Problems in extremal graphs and poset theory

by

Shanise Walker

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Mathematics

Program of Study Committee:
Ryan R. Martin, Major Professor
Steve Butler
Leslie Hogben
Elgin Johnston
Bernard Lidický

Iowa State University

Ames, Iowa

2018

DEDICATION

To Zariah M. Lane for always being loving.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

ACKNOWLEDGMENTS

I would like to take this opportunity to express my gratitude and thanks to those who supported me during this process.

First and foremost, I would like to thank my adviser Dr. Ryan R. Martin for his guidance and support throughout the research and dissertation writing process. I have learned so much under your guidance and direction. Thank you for passing on your wealth of knowledge and giving me the necessary tools to succeed. I would like to thank each person on my committee for their efforts in guiding me through various steps on the graduate process: Dr. Steve Butler, Dr. Leslie Hogben, Dr. Elgin Johnston, and Dr. Bernard Lidický.

I would also like to thank my family for all of their support. There were many days and nights where I leaned on them for extra motivation. I would like to give a special thank you to my aunt, Carmethia Daniels, and brother, Calvin Walker, for always being positive and uplifting. Without the many long nights of listening to me ramble about my thoughts this would not have been possible.

I would also like to thank the Lee family for being my family away from home. I have enjoyed each moment we have shared. Lastly, I would like to thank my friends for always encouraging me.

ABSTRACT

In this dissertation, we present three different research topics and results regarding such topics. We introduce partially ordered sets (posets) and study two types of problems concerning them–forbidden subposet problems and induced-poset-saturation problems. We conclude by presenting results obtained from studying vertex-identifying codes in graphs.

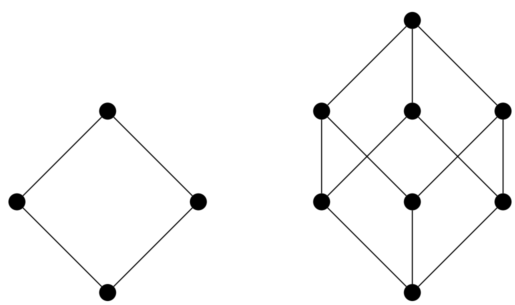
In studying forbidden subposet problems, we are interested in estimating the maximum size of a family of subsets of the n-set avoiding a given subposet. We provide a lower bound for the size of the largest family avoiding the $\mathcal{N}$ poset, which makes use of error-correcting codes. We also provide and upper and lower bound results for a k-uniform hypergraph that avoids a *triangle*. Ferrara et al. introduced the concept of studying the minimum size of a family of subsets of the n-set avoiding an induced poset, called induced-poset-saturation [19]. In particular, the authors provided a lower bound for the size of an induced-antichain poset and we improve on their lower bound result.

Let $G = (V, E)$ be a graph with vertex set V and edge set E. For any nonnegative integer r, let $B_r(v)$ denote the ball of radius r around vertex $v \in V$. For a finite graph G, an r-vertex-identifying code in G is a subset $C \subset V(G)$, with the property that $B_r(u) \cap C \neq B_r(v) \cap C$, for all distinct $u, v \in V(G)$ and $B_r(v) \cap C \neq \emptyset$, for all $v \in V(G)$. We study graphs with large symmetric differences and (p, β)-jumbled graphs and estimate the minimum size of a vertex-identifying code in each graph.

CHAPTER 1. INTRODUCTION

A *partially ordered set* (poset) is a set with a partial order. A *partial order* is a binary relation "$\preceq$" over a set P such that the relation is reflexive, antisymmetric, and transitive. Let x, y, z be elements such that $y \neq x$ and $z \notin \{x, y\}$. We say element y of a poset *covers* element x if there is no element z such that $x \preceq z \preceq y$. To represent a poset, we typically draw its Hasse diagram. A *Hasse diagram* is a graphical diagram of a poset that is displayed using the cover relation of the poset. The Hasse diagram is used to present a poset by drawing its transitive reduction.

The *n-dimensional Boolean lattice*, $\mathcal{B}_n$, denotes the poset $(2^{[n]}, \subseteq)$, where $[n] := \{1, \ldots, n\}$ and, for every finite set S, we denote 2^S to be the set of subsets of S. Figure 1.1 shows Hasse diagrams of the Boolean lattices $\mathcal{B}_2$ and $\mathcal{B}_3$.



