-s} to go back-and-forth to. This requires $k+m-s \geq \lfloor s/2 \rfloor$ or $k+m \geq s + \lfloor s/2 \rfloor$. In the case of $s = 2m$ this gives $k+m \geq 3m$ or $k \geq 2m$.

An independent set $\mathcal{F} \subseteq \mathcal{G}$ in the Cayley graph must therefore have that no two graphs $K', K'' \in \mathcal{F}$ have intersection $2 \leq s \leq 2m$.

Since every $(k+m)$-clique may be represented uniquely by $(k+m)$-subsets of $[n]$, we may use the Frankl-Füredi Theorem 3.16 with $\ell_1 = k+m-(2m+1) = k-m-1$ and $\ell_2 = 2$. This gives rise for sufficiently large n to the existence of a constant c_k such that $|\mathcal{F}| \leq c_k \cdot n^{\max\{k+m-1,2\}}$, and therefore we have similarly to equation (3.8):

$$d_{\mathcal{H}_{2k}}(n) \leq \frac{\max_{\mathcal{F} \subseteq \mathcal{G} \text{ indep}} |\mathcal{F}|}{|\mathcal{G}|} \leq \frac{c_k \cdot n^{\max\{k+m-1,2\}}}{\binom{n}{k+m}} = O_k\left(n^{\max\{-2m-1,-k-m+2\}}\right).$$

□

Theorem 3.22 applies for $k \geq 2m$, hence we have $d_{\mathcal{H}_{2k}}(n) = O_k(n^{-2m-1})$ if $m > 3$, as then

$$-2m-1 > -3m+2 \geq -k-m+2.$$

Furthermore, if for a fixed k we then also set m such that $k+m = n$, then we have

$$d_{\mathcal{H}_{2k}}(n) \leq O_k(n^{-2(n-k)-1}) \leq O_k(n^{-2n+2k-1}).$$

This is upper bound is quite good, yet it pales in comparison to a knwon bound. Morris and Saxton [14] proved that a collection of graphs on n vertices that does not contain a C_{2k}, is of cardinality at most $2^{O(n^{1+1/k})}$. Now let $\mathcal{F}$ be an independent set in the Cayley graph corresponding to $\mathcal{H}_{2k}$, and let $G_0 \in \mathcal{F}$ be a fixed graph. Then f defined by $f(G) = G \oplus G_0$ on the domain $\mathcal{F}$ is an injective function from $\mathcal{F}$ to the collection of C_{2k}-free graphs. Therefore $\mathcal{F}$ has cardinality at most $|\mathcal{F}| \leq 2^{O(n^{1+1/k})}$. Hence we find:

Corollary 3.23.

$$D_{\mathcal{H}_{2k}}(n) \leq 2^{O(n^{1+1/k})} \implies d_{\mathcal{H}_{2k}}(n) \leq 2^{-\binom{n}{2}+O(n^{1+1/k})}.$$

Furthermore, Erdős [8] proved that $\text{ex}(n, C_{2k}) = O(n^{1+1/k})$. Combined with Lemma 3.9, Corollary 3.23 gives equality for both $D_{\mathcal{H}_{2k}}(n)$ and $d_{\mathcal{H}_{2k}}(n)$. This big-O bound decays super-exponentially, much faster than the $O(n^{-2n})$ bound from Theorem 3.22.

3.2.3. Lower bounds for the independence ratio

As mentioned at the start of this section 3.2, Alon [1] describes an approach to give a lower bound for the independence number. In the paper, the definition $\mathcal{H}' = \{H : L \cong H\}$ is used, where L is a graph of choice. Since the new Theorems 3.20 and 3.22 use that $\mathcal{H} = \{H : L \subseteq H\}$, lower bounds for the independence ratio $d_{H'}(n)$ proven by Alon [1] do not give lower bounds for the independence ratio $d_H(n)$ by Remark 3.19. We will however describe the Alon's idea shortly.

In the proof of Theorem 3.17, see Appendix A, it was proven that for vertex transitive graphs G we have that $\alpha(G)\chi_f(G) = |V(G)|$, and hence the independence ratio for vertex transitive graphs G is equal to the reciprocal of its fractional chromatic number $\chi_f(G)$. Therefore, an upper bound for the fractional chromatic number would be sufficient as a lower bound for the independence number. The chromatic number $\chi(G)$ is an upper bound on the fractional chromatic number, since the LP of the fractional chromatic number is a relaxation of the ILP of the chromatic number. Hence, if we let G be the Cayley graph corresponding to $\mathcal{H}$,

$$d_{\mathcal{H}}(n) = \frac{D_{\mathcal{H}}(n)}{2^{\binom{n}{2}}} = \frac{\alpha(G)}{|V(G)|} = \frac{1}{\chi_f(G)} \geq \frac{1}{\chi(G)}. \tag{3.9}$$

To this effect, consider the following: suppose that each edge e in K_n, we have assigned a vector $v_e \in \mathbb{Z}_2^r$ for some positive integer r such that for any $H \in \mathcal{H}'$ we have that the sum of edge-vectors of H is not the all-zero vector, i.e. $\sum_{e\in E(H)} v_e \neq \mathbf{0}\ \forall H \in \mathcal{H}'$.

We now assign a color $c : G[n] \to \mathbb{Z}_2^r$ to every graph-vertex G_i in the Cayley graph, via $c(G_i) = \sum_{e\in E(G_i)} v_e$. This is a valid 2^r-coloring of the Cayley graph. Indeed two graphs G_1, G_2 forms an edge in the Cayley graph if $G_1 \oplus G_2 \in \mathcal{H}'$. Hence if G_1 and G_2 have an edge in the Cayley graph, then for some $H \in \mathcal{H}'$ we get

$$c(G_1) = \sum_{e\in E(G_1)} v_e = \sum_{e\in E(G_2)} v_e + \sum_{e\in E(H)} v_e \neq \sum_{e\in E(G_2)} v_e + \mathbf{0} = c(G_2).$$

The method described here works for any family $\mathcal{H}$, not just the specific family $\mathcal{H}' = \{H : L \cong H\}$. The main problem arises when we try to assign vectors v_e to edges e of K_n with the requirement that $\sum_{e\in E(H)} v_e \neq \mathbf{0}$. As an example, here is a Theorem from Alon [1] which proves the tightness of the upper bound of the independence ratio for stars, as in Theorem 3.18.

Theorem 3.24 ([1]). *Let $\mathcal{H} = \{H : K_{1,2k} \cong H\}$. For all $k \in \mathbb{Z}_{\geq 1}$ we have*

$$d_{\mathcal{H}}(n) = \Omega_k\left(\frac{1}{n^k}\right).$$

Proof. Let s be the smallest integer so that $2^s - 1 \geq n$, so $2^{s-1} - 1 \leq n$. The columns of the parity check matrix of a BCH-code with minimum Hamming distance $2k+1$, show that there is a collection S of $2^s - 1$ binary vectors of length $r = ks$ so that no sum of at most $2k$ of them is the zero vector. Let c be a proper n-coloring of K_n. For each edge e let v_e be the $c(e)$'th element of S. Now for every graph with at most $2k$ edges, particularly all graphs H which are stars with $2k$ edges, we have $\sum_{e\in E(H)} v_e \neq \mathbf{0}$.

Therefore, if G is the Cayley graph corresponding to $\mathcal{H}$, $f(G) = \sum_{e\in E(G)} v_e$ is a proper 2^r-coloring of the Cayley graph. Therefore $\chi(G) \leq 2^r = 2^{ks} = (2^{s-1})^k 2^k \leq (n+1)^k \cdot 2^k$, and hence from equation (3.9) we get

$$d_{\mathcal{H}}(n) \geq \frac{1}{\chi(G)} \geq \frac{1}{(n+1)^k 2^k} = \Omega_k\left(\frac{1}{n^k}\right).$$

This gives the desired lower bound for stars. □

The part where this proof breaks down if we were working with the family $\mathcal{H} = \{H : K_{1,2k} \subseteq H\}$ is that now there are graphs $H \in \mathcal{H}$ with far more than $2k$ edges, e.g. the complete graph K_n contains a $K_{1,2k}$. Therefore, for these graphs H, $\sum_{e\in E(H)} v_e$ might be the all-zero vector $\mathbf{0}$, which would make f an improper coloring, breaking the proof. A different construction, or proof, is needed to adjust the proof to the family $\mathcal{H}$.