$\mathcal{B}_2$ $\mathcal{B}_3$

Figure 1.1 Hasse diagrams of $\mathcal{B}_2$ and $\mathcal{B}_3$.

For posets, $\mathcal{P} = (P, \preceq)$ and $\mathcal{P}' = (P', \preceq)$, we say $\mathcal{P}'$ is a *(weak) subposet* of $\mathcal{P}$ if there exists an injection $f : \mathcal{P}' \to \mathcal{P}$ that preserves the partial ordering. That is, whenever $u \leq' v$ in $\mathcal{P}'$, we have $f(u) \leq f(v)$ in $\mathcal{P}$. If $\mathcal{F}$ is a subposet of $\mathcal{B}_n$ such that $\mathcal{F}$ contains no subposet $\mathcal{P}$, we say $\mathcal{F}$ is $\mathcal{P}$*-free*. The i^{th} *layer* from the bottom of $\mathcal{B}_n$ is the collection of all subsets of $[n]$ of size i. Note that at the i^{th} layer all elements are pairwise disjoint. Let $\mathcal{B}(n, k)$ denote the collection of subsets of $[n]$ of the k middle layers of $\mathcal{B}_n$. If n is a fixed integer, then $\Sigma(n,k) := |\mathcal{B}(n,k)|$, which is the sum of the k largest binomial coefficients of the form $\binom{n}{i}$. To give further insight of the structure of $\mathcal{B}_n$, see Figure 1.2 below, where solid lines in the figure represent the k middle layers of $\mathcal{B}_n$. This diagram is courtesy of Hogenson [29]. A *chain* in a poset is a set of elements in which each pair of elements are pairwise related. The *height* of a poset is the maximum size chain over all chains in the poset.

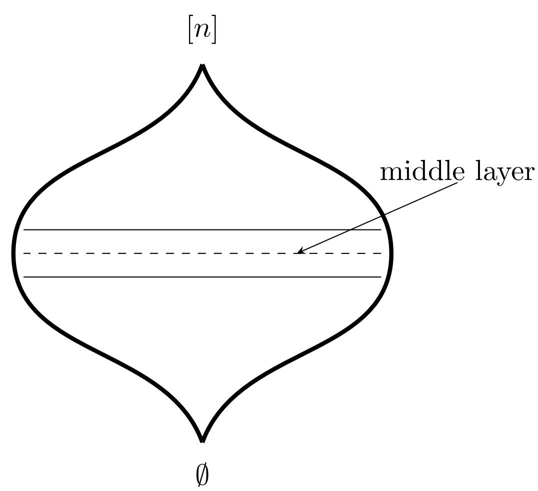


Figure 1.2 The n-dimensional Boolean lattice $\mathcal{B}_n$.

In studying forbidden subposet problems, we want to find the size of a largest family that does not contain $\mathcal{P}$ as a weak subposet of $\mathcal{B}_n$. In 1928, Sperner [45] proved that the size of a largest antichain in $\mathcal{B}_n$ is $\binom{n}{\lfloor n/2 \rfloor}$, where an *antichain* is a chain in which no two elements are pairwise related. To construct an antichain-free family satisfying Sperner's result, choose the middle level of $\mathcal{B}_n$. For n even, this is determined to be the middle level $\binom{[n]}{n/2}$ and for n odd, either of the two middle levels $\binom{[n]}{(n-1)/2}$ or $\binom{[n]}{(n+1)/2}$ work.

In 1945, Erdős [14] generalized Sperner's result to chains. A chain is sometimes called a *path poset* because it appears as a path in its Hasse diagram. Let $\mathcal{P}_k$ denote the path poset on k points. Figure 1.3 shows path posets $\mathcal{P}_2$ and $\mathcal{P}_k$. Erdős proved that the size of the largest path poset, $\mathcal{P}_k$, is $\Sigma(n, k-1) \approx (k-1)\binom{n}{\lfloor n/2 \rfloor}$. Let $\mathrm{La}(n, \mathcal{P})$ denote the size of a largest $\mathcal{P}$-free family in $\mathcal{B}_n$. Using the notation $\mathrm{La}(n, \mathcal{P})$, we formally state the results of Sperner and Erdős.