4

Bounds on the independence number, with the use of eigenvalues

This chapter attempts to find upper bounds for $D_{\mathcal{H}}(n)$. By realizing that $D_{\mathcal{H}}(n)$ is the independence number of the Cayley graph associated with the $\mathcal{H}$-code, we can use upper bounds that are known for the independence number. Two bounds will be used, as in Section 4.1, and also proven as the literature likes to use the bounds, but never give the proof. These upper bounds use the eigenvalues of that Cayley graph. The way these eigenvalues are found using character sums. The theory behind characters, and how to use these to find eigenvalues, will be explained in Section 4.2. Finally, we compare the used bounds against known values in Section 4.3

4.1. Some upper bounds on the independence number, using eigenvalues

Two bounds for the independence number α will be presented here, both of which require us to know the eigenvalues of the Cayley graph associated with the $\mathcal{H}$-code, which was presented in the proof of Lemma 3.3. Both bounds will be proven here since the literature for the proofs of the bounds, especially for the Cvetković bound, is mostly lacking. For the following theorem we denote $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m$ as the eigenvalues of a graph G on m vertices in descending order.

Theorem 4.1 (Hoffman bound)**.** *For any d-regular graph G on m vertices we have:*

$$\alpha(G) \leq \frac{-m \cdot \lambda_m}{d - \lambda_m}.$$

Proof. Since G is d-regular, we know for its adjacency matrix A that $A\mathbf{1} = d\mathbf{1}$, and one can prove, for instance with Gerschgorin's circle theorem, that d is the largest eigenvalue λ_1 of A. We also get that A and the all-ones $m \times m$ matrix $J_m = \mathbf{1}\mathbf{1}^T$ share the all ones vector $\mathbf{1}$ as an eigenvector, and therefore share a basis of eigenvectors. One of which is the all one vector $\mathbf{1}$, and $(m-1)$ of which, label them $u_2, ..., u_m$, are orthogonal to $\mathbf{1}$. That is, for $2 \leq i \leq m$ we have $Au_i = \lambda_i u_i$ and $Ju_i = \mathbf{1}\mathbf{1}^T u_i = \mathbf{1} \cdot 0 = 0$.

Now consider the matrix $D = A - \frac{d-\lambda_m}{m} J_m$. The matrix D has the exact same eigenvectors as A and J_m, and we find

$$D\mathbf{1} = A\mathbf{1} - \frac{d-\lambda_m}{m} J_m \mathbf{1} = d\mathbf{1} - \frac{d-\lambda_m}{m} \cdot m\mathbf{1} = d\mathbf{1} - (d-\lambda_m)\mathbf{1} = \lambda_m \mathbf{1},$$

and for $2 \leq i \leq m$ we have

$$Du_i = Au_i - \frac{d-\lambda_m}{m} J_m u_i = \lambda_i u_i + \frac{d-\lambda_m}{m} \cdot 0 = \lambda_i u_i.$$

We see that the smallest eigenvalue of D is λ_m, and therefore the matrix $E = D - \lambda_m I_m$ has a smallest eigenvalue of 0 and is therefore positive semi-definite. Let I be a set of vertices which form a maximum independent set in A, so $|I| = \alpha(G)$. Now consider the principal submatrix $E_{(I,I)}$ of E considering only the rows and columns of E which belong to the vertices of I. This principal submatrix is positive semidefinite because E is. Combined with the fact that $A_{(I,I)}$ is an all zero matrix, we find that the eigenvalues of

$$E_{(I,I)} = D_{(I,I)} - \lambda_m I_{m,(I,I)} = -\frac{d-\lambda_m}{m} J_{m,(I,I)} - \lambda_m I_{m,(I,I)}$$

are positive. Specifically, for the all-ones vector on I we get:

$$E_{(I,I)}\mathbf{1}_I = -\frac{d-\lambda_m}{m} J_{m,(I,I)}\mathbf{1}_I - \lambda_m I_{m,(I,I)}\mathbf{1}_I = -\frac{d-\lambda_m}{m} \cdot |I|\mathbf{1}_I - \lambda_m \mathbf{1}_I = \left(-\frac{d-\lambda_m}{m} \cdot \alpha(G) - \lambda_m\right)\mathbf{1}_I,$$

and so finally

$$-\frac{d-\lambda_m}{m} \cdot \alpha(G) - \lambda_m \geq 0 \implies \alpha(G) \leq \frac{-m\lambda_m}{d-\lambda_m}.$$

□

For the Cayley graph belonging to the collection $\mathcal{H}$, see Remark 3.8, we can apply the Hoffman bound as the Cayley graph is vertex-transitive, see Lemma 3.3, and therefore regular. The eigenvalues can be calculated, but that will be postponed until section 4.2. The only other parameter that is not immediately apparent is the value of d, the degree of the Cayley graph. To calculate the degree of the Cayley graph, we can use that the graph is regular. Hence, we only need to consider for example the number of edges adjacent to the vertex in the Cayley graph belonging to the empty graph $\overline{K_n}$. Namely, the degree would be equal to the how many graphs G exist such that $\overline{K_n} \oplus G = G$ is an element of $\mathcal{H}$. That is, the degree of the Cayley graph is precisely the cardinality of $\mathcal{H}$.

For the case where $\mathcal{H}$ is the local graph class $\{H' : H \subseteq H'\}$ for some graph H, we get that the the degree d is the number of graphs which contain H. This is equivalent to the total number of graphs minus the total number of graphs which are H-free.

There is an easy lower bound to the number of $\mathcal{H}$-free graphs, which uses the same idea of the proof of Lemma 3.9: Let G be a maximum edge $\mathcal{H}$-free graph, then $|E(G)| = \mathrm{ex}(n,\mathcal{H})$ and each of its subgraphs is also $\mathcal{H}$-free. Therefore the number of $\mathcal{H}$-free graphs is at least $2^{\mathrm{ex}(n,\mathcal{H})}$.

No exact upper bound for the number of $\mathcal{H}$-free graphs is known, and for some cases only asymptotic results are known. For instance, for the case when $\mathcal{H} = \{H : K_3 \subseteq H\}$, Erdős, Kleitman and Rothschild [15] found that the upper bound for the number of triangle free graphs is of the same order as the lower bound, namely $2^{\lfloor n^2/4 \rfloor + o(n^2)}$. This gives us both bounds for the degree d:

$$2^{\binom{n}{2}} - 2^{\lfloor n^2/4 \rfloor} \geq d \geq 2^{\binom{n}{2}} - 2^{\lfloor n^2/4 \rfloor + o(n^2)}.$$

This is useful information, and could be expanded upon. For now we will leave it as is, because the eigenvalues have to be calculated manually.

We continue with another bound that requires knowledge of every eigenvalue of a graph G on m vertices, opposed to just the smallest eigenvalue. We denote m_{R0} as the number of eigenvalues of G with the relation R to 0 where $R \in \{<,>,=,\leq,\geq\}$; e.g. $m_{>0}$ is the number of eigenvalues bigger than 0.

Theorem 4.2 (Cvetković bound)**.** *For any graph G (on m vertices), we have*

$$\alpha(G) \leq \min\{m_{>0}, m_{<0}\} + m_{=0}$$

Proof. Let I be the vertices of a maximum independent set in G. For the adjacency matrix A of G, we know then that the principal submatrix $A_{(I,I)}$ is the zero matrix, and thus $A_{(I,I)}$ has $\alpha(G)$ eigenvalues equal to 0.

We then use Cauchy's interlacing theorem, and find that for the eigenvalues $(\lambda_k)_k$ of G for $i \in \{1,2,\ldots,\alpha(G)\}$:

$$\lambda_i \geq 0 \geq \lambda_{m-\alpha(G)+i}.$$

Therefore, there are at least $\alpha(G)$ eigenvalues that are greater or equal to 0, and similarly there are at least $\alpha(G)$ eigenvalues that are smaller or equal to 0. Hence we get

$$\min\{m_{\geq 0}, m_{\leq 0}\} \geq \alpha(G) \iff \alpha(G) \leq \min\{m_{>0}, m_{<0}\} + m_{=0}.$$

□

We will now proceed with the calculation of the eigenvalues.

4.2. Finding eigenvalues of the Cayley graph using character sums

From Remark 3.8, we know that $G_{\mathcal{H}}$ from the proof of Lemma 3.3 is a Cayley graph $Cay(\Gamma, S)$ on the group $\Gamma = \mathbb{Z}_2^{\binom{n}{2}}$ with bitwise-XOR $\oplus$ as the operator, and the generating set S is equal to the set of binary edge representations of all graphs in $\mathcal{H}$. Since $G_{\mathcal{H}}$ is a Cayley graph, the eigenvalues can be expressed using character-sums. For this we will diverge from the story to explain characters.

4.2.1. Background information on characters

We start off with the general definition of characters and build up to Theorem 4.7, which, combined with Theorem 4.8, allows for quick calculation of all the eigenvalues of the Cayley graph. This information is part of the folklore of spectral graph theory, and can be found in a book on the topic.

Definition 4.3. For a group Γ, a function $\chi : \Gamma \to \mathbb{C}$ is called a **character** of Γ if χ is a group homomorphism of the multiplicative group $\mathbb{C}^*$, i.e. for any $a, b \in \Gamma$, we have $\chi(a+b) = \chi(a)\chi(b)$.

For any character χ it follows from the definition that $\chi(0) = \chi(0+0) = \chi(0)\chi(0)$, which implies $\chi(0) = 1$. For finite groups Γ we have the added property that every element a of Γ has a finite order, and that order must divide the size of the group $|\Gamma|$. Hence we get that since $|\Gamma| \cdot a = 0$, we also must have

$$1 = \chi(0) = \chi(|\Gamma| \cdot a) = \chi(a+a+\ldots+a) = \chi(a)\chi(a)\cdots\chi(a) = \chi(a)^{|\Gamma|},$$

and therefore each $\chi(a)$ is an $|\Gamma|$-th root of unity. Therefore characters are typically of the form $\chi(a) = e^{2\pi i a/|\Gamma|}$ or some power of it.

The function $\chi : \Gamma \to \mathbb{C}$ that maps everything to 1 is a character for any group Γ. This specific character is called the **principal character** and is denoted by χ_0. For any character which is not χ_0, we find the following property:

Lemma 4.4. *For any group Γ, if χ is a character of Γ which is not χ_0, then $\sum_{a\in\Gamma} \chi(a) = 0$.*

Proof. Since $\chi \neq \chi_0$, there exists a $b \in \Gamma$ such that $\chi(b) \neq 1$. Then:

$$\chi(b)\sum_{a\in\Gamma}\chi(a) = \sum_{a\in\Gamma}\chi(a+b) = \sum_{c\in\Gamma}\chi(c) = \sum_{a\in\Gamma}\chi(a).$$

Therefore $(\chi(b)-1)\sum_{a\in\Gamma}\chi(a) = 0$, and since $\chi(b) \neq 1$ we have $\sum_{a\in\Gamma}\chi(a) = 0$. □

Using the complex inner product $\langle \chi_1, \chi_2 \rangle = \sum_{a\in\Gamma} \chi_1(a)\overline{\chi_2(a)}$, Lemma 4.4 is used to prove the following lemma:

Lemma 4.5. *For any group Γ, if χ_1, χ_2 are two distinct characters of Γ, then $\langle \chi_1, \chi_2 \rangle = 0$.*

Proof. Note that $\chi = \chi_1 \cdot \overline{\chi_2}$ is a character of Γ, as χ_1 and χ_2 are and $\overline{\chi_2(a+b)} = \overline{\chi_2(a)\chi_2(b)} = \overline{\chi_2(a)} \cdot \overline{\chi_2(b)}$. Since $\chi_1 \neq \chi_2$, there is a $b \in \Gamma$ such that $\chi_1(b) \neq \chi_2(b)$. Therefore $\chi(b) = \chi_1(b) \cdot \overline{\chi_2(b)} \neq \chi_2(b) \cdot \overline{\chi_2(b)} = 1$. Therefore χ is not the principal character and from Lemma 4.4 we get

$$\langle \chi_1, \chi_2 \rangle = \sum_{a\in\Gamma} \chi_1(a) \cdot \overline{\chi_2(a)} = \sum_{a\in\Gamma} \chi(a) = 0.$$

□

Lemma 4.5 tells us that the characters of Γ are orthogonal to the inner product and therefore linearly independent. Hence the set of all characters of Γ forms a linear basis of $\mathbb{C}^{\Gamma}$. Particularly, for a finite group Γ, the set of all characters of Γ forms a $|\Gamma|$-dimensional space. Hence:

Corollary 4.6. *For any finite group Γ, it has at most $|\Gamma|$ characters.*

This gives an upper bound on the number of characters of Γ if Γ is finite. A more powerful statement can be made when Γ is also abelian, and its result will be used later:

Theorem 4.7. *For any finite abelian group Γ, it has exactly $|\Gamma|$ characters.*

Proof. Any finite abelian group Γ is isomorphic to a finite product of cyclic groups $\bigtimes_{i\in[k]} \mathbb{Z}/n_i\mathbb{Z}$. We now prove two points:

- $\mathbb{Z}/n\mathbb{Z}$ has n characters. Indeed, for any $x \in \{0,1,\dots,n-1\}$ consider $\chi_x(a) = e^{2\pi i x a/n}$. We get χ is a character of $\mathbb{Z}/n\mathbb{Z}$ for all x, as for any $a,b \in \mathbb{Z}/n\mathbb{Z}$ we have

$$\chi_x(a+b \mod n) = e^{2\pi i x(a+b+pn)/n} = e^{2\pi i x a/n} \cdot e^{2\pi i x b/n} \cdot e^{2\pi i x p} = \chi_x(a) \cdot \chi_x(b) \cdot 1.$$

  Furthermore, $\chi_x(1)$ gives different values for all $x \in \{0,1,\dots,n-1\}$, hence $\mathbb{Z}/n\mathbb{Z}$ has at least n characters, and no more due to Corollary 4.6.

- If Γ_1, Γ_2 are two finite abelian groups with $|\Gamma_1|$ and $|\Gamma_2|$ characters, then $\Gamma_1 \times \Gamma_2$ has exactly $|\Gamma_1| \cdot |\Gamma_2|$ characters. Indeed, if χ_1, χ_1' are two distinct characters of Γ_1 and χ_2, χ_2' are two distinct characters of Γ_2, then $\chi(a,b) = \chi_1(a) \cdot \chi_2(b)$ and $\chi'(a,b) = \chi_1'(a) \cdot \chi_2'(b)$ are two distinct characters of $\Gamma_1 \times \Gamma_2$. It follows that both χ and χ' are characters as χ_1, χ_1' and χ_2, χ_2' are characters, respectively. To show they are distinct: in one case we have an $a \in \Gamma_1$ such that $\chi_1(a) \neq \chi_1'(a)$, and so

$$\chi(a,0) = \chi_1(a) \cdot 1 \neq \chi_1'(a) \cdot 1 = \chi'(a,0),$$

  and in the other case we have a $b \in \Gamma_2$ such that $\chi_2(b) \neq \chi_2'(b)$, and so

$$\chi(0,b) = 1 \cdot \chi_2(b) \neq 1 \cdot \chi_2'(b) = \chi'(0,b).$$

  Therefore $\Gamma_1 \times \Gamma_2$ has at least $|\Gamma_1| \cdot |\Gamma_2|$ characters, and no more due to Corollary 4.6.

Together we find that Γ has exactly $\prod_{i \in [k]} |\mathbb{Z}/n_i\mathbb{Z}| = |\Gamma|$ characters. □

4.2.2. Calculation of the eigenvalues

Now that we have sufficient background on characters, we continue with finding the eigenvalues of the Cayley graph in question.

Theorem 4.8. *Let Γ be a finite abelian group, $\chi : \Gamma \to \mathbb{C}$ a character of Γ, and $S \subseteq \Gamma$ a set closed under taking inverses and not containing the identity of Γ. Let M be the adjacency matrix of the Cayley graph $G = Cay(\Gamma, S)$. Consider the vector $\mathbf{x} \in \mathbb{C}^\Gamma$ such that $\mathbf{x}_a = \chi(a)$. Then $\mathbf{x}$ is an eigenvector of G, with eigenvalue $\sum_{s \in S} \chi(s)$.*

Proof. Considering the a-th entry of the vector $M\mathbf{x}$ we get

$$\begin{aligned}(M\mathbf{x})_a &= \sum_{b \in \Gamma} M_{a,b}\mathbf{x}_b = \sum_{b : b-a \in S} \mathbf{x}_b \\ &= \sum_{b : b-a \in S} \chi(b) = \sum_{s \in S} \chi(a+s) = \sum_{s \in S} \chi(s)\chi(a) \\ &= \left(\sum_{s \in S} \chi(s)\right) \cdot \chi(a) = \left(\sum_{s \in S} \chi(s)\right) \cdot \mathbf{x}_a\end{aligned}$$

□

In our case, we are working with the Cayley graph $Cay(\Gamma, S)$, where Γ is the finite abelian group $\mathbb{Z}_2^{\binom{n}{2}}$. From Theorem 4.8 we get that any character is an eigenvector. Furthermore we know that the characters themselves are linearly independent from Lemma 4.5, and there are exactly $|\Gamma|$ characters of Γ, which is equal to the number of vertices in the Cayley graph $Cay(\Gamma, S)$. Hence we can enumerate all eigenvalues of the Cayley graph using Theorem 4.8.