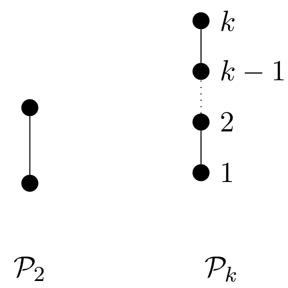


Figure 1.3 Hasse diagrams of $\mathcal{P}_2$ and $\mathcal{P}_k$.

Theorem 1.0.1 (Sperner [45])**.** $\mathrm{La}(n, \mathcal{P}_2) = \binom{n}{\lfloor n/2 \rfloor}$.

Theorem 1.0.2 (Erdős [14])**.** *For* $n \geq k-1 \geq 1$, $\mathrm{La}(n, \mathcal{P}_k) = \sum(n, k-1) \approx (k-1)\binom{n}{\lfloor n/2 \rfloor}$. *Moreover, the* $\mathcal{P}_k$*-free families of maximum size in* $\mathcal{B}_n$ *are given by* $\mathcal{B}(n, k-1)$.

In addition to studying forbidden subposet problems, we study induced-poset-saturation problems.

1.1 Notation and terminology

In this section, we define notation and terminolgy used throughout this work. A *hypergraph* H is a pair $H = (V(H), E(H))$ where $V(H)$ is a set of vertices and $E(H)$ is a set of non-empty subsets of $V(H)$ called hyperedges. The *complement* of a hypergraph H is defined to be the hypergraph with vertex set $V(H)$ whose edge set consists of subsets of $V(H)$ which do not lie in $E(H)$. A *subhypergraph* of a hypergraph H is a hypergraph whose vertex set and edge set consists of a subset (not necessarily proper) of vertices and edges that belong to H. A *spanning subhypergraph* of a hypergraph H is a subhypergraph that contains all vertices in H. An *induced subhypergraph* of H is a hypergraph $H' = (V(H'), E(H'))$, whose vertex set $V(H')$ is a subset of $V(H)$ and ahow hyperedges are completely contained in $V(H')$ form the hyperedge set $E(H')$. The *order* of a hypergraph H is the cardinality of its vertex set, denoted $|V(H)|$, and its *size* is the cardinality of its edge set, denoted $|E(H)|$. We say that a hypergraph is *finite* if it has a finite number of vertices and hyperedges.

Two vertices u and v in a hypergraph are said to be *adjacent* if there is an edge containing u and v. The *neighborhood* of a vertex v, denoted $N(v)$, is the set of all vertices adjacent to it. The *degree* of a vertex v is $|N(v)|$. The *minimum degree of a hypergraph* H, denoted $\delta(H)$, is the minimum of the vertex degrees over all vertices in H. The *maximum degree of a hypergraph* H, denoted $\Delta(H)$, is the maximum of the vertex degrees over all vertices in H. We define the *closed neighborhood* of a vertex v, denoted $N[v]$, to be $N[v] := N(v) \cup \{v\}$. A *k-uniform hypergraph* (*k-graph*) is a hypergraph in which every hyperedge has cardinality k. When $k = 2$, we call H a *graph.*

In a graph (2-graph), we say the vertices that lie in an edge of a graph are *endpoints* of the edge. A *loop* in a graph (2-graph) is an edge in which its two endpoints are the same. *Multiple edges* are edges having the same pair of endpoints. We say a graph is *simple* if it has no loops and no multiple edges.

Let K_n denote the *complete graph* on n vertices, which is the graph in which every vertex is adjacent to all other vertices. A *k-partite hypergraph* is a hypergraph whose vertices are written as a disjoint union of k sets, $V = V_1 \cup V_2 \cup \cdots \cup V_k$, and where no hyperedge contains two elements from the same class V_i for all $1 \leq i \leq k$. We denote K_k^n to be the complete k-partite graph with partite sets of size n.

In this work, many of the results use "$O()$", "$o()$", and "$\Omega()$" notation, which we define as "Big-O", "Little-o", and "Big-Omega", respectively. Let $f(n)$ and $g(n)$ be functions defined on a subset of the real numbers. For C a positive constant and n_0 a positive integer, we say $f(n) = O(g(n))$ if $|f(n)| \leq C \cdot g(n)$ for all $n \geq n_0$. We say $f(n) = o(g(n))$ if for any choice of C there is an n_0 so that $|f(n)| < C \cdot g(n)$ for all $n \geq n_0$. For C a positive constant and n_0 a positive integer, we say $f(n) = \Omega(g(n))$ if $f(n) \geq C \cdot g(n)$ for all $n \geq n_0$. Lastly, if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$, then $f(n) = \Theta(g(n))$.