All we need now is a character for every vertex of the Cayley graph. For $\mathbb{Z}_2$, we can make a character for any $x \in \{0,1\}$ of the form

$$\chi_x(a) = e^{2\pi i x a/2} = e^{\pi i x a} = (-1)^{xa}.$$

Based on the proof of Theorem 4.7, a working character for $\mathbb{Z}_2^{\binom{n}{2}}$ would be the product of the characters from $\mathbb{Z}_2$, i.e. if we let $d = \binom{n}{2}$, for any $x \in \mathbb{Z}_2^d$ we have a character of the form

$$\chi_x(a) = \prod_{i \in [d]} (-1)^{x_i a_i} = (-1)^{\sum_{i \in [d]} x_i a_i}.$$

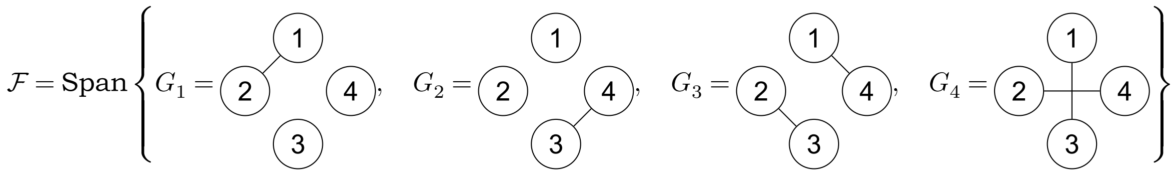
This is a useful, and easy to program, representation of the character, since the sum $\sum_{i \in [d]} x_i a_i$ is equal to the number of corresponding ones between the two binary strings x and a of length d. The code used can be found in Appendix B.

4.3. Comparing the upper bounds against known lower bounds

From Section 4.2.2 we showed for a Cayley graph $Cay(\Gamma, S)$ that a character χ of Γ gives the eigenvalues of the Cayley graph, via the character sum $\sum_{s\in S}\chi(s)$. In this section we use this method to obtain the eigenvalues and apply them to the Hoffman bound, Theorem 4.1, and Cvetković bound, Theorem 4.2, respectively. These bounds will be applied to the local graph class $\mathcal{H} = \{H : L \subseteq H\}$ and the collection $\mathcal{H} = \{H : L \cong H\}$ where $L \in \{K_3, K_4, P_3, C_4\}$ and the number of vertices range from 3 to 6. See Tables 4.1, 4.2, 4.3 and 4.4.

The lower bounds for $D_{\mathcal{H}}(n)$ for $\mathcal{H} = \{H : K_4 \subseteq H\}$ come from Lemma 3.9 combined with the fact that, denoting the Turán graph as $T(n,r)$, $\mathrm{ex}(5,K_4) = |E(T(5,3))| = 8$ and $\mathrm{ex}(6,K_4) = |E(T(6,3))| = 12$ via Turán's Theorem [19]. Lower bounds for $D_{\mathcal{H}}(n)$ for $\mathcal{H} = \{H : C_4 \subseteq H\}$ come from Lemma 3.9 combined with the results from Clapham, Flockhart, Sheehan [5], which give the extremal numbers $\mathrm{ex}(n, C_4)$ for $1 \le n \le 21$.

For the case $n = 4$ and $\mathcal{H} = \{H : P_3 \cong H\}$, we can bound $D_{\mathcal{H}}(n)$ by noting that the collection $\mathcal{F}$ as in (4.1) is valid for the independence problem, and is of size $2^4 = 16$. The span in this context is just the collection of all possible linear combinations of the graphs in the span under the symmetric difference operator $\oplus$.



(4.1)

Indeed, P_3 has only two edges. Since none of the graphs G_i in the span have overlapping edges, the only linear combinations of the graphs in the span that have 2 edges are $G_1 \oplus G_2$, G_3 or G_4, and none of these are paths of length 3.

Some other lower bounds can be lifted via Remark 3.19.

The question marks in Tables 4.3 and 4.4 are unknown as far as the literature is concerned. One can use Remark 3.19 to get a lower bound, but the difference between the upper and lower bounds in most of the cases is too big to be meaningful. Interestingly, for $n = 5$ and $\mathcal{H} = \{H : L \cong H\}$ and $L \in \{P_3, C_4, K_4\}$, the Hoffman bounds gives upper bounds which are powers of two. This could indicate that a linear solution of possible for these problems. Due to Example 3.6, we already know the exact value for $L = K_3$, and the Hoffman upper bound agrees with the exact value for every $n \in \{3,4,5,6\}$.

The Cvetković bound gives in most cases worse upper bounds compared to the Hoffamn bound for $D_{\mathcal{H}}(n)$. The biggest difference in the bounds in our results can be found in Table 4.4: for $n = 6$, and $\mathcal{H} = \{H : P_3 \subseteq H\}$. Here the Hoffman bound gives an upper bound of 57 compared to the upper bound of 19565 the Cvetković bound gives. However, interestingly, for $n = 6$ we get that $D_{\mathcal{H}}(n)$ for $\mathcal{H} = \{H : K_4 \cong H\}$ is smaller than 12952 according to the Cvetković bound. This is also the only occurrence of the Cvetković bound being a better upper bound than the Hoffman bound in our results. On the other hand:

Theorem 4.9. *For any $m \in \mathbb{Z}_{\ge 2}$ and for all $n \ge m$, if $\mathcal{H} = \{H : K_m \cong H\}$, then $M_{\mathcal{H}}(n) = 2$.*

Proof. We will first make a collection $\mathcal{F}$ which adheres to the clique problem $M_{\mathcal{H}}(n)$. Since the Cayley graph corresponding to the $\mathcal{H}$-code is vertex transitive, we can always put the empty graph on n vertices $\overline{K_n}$ in the collection. Any new graph we put in $\mathcal{F}$ can now only be a K_m. We can choose any K_m by symmetry, hence let K be the K_m on the vertices $1,2,\ldots,m$. Now $\mathcal{F} = \{\overline{K_n}, K\}$ is valid for the clique problem, and $M_{\mathcal{H}}(n) \ge 2$.

For the sake of contradiction, assume we can put another K_m, say $K' \ne K$ in the collection $\mathcal{F}$ with it being valid for the clique problem. Hence the symmetric difference of K' and K should produce a new K_m. Denote the number of vertices than intersect between K and K' by k. The number of edges in the symmetric difference $K \oplus K'$ is equal to

$$|E(K)| + |E(K')| - 2|E(K) \cap E(K')| = \binom{m}{2} + \binom{m}{2} - 2\binom{k}{2}$$

(see also Example 3.6).

Since $K \oplus K'$ is a K_m, k needs to be such that $2\binom{k}{2} = \binom{m}{2}$. This will give a single positive solution:

$$k = \frac{1}{2} + \frac{1}{2}\sqrt{1+2m(m-1)},$$

which, if integer, could imply $K \oplus K'$ is a K_m. (This is possible for e.g. $n = 4$ or $n = 21$.)

To this effect let m be such that k is integer and, WLOG, let K' be the K_m on the vertices $1, 2, \ldots, k$ and $v_{k+1}, v_{k+2}, \ldots, v_m$, where $v_{k+1}, v_{k+2}, \ldots, v_k$ are distinct and not from $\{k+1, k+2, \ldots, m\}$. The symmetric difference $K \oplus K'$ now contains the edges $\{1, v_m\}$ and $\{k, v_m\}$ but not the edge $\{1, k\}$. Since the number of edges in the symmetric difference is exactly $\binom{m}{2}$, $K \oplus K'$ is not a K_m, which is a contradiction. □

This would implicate that the inequality from Lemma 3.3 is not tight for this $\mathcal{H}$, as we now have

$$M_{\mathcal{H}}(n)D_{\mathcal{H}}(n) \leq 2 \cdot 12952 = 25904 < 32768 = 2^{\binom{6}{2}}.$$

Finally we note that for $\mathcal{H} = \{H : K_3 \subseteq H\}$, which is the open problem mentioned in section 3.1, the difference between the upper bounds and the known lower bound of $2^{\lfloor n^2/4 \rfloor}$ seems to become larger as n increases. Even worse is that Alon et al. [2] proved for $n = 3, 4, 5, 6$ vertices that the lower bound was tight. Therefore, this approach will most likely not give useful bounds for a higher number of vertices.

$\mathcal{H}$	d	λ_m	$m_{<0}$, $m_{=0}$, $m_{>0}$	Hoffman	Cvetković	Known lower bound	Spectrum
$K_3 \subseteq$	1	-1	$4,0,4$	4	4	$m/2=4$, by Example 3.5	$(1)^4(-1)^4$
$K_3 \cong$	1	-1	$4,0,4$	4	4	$m/2=4$, by Example 3.5	$(1)^4(-1)^4$
$P_3 \subseteq$	4	-2	$3,3,2$	$\frac{8}{3}<3$	5	$2^{\lfloor 3/2 \rfloor}=2$, by Example 3.10	$(4)^1(-2)^3(0)^3(2)^1$
$P_3 \cong$	3	-1	$6,0,2$	2	2	2, by Remark 3.19	$(3)^2(-1)^6$

Table 4.1: For $n=3$ vertices, an overview of the Cayley graphs for the collections $\mathcal{H}$ given. The Cayley graph is every case has $m=2^{\binom{3}{2}}=2^3=8$ vertices. The degree of the Cayley graph is denoted by d, λ_m is the smallest eigenvalue.