Other important notations that we use are the following. We use $\subseteq$ to denote a set A is a subset of a set B and $\subset$ to denote that A is a proper subset of B. For two sets A and B, we say $A \triangle B$ is the set of elements that are in either A or B but not both, and we call $A \triangle B$ the symmetric difference of A and B.

We use $\log_2 x$ to denote the logarithm to base 2, $\ln x$ to denote the natural logarithm and e to be base of the natural logarithm.

1.2 Forbidden subposet problems

Since Sperner's and Erdős' results, $\mathcal{P}$-free posets (or $\mathcal{P}$-free families) have been studied extensively by various authors. The *r-fork poset*, denoted $\mathcal{V}_r$, where $r \geq 2$, consists of $k+1$ distinct sets $A, B_1, B_2, \ldots, B_k$ where $A < B_i$ for $1 \leq i \leq k$. See Figure 1.4 for Hasse diagrams of $\mathcal{V}_2$ and $\mathcal{V}_k$. The first results for $\mathrm{La}(n, \mathcal{V}_r)$ were obtained for $r = 2$ by Katona and Tarján in 1983.

Theorem 1.2.1 (Katona and Tarján [33])**.**

$$\binom{n}{\lfloor n/2 \rfloor}\left(1 + \frac{1}{n} + O\left(\frac{1}{n^2}\right)\right) \leq \mathrm{La}(n, \mathcal{V}_2) \leq \binom{n}{\lfloor n/2 \rfloor}\left(1 + \frac{2}{n}\right)$$

Debonis and Katona [11] give an upper bound for the size of the largest $\mathcal{V}_r$-free family, and Thanh [46] gives the lower bound, which is detailed in Theorem 1.2.2.

Theorem 1.2.2 (Thanh [46] and Debonis and Katona [11])**.**

$$\binom{n}{\lfloor n/2 \rfloor}\left(1 + \frac{r-1}{n} + O\left(\frac{1}{n^2}\right)\right) \leq \mathrm{La}(n, \mathcal{V}_r) \leq \binom{n}{\lfloor n/2 \rfloor}\left(1 + 2 \cdot \frac{r-1}{n} + O\left(\frac{1}{n^2}\right)\right)$$

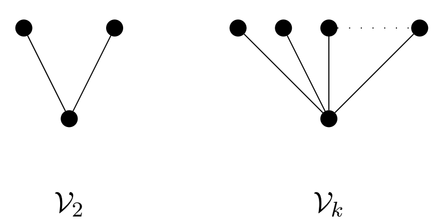


Figure 1.4 Hasse diagrams of the $\mathcal{V}_2$ and $\mathcal{V}_k$ posets.

Notice that in the result of Erdős the asymptotic value of $\mathrm{La}(n, \mathcal{P}_k)/\binom{n}{\lfloor n/2 \rfloor}$ is determined to be the integer $(k-1)$, and in Sperner's result $\mathrm{La}(n, \mathcal{P}_2)/\binom{n}{\lfloor n/2 \rfloor}$ is determined to be the integer 1. In 2009, Griggs and Lu conjectured the following.

Conjecture 1.2.3 (Griggs and Lu [26])**.** *For any finite poset $\mathcal{P}$, the limit $\pi(\mathcal{P}) := \lim_{n\to\infty} \frac{\mathrm{La}(n,\mathcal{P})}{\binom{n}{\lfloor n/2 \rfloor}}$ exists and is an integer.*

Griggs and Lu [26] proved that for any tree poset, $\mathcal{T}$, of height 2, we have $\pi(\mathcal{T}) = 1$. Later, Bukh proved that $\pi(\mathcal{T})$ is an integer for any general tree poset $\mathcal{T}$ [6]. It is believed that Griggs and Lu's conjecture is true, for most posets their conjecture has not been proven.