$\mathcal{H}$	d	λ_m	$m_{<0}$, $m_{=0}$, $m_{>0}$	Hoffman	Cvetković	Known lower bound	Spectrum
$K_3 \subseteq$	23	-9	$34,0,30$	18	30	$2^{\lfloor 16/4 \rfloor}=16$	$(23)^1(-9)^6(3)^{29}(-5)^7(-1)^{21}$
$K_3 \cong$	4	-4	$8,48,8$	32	56	$m/2=32$, by Example 3.6	$(4)^8(0)^{48}(-4)^8$
$P_3 \subseteq$	54	-6	$36,0,28$	$\frac{32}{5}<7$	28	$2^{\lfloor 4/2 \rfloor}=4$, by Example 3.10	$(54)^1(-6)^9(-2)^{27}(2)^{27}$
$P_3 \cong$	12	-4	$18,32,14$	16	46	16, by (4.1)	$(12)^2(4)^{12}(0)^{32}(-4)^{18}$
$C_4 \subseteq$	10	-6	$31,0,33$	24	31	$2^4=16$, see [5]	$(10)^1(-6)^6(2)^{29}(6)^3(-2)^{25}$
$C_4 \cong$	3	-1	$48,0,16$	16	16	16, by Remark 3.19	$(3)^{16}(-1)^{48}$
$K_4 \subseteq$	1	-1	$32,0,32$	32	32	$m/2=32$, by Example 3.5	$(1)^{32}(-1)^{32}$
$K_4 \cong$	1	-1	$32,0,32$	32	32	$m/2=32$, by Example 3.5	$(1)^{32}(-1)^{32}$

Table 4.2: For $n=4$ vertices, an overview of the Cayley graphs for the collections $\mathcal{H}$ given. The Cayley graph is every case has $m=2^{\binom{4}{2}}=2^6=64$ vertices. The degree of the Cayley graph is denoted by d, λ_m is the smallest eigenvalue.

$\mathcal{H}$	d	λ_m	$m_{<0}$, $m_{=0}$, $m_{>0}$	Hoffman	Cvetković	Known lower bound
$K_3 \subseteq$	636	-106	498, 150, 376	$\frac{1024}{7} < 147$	526	$2^{\lfloor 25/4 \rfloor} = 64$
$K_3 \cong$	10	-10	416, 192, 416	512	608	$m/2 = 512$, by Example 3.6
$P_3 \subseteq$	998	-18	578, 0, 446	$\frac{2304}{127} < 19$	446	$2^{\lfloor 5/2 \rfloor} = 4$, by Example 3.10
$P_3 \cong$	30	-10	542, 0, 482	256	482	?
$C_4 \subseteq$	476	-120	497, 60, 467	$\frac{30720}{149} < 207$	527	$2^6 = 64$, see [5]
$C_4 \cong$	15	-5	672, 0, 352	256	352	?
$K_4 \subseteq$	66	-36	480, 120, 424	$\frac{6144}{17} < 362$	544	$2^8 = 256$, by Turán's Theorem [19]
$K_4 \cong$	5	-5	512, 0, 512	512	512	?

Table 4.3: For $n = 5$ vertices, an overview of the Cayley graphs for the collections $\mathcal{H}$ given. The Cayley graph is every case has $m = 2^{\binom{5}{2}} = 2^{10} = 1024$ vertices. The degree of the Cayley graph is denoted by d, λ_m is the smallest eigenvalue. Spectrum has been left out due to its size.

$\mathcal{H}$	d	λ_m	$m_{<0}$, $m_{=0}$, $m_{>0}$	Hoffman	Cvetković	Known lower bound
$K_3 \subseteq$	26979	-1815	17407, 0, 15361	$\frac{9912320}{4799} < 2066$	15361	$2^{\lfloor 36/4 \rfloor} = 512$
$K_3 \cong$	20	-20	10112, 12544, 10112	16384	22656	$m/2 = 16384$, by Example 3.6
$P_3 \subseteq$	32692	-56	13203, 6825, 12740	$\frac{458752}{8187} < 57$	19565	$2^{\lfloor 6/2 \rfloor} = 8$, by Example 3.10
$P_3 \cong$	60	-12	15020, 6264, 11484	$\frac{16384}{3} < 5462$	17748	?
$C_4 \subseteq$	24784	-2412	14764, 1410, 16594	$\frac{19759104}{6799} < 2907$	16174	$2^7 = 128$, see [5]
$C_4 \cong$	45	-15	18624, 0, 14144	8192	14144	?
$K_4 \subseteq$	5142	-1666	17441, 0, 15327	$\frac{6823936}{851} < 8019$	15327	$2^{12} = 4096$, by Turán's Theorem [19]
$K_4 \cong$	15	-13	19816, 0, 12952	$\frac{106496}{7} < 15214$	12952 (!)	?

Table 4.4: For $n = 6$ vertices, an overview of the Cayley graphs for the collections $\mathcal{H}$ given. The Cayley graph is every case has $m = 2^{\binom{6}{2}} = 2^{15} = 32768$ vertices. The degree of the Cayley graph is denoted by d, λ_m is the smallest eigenvalue. Spectrum has been left out due to its size.

5

Open questions and future research

This chapter pertains topics from this thesis which contain unanswered questions, which can be further explored in future research. Section 5.1 will also explain an approach made to an algorithm solving the problem for small inputs.

5.1. Determining $D_{\mathcal{H}}(n)$ for $\mathcal{H} = \{H : K_3 \subseteq H\}$

As mentioned in section 3.1, the maximum cardinality collection for which no two graphs have their symmetric difference contain a K_3 is still open. Alon et al. proved for $n = 3,4,5,6$ that the lower bound of $2^{\lfloor n^2/4 \rfloor}$, which stems from Lemma 3.9, is tight, by giving constructions for the clique problem $M_{\mathcal{H}}(n)$ and using Lemma 3.3. For more than 6 vertices, the independence problem is still open, as per Conjecture 5.1.

Conjecture 5.1 ([2]). *Let $\mathcal{H} = \{H : K_3 \subseteq H\}$. For all $n \geq 7$,*

$$M_{\mathcal{H}}(n) \geq 2^{\binom{n}{2} - \lfloor n^2/4 \rfloor}$$

Conjecture 5.1 would give equality to both $D_{\mathcal{H}}(n)$ and $M_{\mathcal{H}}(n)$ by Lemma 3.3.

Finding the value of $M_{\mathcal{H}}(n)$ for $n = 7$ would already be an improvement, and could be able to be computed. One would brute force this problem by solving the clique problem on the Cayley graph corresponding to $\mathcal{H}$ explicitly. As a reminder, the ILP for the clique number of a graph on m vertices is:

$$\omega(G) = \max \left\{ \sum_{i=1}^{m} x_i : \quad x_i + x_j \leq 1 \ \ \forall ij \notin E(G), \quad x_i \in \{0,1\} \ \ \forall i \in [m] \right\}. \tag{C}$$

Note that if we were to brute force the clique problem in its current state, there are $2^{\binom{7}{2}} = 2^{21} = 2097152$ different graphs on 7 vertices, and therefore there are $\binom{2097152}{2} \approx 2.199 \cdot 10^{12}$ pairs of graphs to compare. This is too much for any computer to do in a reasonable amount of time.

On the other hand, the check whether the symmetric difference of two pairs of graphs on n vertices contains a triangle can be done in $O(|V(K_n)||E(K_n)|) = O(n^3)$ time which, for a constant number of vertices n, is quickly computable. The main issue is we would need a way to drastically reduce the number of constraints we are adding.

Since the graph for which we are trying to calculate the independence number of is vertex transitive, any vertex is part of a maximum clique. This means that we could start at a random graph-vertex, say the empty graph, and determine locally which graphs we can add to our clique collection. This gives rise to a branch-and-cut approach, which will be explained in Section 5.1.1.

5.1.1. A potential branch-and-cut approach

This approach is an iterative algorithm which should give an optimal solution, and should reduce the number of constraints added.

Denote the number of vertices of the Cayley graph G on n vertices corresponding to $\mathcal{H}$, by m. First we relax the ILP for the clique number (**C**), by removing the condition that x_i is integer by letting $x_i \in [0,1]\ \ \forall i \in [m]$. We also remove the non-edge constraints $x_i + x_j \leq 1$ for every edge ij in the Cayley graph and define for some selection of non-edges $P \subseteq E(K_m) \setminus E(G)$, and two selections of vertices $Y, Z \subseteq V(G)$:

$$\omega_{\text{relax}}(P,Y,Z) = \max\left\{ \sum_{i=1}^{m} x_i : \quad x_i + x_j \leq 1 \ \ \forall ij \in P, \quad x_i \geq 1 \ \ \forall i \in Y, \right. \\ \left. x_i \leq 0 \ \ \forall i \in Z, \quad x_i \in [0,1] \ \ \forall i \in [m] \right\}. \tag{$\mathbf{C_r}$}$$

Here P will become a collection of local cuts, Y will be the collection of vertices that are forced to be in the solution, and Z will be the collection of vertices that are forced to not be in the solution.