The most well-studied poset in which the conjecture is not known is the diamond poset. Let the *k-diamond poset*, denoted $\mathcal{D}_k$, where $k \geq 2$, consist of $k+2$ distinct sets $A < B_1, B_2, \ldots, B_k < C$. The 2-diamond poset is just the diamond. See Figure 1.5 for the Hasse diagram of $\mathcal{D}_k$. When $k = 2$, Griggs and Lu [26] made the observation that $\pi(\mathcal{D}_2) \in [2, 2.296]$ and conjectured $\pi(\mathcal{D}_2) = 2$. Table 1.2 shows a list of results for $\pi(\mathcal{D}_2)$. In an effort to resolve $\pi(\mathcal{D}_2)$, several techniques have been used. Griggs, Li, and Lu [25] used the Lubell function method to try and resolve $\pi(\mathcal{D}_2)$, and Axenovich, Martin, and Manske [2] used counting full chains. Other methods used include flag algebras by Kramer, Martin, and Young [36] and cycle decompositions by Grósz, Methuku, and Tompkins [27].

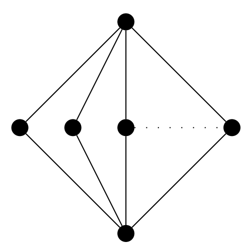
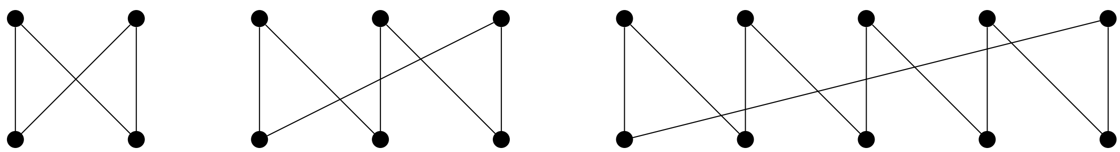


Figure 1.5 The k-diamond poset, $\mathcal{D}_k$.

Table 1.1 $\pi(\mathcal{D}_2)$ results.

$\pi(\mathcal{D}_2)$	Authors	Year
$\pi(\mathcal{D}_2) \leq 2.28327$	Axenovich-Manske-Martin [2]	2012
$\pi(\mathcal{D}_2) \leq 2.27274$	Griggs-Li-Lu [25]	2012
$\pi(\mathcal{D}_2) \leq 2.25$	Kramer-Martin-Young [36]	2013
$\pi(\mathcal{D}_2) \leq 2.20711$	Grósz-Methuku-Tompkins [27]	2018

Like the diamond poset, the crown poset has been extensively studied by various authors. For $t \geq 2$, the *crown poset*, denoted $\mathcal{O}_{2t}$, is a poset of height 2 in which its Hasse diagram is a cycle of length $2t$. When $t = 2$, the poset $\mathcal{O}_4$ is also referred to the as the butterfly poset. See Figure 1.6 for the Hasse diagrams of the butterfly poset, $\mathcal{O}_6$, and $\mathcal{O}_{10}$. Table 1.2 provides a list of results for the size of the largest $\mathcal{O}_{2t}$-free families.



$\mathcal{O}_4$ poset $\qquad$ $\mathcal{O}_6$ poset $\qquad$ $\mathcal{O}_{10}$ poset

Figure 1.6 Hasse diagrams of $\mathcal{O}_4$, $\mathcal{O}_6$, and $\mathcal{O}_{10}$.

Table 1.2 $\mathrm{La}(n, \mathcal{O}_{2t})$ results.

$\mathrm{La}(n, \mathcal{O}_{2t})$	Authors	Year
$\mathrm{La}(n, \mathcal{O}_4) = \sum(n, 2)$	De Bonis-Katona-Swanepoel [12]	2005
$\mathrm{La}(n, \mathcal{O}_{2t}) = (1 + o(1))\binom{n}{\lfloor n/2 \rfloor}$ for even $t \geq 4$	Griggs-Lu [26]	2009
$\mathrm{La}(n, \mathcal{O}_{2t}) \leq (1 + \frac{1}{\sqrt{2}} + o(1))\binom{n}{\lfloor n/2 \rfloor}$ for odd $t \geq 3$	Griggs-Lu [26]	2009
$\mathrm{La}(n, \mathcal{O}_{2t}) = (1 + o(1))\binom{n}{\lfloor n/2 \rfloor}$ for odd $t \geq 7$	Lu [38]	2014

The results of Lu [38] in 2014 resolved $\mathrm{La}(n, \mathcal{O}_{2t})$ for odd $t \geq 7$ by using a result of Conlon [10], We state Conlon's result for completeness. Recall the definitions of a hypergraph, an r-uniform hypergraph, and a k-partite hypergraph from Section 1. Using these definitions, we provide a formal definition of a k-partite representation, which will be used in the theorem provided that leads to resolving $\mathrm{La}(n, \mathcal{O}_{2t})$ asymptotically for all t except $t = 2, 3, 5$.