We initialize the branch-and-cut algorithm with the parameters $(P,Y,Z) = (\varnothing, \{\varnothing\}, \varnothing)$, where we put the vertex belonging to the empty graph in our solution.

Once we solve ($\mathbf{C_r}$) for the Cayley graph once, it will give an optimal solution C. For each pair of elements i,j in the solution C such that $x_i + x_j > 1$, we check whether the symmetric difference contains a triangle. If the check is negative, we add the edge ij to the collection P so that the inequality $x_i + x_j > 1$ is is no longer possible. Once we have checked every pair in the solution C, if there was a cut added to P, we solve ($\mathbf{C_r}$) again, and if not, we continue with the branching algorithm.

We now have a solution $\widehat{C}$ to the relaxed problem, where no more cutting planes can be added. If the solution is integer, we are done and have found a solution for the clique ILP (**C**). Indeed, it satisfies the integer requirement of (**C**) and if there was a non-edge $ij \notin E(G)$ such that $x_i + x_j > 1$, it would have been added to P, and so $\widehat{C}$ is a solution for (**C**).

If the solution $\widehat{C}$ is not integer, we can find an element k in our solution $\widehat{C}$ for which x_k is non-integer. Next we branch in two paths, one with input $(P, Y \cup \{k\}, Z)$, i.e. $x_k = 1$ is chosen, and one with input $(P, Y, Z \cup \{k\})$, i.e. $x_k = 0$ is chosen.

For each of the new branches we go over the entire process again, until all branches have been pruned due to (**C**) returning an integer solution, or pruned due to the problem becoming infeasible. An optimal solution has either $x_k = 1$ or $x_k = 0$ for every vertex k, so that optimal solution must lie on one of the branches $(P, Y \cup \{k\}, Z\}$ or $(P, Y, Z \cup \{k\})$. Therefore, once every branch has been explored, the algorithm must have found that optimal solution. Hence the maximum objective value of all solutions found via the branch-and-cut method is optimal for the maximum value for (**C**).

An attempt on building this algorithm was made, and it gives the same results as Alon [2] for small inputs. Only issue is that, when branching in the branch-and-cut approach, we only add one vertex to the solution at a time. However, the number $2^{\binom{n}{2} - \lfloor n^2/4 \rfloor}$ will grow exponentially, hence the program slows down in efficiency when the number of vertices increases.

Here we explain a speedup that could be applied:. If we assume the optimal solution to the clique problem is a linear collection, i.e. closed under taking the symmetric difference, then we can apply the following: For every new element k that we add to Y, add also the elements $\{k \oplus y : y \in Y\}$ to Y. Similarly, for every k that we add to Z, add also the elements $\{k \oplus z : z \in Z\}$ to Z. Denoting the number of branches the original method needs to explore by B, the speedup would only have to explore $\log_2(B)$ branches. Using this speedup would therefore drastically increase the efficiency, and would make it possible to calculate the value $M_{\mathcal{H}}(7)$.

The big downside of this speedup is that the optimal solutions need not be linear, in which case this speedup would only show that no linear solution exists. If one could prove that $M_{\mathcal{H}}(n)$ has a linear solution for any n, which the author doubts is the case, this issue would be non-existent.

5.2. General bounds on $D_{\mathcal{H}}(n)$ and $d_{\mathcal{H}}(n)$ for $\mathcal{H} = \{H : L \cong H\}$

In the calculations of the upper bound for $D_{\mathcal{H}}(n)$ using eigenvalues that were done in Section 4.3, we found that for the isomorphism classes $\mathcal{H} = \{H : L \cong H\}$ with $L \in \{P_3, K_4, C_4\}$, the Hoffman bounds returned a power of 2 as an upper bound for $n = 5$ and $n = 6$, while lower bounds for $D_{\mathcal{H}}(n)$ were unknown. Constructions for the independence problem $D_{\mathcal{H}}(n)$ for these L would be very helpful to show the accuracy of the upper bounds in these cases. Brute force solutions are also possible to do here since the number of vertices is small, so exact values can be obtained, and can be compared against the upper bounds.

As noted in Section 4.3, when $L = K_4$, we know that $M_{\mathcal{H}}(n) = 2$ from Theorem 4.9, and the upper bound on $D_{\mathcal{H}}(n)$ for $n = 6$ showed that the inequality from Lemma 3.3 was not tight. One could alter the proof of Theorem 3.18 such that it could be applied for the family $\mathcal{H} = \{H : K_4 \cong H\}$, and give a bound for $d_{\mathcal{H}}(n)$, which would answer an open question asked by Alon [1]. One would have to find a family of graphs for which a symmetric difference of two graphs within that collection is a K_4, the existence of which is unknown to the author.

Furthermore, from Corollary 3.14 we know that for the families $\mathcal{H}'_{2k} = \{H : C_{2k} \cong H\}$ with $k \geq 2$, we have $d_{\mathcal{H}'_{2k}}(n) = O(n^{-1})$. Theorems 3.20 and 3.22 cannot be applied to these families since they were proven for $\mathcal{H}_{2k} = \{H : C_{2k} \subseteq H\}$

5.3. Lower bounds on $d_{\mathcal{H}}(n)$ for $\mathcal{H} = \{H : L \subseteq H\}$

In Section 3.2.3 we discussed the approach Alon used to prove a lower bound on $d_{\mathcal{H}'}(n)$ in the case that $\mathcal{H}' = \{H : K_{1,2k} \cong H\}$. We found that the proof of Theorem 3.24 was insufficient to prove the same lower bound for the independence ratio $d_{\mathcal{H}}(n)$ of the subgraph family $\mathcal{H} = \{H : L \subseteq H\}$. For this family $\mathcal{H}$ we already get a lower bound from Lemma 3.9, but the approach from Section 3.2.3 might tighten the bound even further. Hence, altering the method used here to workk for the family $\mathcal{H}$ might be interesting research for later.

The proof also relied on the fact that the columns of a BCH-code with a certain minimum Hamming distance gave rise to a collection of binary vectors which the vectors v_e could be assigned to. Perhaps a different type of code could be used that has different properties pertaining its columns or rows, which could be used for the subgraph problem.

5.4. New upper bounds on $d_{\mathcal{H}}(n)$

We were able to prove an upper bound for $d_{\mathcal{H}_{2k}}(n)$ where $\mathcal{H}_{2k} = \{H : C_{2k} \subseteq H\}$ of approximate order $O(n^{-2n})$, using an approach by Alon [1]. This bound turned out to be worse than the upper bound indirectly found by Morris and Saxton [14], see Corollary 3.23 and the remarks before. The method used by Alon could most certainly be expanded upon for families of graphs $\mathcal{H}$ for which we do not know any upper bound for $d_{\mathcal{H}}(n)$.

An earlier approach was made by the author to determine the independence ratio for $\mathcal{H}'_{2k} = \{H : C_{2k} \cong H\}$ where, in the style on Theorems 3.20 and 3.22, we let $\mathcal{G}$ be a subgraph of the Cayley graph corresponding to $\mathcal{H}'_{2k}$ equal to the collection of $(k+1)$-cycles. It turned out that $\mathcal{F} \subseteq \mathcal{G}$ was independent in the Cayley graph if and only if two graphs $F_1, F_2 \in \mathcal{F}$ did not have exactly 2 vertices intersect **and** those two vertices had an edge in both F_1 and F_2. It would not be completely possible to use the Frankl-Füredi Theorem 3.16 immediately in this case. If one could alter the statement of the Frankl-Füredi Theorem, one might be able to give a better upper bound than the inverse linear bound of Corollary 3.14.

A

Proof of Theorem 3.17

To restate the theorem in question for clarity:

Theorem. *Let G be a vertex transitive graph, and let $H \subseteq G$ be a subgraph of G, then*

$$\frac{\alpha(G)}{|V(G)|} \leq \frac{\alpha(H)}{|V(H)|}.$$

For the proof we will need the definition of the fractional chromatic number.

Definition A.1 (Fractional chromatic number)**.** Let G be a graph and denote $\mathcal{I}(G)$ as the collection of all independent sets in G. The **fractional chromatic number** $\chi_f(G)$ of a graph G is the solution of the following linear optimization problem:

$$\chi_f(G) = \min\left\{\sum_{I\in\mathcal{I}(G)} \lambda_I : \quad \sum_{I\in\mathcal{I}(G), v\in I} \lambda_I \geq 1 \ \forall v \in V(G), \quad \lambda_I \geq 0 \ \forall I \in \mathcal{I}(G)\right\}, \qquad \textbf{(P)}$$

or equivalently, the solution to its dual linear problem:

$$\chi_f(G) = \max\left\{\sum_{v\in V(G)} w_v : \quad \sum_{v\in I} w_v \leq 1 \ \forall I \in \mathcal{I}(G), \quad w_v \geq 0 \ \forall v \in V(G)\right\}. \qquad \textbf{(D)}$$

The equality of both forms of $\chi_f(G)$ comes from strong duality: both the primal (**P**) and the dual (**D**) are feasible, and bounded. Indeed, for the primal (**P**), $\lambda = (\lambda_I)_{I\in\mathcal{I}(G)} = \mathbf{1}$ is a feasible solution, and for the dual (**D**), $w = (w_v)_{v\in V(G)} = \mathbf{0}$ is feasible.

Furthermore, in the primal (**P**) we can clearly see that since $\lambda_I \geq 0$ for all independent sets I in $\mathcal{I}(G)$, that the objective value $\sum_{I\in\mathcal{I}(G)} \lambda_I$ is at least $\sum_{I\in\mathcal{I}(G)} 0 = 0$. Finally in the dual (**D**), since $\sum_{v\in I} w_v \leq 1$ for all independent sets I in $\mathcal{I}(G)$, $w_v \leq 1$ for all $v \in V(G)$ as each individual vertex is an independent set in G. So the objective value $\sum_{v\in V(G)} w_v$ is at most $\sum_{v\in V(G)} 1 = |V(G)|$.

Now we can start with the proof.

Proof of the theorem. We split the proof in four parts:

- First we prove $\alpha(G)\chi_f(G) \geq |V(G)|$. Consider the dual (**D**) of the fractional chromatic number. We claim that the vector w where $w_v = 1/\alpha(G)$ for all $v \in V(G)$ is a feasible solution. Indeed, clearly $w_v \geq 0$ for all $v \in V(G)$, and for each independent set I in $\mathcal{I}(G)$ we have

  $$\sum_{v\in I} w_v = \sum_{v\in I} \frac{1}{\alpha(G)} = \frac{|I|}{\alpha(G)} \leq 1.$$

  Hence, since w is feasible, we must have

  $$\chi_f(G) \geq \sum_{v\in V(G)} w_v = \sum_{v\in V(G)} \frac{1}{\alpha(G)} = \frac{|V(G)|}{\alpha(G)} \iff \alpha(G)\chi_f(G) \geq |V(G)|.$$

- Next we prove, for vertex transitive graphs G, that $\alpha(G)\chi_f(G) \leq |V(G)|$. For vertex transitive graphs, every vertex is contained at least one, and in the same amount of maximum independent sets. Call this number k.

  In the context of the primal problem (**P**), let $\lambda = (\lambda_I)_{I \in \mathcal{I}(G)}$ where $\lambda_I = 1/k$ if I is of maximum size, and 0 otherwise. This is a feasible solution to the primal problem as each λ_I is greater or equal to 0, and for all vertices $v \in V(G)$ we have

$$\sum_{I \in \mathcal{I}(G), v \in I} \lambda_I = \sum_{I \in \mathcal{I}(G) \text{ max indep.}, v \in I} \lambda_I = k \cdot \frac{1}{k} = 1.$$

  If we denote the total number of maximum independent sets as ℓ, we find that

$$\chi_f(G) \leq \sum_{I \in \mathcal{I}(G)} \lambda_I = \sum_{I \in \mathcal{I}(G) \text{ max indep.}} \lambda_I = \frac{\ell}{k}.$$

  We prove that the ratios ℓ/k and $|V(G)|/\alpha(G)$ are the same, via a double counting argument. Let S be the set of pairs (v, I) where I is a maximum independent set, and $v \in I$. On one side, we have

$$\begin{aligned} |S| &= |\{(v, I) : v \in V(G), I \in \mathcal{I}(G) \text{ max indep}, v \in I\}| \\ &= \sum_{I \in \mathcal{I}(G) \text{ max indep}} \left| \bigcup_{v \in I} \{(v, I)\} \right| = \sum_{I \in \mathcal{I}(G) \text{ max indep}} \alpha(G) = \ell\alpha(G), \end{aligned}$$

  and on the other side we have

$$\begin{aligned} |S| &= |\{(v, I) : v \in V(G), I \in \mathcal{I}(G) \text{ max indep}, v \in I\}| \\ &= \sum_{v \in V(G)} \left| \bigcup_{I \in \mathcal{I}(G) \text{ max indep}, v \in I} \{(v, I)\} \right| = \sum_{v \in V(G)} k = |V(G)|k. \end{aligned}$$

  Together we get $\ell/k = |V(G)|/\alpha(G)$, and so

$$\chi_f(G) \leq \frac{|V(G)|}{\alpha(G)} \iff \alpha(G)\chi_f(G) \leq |V(G)|.$$

- Now we prove that for any subgraph $H \subseteq G$, we have $\chi_f(G) \geq \chi_f(H)$. In the context of the dual problem (**D**), consider an optimal solution w^H for $\chi_f(H)$. We can make a solution w^G for $\chi_f(G)$ by letting

$$w_v^G = \begin{cases} w_v^H & \text{if } v \in H \\ 0 & \text{else} \end{cases}$$

  This is a feasible solution for the dual problem as $w_v^G \geq 0$ for all $v \in V(G)$ and for any independent set $I \in \mathcal{I}(G)$ we know $I \cap V(H)$ is independent in H and so

$$\sum_{v \in I} w_v^G = \sum_{v \in (I \cap V(H))} w_v^H \leq 1.$$

  Therefore

$$\chi_f(G) \geq \sum_{v \in V(G)} w_v^G = \sum_{v \in V(H)} w_v^H = \chi_f(H).$$

Finally, putting everything together we have that for any graph G we have $\alpha(G)\chi_f(G) \geq |V(G)|$, with equality for vertex transitive graphs, and for any subgraph $H \subseteq G$ we have $\chi_f(G) \geq \chi_f(H)$. Therefore, for any vertex transitive graph G and subgraph $H \subseteq G$ we find:

$$\frac{\alpha(G)}{|V(G)|} = \frac{1}{\chi_f(G)} \leq \frac{1}{\chi_f(H)} \leq \frac{\alpha(H)}{|V(H)|}.$$

□

B

Sagemath code for finding the eigenvalues of the Cayley graph

```
from itertools import product
from collections import Counter

# This loads the number of vertices n, the number of edges of K_n and
# the collection of forbidden/necessary structures
n = 6
num_edges = n * (n - 1) // 2  # Number of edges in the complete graph K_n
subgraph = False # "subgraph = False" is for isomorphism

# H, num_ver_H, num_edg_H = graphs.CompleteGraph(Integer(3)), 3, 3 # check for K_3
H, num_ver_H, num_edg_H = graphs.CompleteGraph(Integer(4)), 4, 6 # check for K_4
# H, num_ver_H, num_edg_H = graphs.CycleGraph(Integer(4)), 4, 4 # check for C_4
# H, num_ver_H, num_edg_H = graphs.PathGraph(Integer(3)), 3, 2 # check for P_3

# This contains helper functions for the calculation of the eigenvalues.
global memo
memo = set()

def add_in_zero_pos(binary_string) -> set:
    """Given a binary string s, returns all binary strings s' such
    that the intersection of s and s' is precisely s.
    Equivalently, converting s to a graph G_s, this function returns all graphs
    containing G_s as a subgraph."""
    # Positions where S has 0
    zero_positions = [i for i, b in enumerate(binary_string) if b == '0']
    k = len(zero_positions)
    result_set = set()

    for combo in product("01", repeat=k):
        s_list = list(binary_string)
        for pos, bit in zip(zero_positions, combo):
            s_list[pos] = bit
        result_set.add(int("".join(s_list),2))
    return result_set

def int_graph_subgraph_check(H_forbid, G_int) -> bool:
    """Given a integer s (as a binary s_bin), the function checks if the according graph
    contains the forbidden graph H.
```

```
    In the case of not containing H as a subgraph, this function adds all binary strings
    s' such that the intersection of s' and s_bin is
    precisely s_bin to a memo for speedup."""
    global memo
    if G_int in memo:
        return True
    G_graph = Graph(n)
    idx_counter = 0
    pre = bin(G_int)[2:]
    G_binary = '0'*(num_edges - len(pre)) + pre
    for i in range(n):
        for j in range(i+1,n):
            if G_binary[idx_counter] == '1':
                G_graph.add_edge((i,j))
            idx_counter += 1
    if H.is_subgraph(G_graph, induced = False, up_to_isomorphism = True):
        memo = memo.union(add_in_zero_pos(G_binary))
        return True
    return False

def int_graph_isomorphism_check(H, G_int) -> bool:
    """Given a integer s (as a binary s_bin), the function checks if the according graph
    is the forbidden graph H. The only check needed is graphs on num_edg_H vertices"""
    if G_int.bit_count() != num_edg_H:
        return False
    G_graph = Graph(n)
    idx_counter = 0
    pre = bin(G_int)[2:]
    G_binary = '0'*(num_edges - len(pre)) + pre
    for i in range(n):
        for j in range(i+1,n):
            if G_binary[idx_counter] == '1':
                G_graph.add_edge((i,j))
            idx_counter += 1
    if H.is_subgraph(G_graph, induced = False, up_to_isomorphism = True):
        return True
    return False

# This contains the calculation of the eigenvalues
def eigenvalues_of_cayley_graph() -> tuple:
    """As per the theory, there is a character chi for the Cayley graph Cay(Gamma,S) of
    the group Gamma = {0,1}^d with bitwise xor addition (d = 2**(n(n-1)//2)) and
    generating set $S$ of the form: chi(x) = (-1)**(sum r_i*x_i)
    for any element r in {0,1}^d.  Let x in ({-1,1})^({0,1}^d) be such that x_a = chi(a),
    then x is an eigenvector with eigenvalue sum_{s in S} chi(s).

    "sum r_e*x_e" is equal to the number of edges in the intersection of a graph s from S
    (x_e) and an arbitrary graph G (r_e).
    Function returns dictionary of the eigenvalues, and the size of the
    generating set (the degree of the Cayley graph)"""

    if subgraph:
        S = [G_int for G_int in range(0, 2**num_edges) \
             if int_graph_subgraph_check(H, G_int)] # For subgraph
    else:
        S = [G_int for G_int in range(0, 2**num_edges) \
             if int_graph_isomorphism_check(H, G_int)] # For isomorphism
    S_size = len(S)
```

```
    eigenvalues = Counter()
    for i, graph_1 in enumerate(range(0,2**num_edges)):
        if i == 1 << (num_edges - 1):
            print("halfway")
        dum_sum = 0
        for graph_2 in S:
            # Bitwise AND, parity check
            parity = (graph_1 & graph_2).bit_count() & 1
            dum_sum += 1 - 2 * parity
        eigenvalues[dum_sum] += 1
    return dict(eigenvalues), S_size

# This last part prints the results
eigenvals, degree = eigenvalues_of_cayley_graph()
zeroes = 0
negative = 0
positive = 0

least = min(eigenvals.keys())
for eigval, mult in eigenvals.items():
    if eigval < 0:
        negative += mult
    elif eigval == 0:
        zeroes += mult
    else:
        positive += mult
print(f"n = {n}")
print(f"H is a(n) {H} on {num_ver_H} vertices, subgraph = {subgraph}")
print(f"eigenvalues of Cayley graph = {eigenvals})
print(f"number of eigenvalues = {2**(num_edges)}")
print(f"zeroes = {zeroes}, positive = {positive}, negative = {negative}")
print(f"smallest eigenvalue = {least}, degree of Cayley graph = {degree}")

Cvetkovic = zeroes + min(positive, negative)
Hofmann = 2**(num_edges) * (-least/(degree - least))
print(f"Cvetkovic bound = {Cvetkovic}, Hofmann bound = {Hofmann}")
```

Bibliography

[1] Noga Alon. “Graph-codes”. In: *European Journal of Combinatorics* 116 (2024), p. 103880. ISSN: 0195-6698. DOI: https://doi.org/10.1016/j.ejc.2023.103880. URL: https://www.sciencedirect.com/science/article/pii/S0195669823001981.

[2] Noga Alon et al. “Structured Codes of Graphs”. In: *SIAM Journal on Discrete Mathematics* 37.1 (2023), pp. 379–403. DOI: 10.1137/22M1487989. eprint: https://doi.org/10.1137/22M1487989. URL: https://doi.org/10.1137/22M1487989.

[3] Aaron Berger and Yufei Zhao. “K4-intersecting families of graphs”. In: *Journal of Combinatorial Theory, Series B* 163 (2023), pp. 112–132. ISSN: 0095-8956. DOI: https://doi.org/10.1016/j.jctb.2023.07.005. URL: https://www.sciencedirect.com/science/article/pii/S0095895623000606.

[4] Fan R. K. Chung et al. “Some intersection theorems for ordered sets and graphs”. In: *J. Comb. Theory A* 43.1 (1986), pp. 23–37. DOI: 10.1016/0097-3165(86)90019-1. URL: https://doi.org/10.1016/0097-3165(86)90019-1.

[5] C. R. J. Clapham, Andrew D. Flockhart, and John Sheehan. “Graphs without four-cycles”. In: *J. Graph Theory* 13 (1989), pp. 29–47. URL: https://api.semanticscholar.org/CorpusID:206079774.

[6] David Ellis, Yuval Filmus, and Ehud Friedgut. *Triangle-Intersecting Families of Graphs*. 2012. arXiv: 1010.4909 [math.CO]. URL: https://arxiv.org/abs/1010.4909.

[7] P. ERDÓS, CHAO KO, and R. RADO. “INTERSECTION THEOREMS FOR SYSTEMS OF FINITE SETS”. In: *The Quarterly Journal of Mathematics* 12.1 (Jan. 1961), pp. 313–320. ISSN: 0033-5606. DOI: 10.1093/qmath/12.1.313. eprint: https://academic.oup.com/qjmath/article-pdf/12/1/313/7288156/12-1-313.pdf. URL: https://doi.org/10.1093/qmath/12.1.313.

[8] Paul Erdös. “Extremal problems in graph theory”. In: *Publ. House Cszechoslovak Acad. Sci., Prague* (1964), pp. 29–36.

[9] Peter Frankl and Zoltán Füredi. “Forbidding just one intersection”. In: *Journal of Combinatorial Theory, Series A* 39.2 (1985), pp. 160–176. ISSN: 0097-3165. DOI: https://doi.org/10.1016/0097-3165(85)90035-4. URL: https://www.sciencedirect.com/science/article/pii/0097316585900354.

[10] C. Godsil and K. Meagher. *Erdos-Ko-Rado Theorems: Algebraic Approaches*. CAMBRIDGE STUDIES IN ADVANCED. Cambridge University Press, 2016. ISBN: 9781107128446. URL: https://books.google.nl/books?id=P0XjCgAAQBAJ.

[11] Gy. Katona. “Intersection theorems for systems of finite sets”. In: *Acta Mathematica Academiae Scientiarum Hungarica* 15.3 (1964), pp. 329–337. ISSN: 1588-2632. DOI: 10.1007/BF01897141. URL: https://doi.org/10.1007/BF01897141.

[12] Swastik Kopparty, Aditya Potukuchi, and Harry Sha. *Error-Correcting Graph Codes*. 2024. arXiv: 2406.13867 [cs.IT]. URL: https://arxiv.org/abs/2406.13867.

[13] A. Kotzig. “Hamilton graphs and Hamilton circuits”. In: *Theory of Graphs and its Applications.* Proceedings of the Symposium held in Smolenice: Academic Press, New York, June 1964, 63–82 and 162.

[14] Robert Morris and David Saxton. *The number of C_{2l}-free graphs*. 2015. arXiv: 1309.2927 [math.CO]. URL: https://arxiv.org/abs/1309.2927.

[15] B. L. Rothschild P. Erdős D. J. Kleitman. “Asymptotic Enumeration of K_n-free Graphs”. In: *Colloquio Internazionale sulle Teorie Combinatorie* (1976).

[16] R. M. Wilson P. Frankl. “Intersection theorems with geometric consequences”. In: *Combinatorica* 1 (4 1981), pp. 357–368. ISSN: 1439-6912. URL: `https://doi.org/10.1007/BF02579457`.

[17] D. J. S. Robinson. *An introduction to abstract algebra.* De Gruyter, Inc., 2003.

[18] Vladimir D. Tonchev. “Error-correcting codes from graphs”. In: *Discrete Mathematics* 257.2 (2002). Kleitman and Combinatorics: A Celebration, pp. 549–557. ISSN: 0012-365X. DOI: `https://doi.org/10.1016/S0012-365X(02)00513-7`. URL: `https://www.sciencedirect.com/science/article/pii/S0012365X02005137`.

[19] Paul Turán. “On an extremal problem in graph theory”. In: *Matematikai és Fizikai Lapok (in Hungarian)* 48 (1941), pp. 436–452.

[20] Leo Versteegen. “Upper Bounds for Linear Graph Codes”. In: *Random Structures & Algorithms* 66.1 (2025), e21263. DOI: `https://doi.org/10.1002/rsa.21263`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.21263`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.21263`.