Definition 1.2.4. *A poset $\mathcal{P}$ of height 2 has a k-*partite representation *if there exist two integers k, ℓ, and a family $\mathcal{F} \subseteq \binom{[\ell]}{k-1} \cup \binom{[\ell]}{k}$ such that*

- *the poset $(\mathcal{F}, \subseteq)$ contains $\mathcal{P}$ as a subposet,*
- *and $H(\mathcal{F}) := \left([\ell],\ \mathcal{F} \cap \binom{[\ell]}{k}\right)$ is an k-uniform k-partite hypergraph.*

For better clarification of a k-partite representation, we provide an example of a 3-partite representation of $\mathcal{O}_{14}$ in Figure 1.7 below.

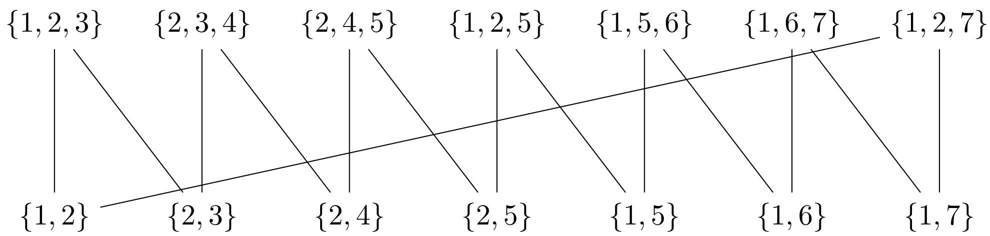


Figure 1.7 3-partite representation of $\mathcal{O}_{14}$.

Now that we have established important definitions regarding Conlon's theorem, we can formally state the results of Conlon and Lu.

Theorem 1.2.5 (Conlon [10])**.** *Except for* $t = 2, 3, 5$ *all crowns* $\mathcal{O}_{2t}$ *have* k*-partite representations for some* k*.*

Theorem 1.2.6 (Lu [38])**.** *Suppose that a poset* $\mathcal{P}$ *of height* 2 *has a* k*-partite representation for some* $k \geq 2$*. Then,* $\mathrm{La}(n, P) = (1 + o(1))\binom{n}{\lfloor n/2 \rfloor}$.

DeBonis, Katona, and Swanepoel [12] solved exactly $\mathrm{La}(n, \mathcal{O}_{2t})$, for $t = 2$. Griggs and Lu [26] resolved $\mathrm{La}(n, \mathcal{O}_{2t})$ for even $t \geq 4$. Thanks to Theorem 1.2.6, all other crowns have been resolved asymptotically with the exception of $\mathrm{La}(n, \mathcal{O}_6)$ and $\mathrm{La}(n, \mathcal{O}_{10})$. For a more extensive survey of the progress on $\mathcal{P}$-free families see Griggs and Li [24].

1.3 Induced-poset-saturation problems

Let G and H be graphs. A spanning subgraph F of G is (G, H)*-saturated* if F contains no copy of H but $F + e$ contains a copy of H for every edge $e \in E(G) \setminus E(F)$. Let $\mathrm{sat}(G, H)$ denote the minimum number of edges in a (G, H)-saturated graph.

Saturation was first introduced independently by Zykov [48] and Erdős, Hajnal, and Moon [15], where the authors considered $\mathrm{sat}(K_n, H)$. Erdős, Hajnal, and Moon showed that for $n \geq k \geq 2$,

$\text{sat}(K_n, K_k) = \binom{n-k+2}{2}$. Since its introduction, saturation has been studied extensively. Let Q_d denote the d-dimensional hypercube. Choi and Guan [9] studied saturation of 4-cycles, Q_2, in the d-dimensional hypercube and proved that $\text{sat}(Q_d, Q_2) \leq \left(\frac{1}{4} + o(1)\right) |E(Q_d)|$. Ferrara et al. [18] studied $\text{sat}(K_k^n, K_t)$. The authors resolved the size of $\text{sat}(K_3^n, K_3)$ for all n and showed the following for $\text{sat}(K_k^n, K_t)$ for $k \geq 3$ and n sufficiently large.

Theorem 1.3.1 (Ferrara et al. [18])**.** *If $k \geq 3$ and $n \geq 100$, then*

$$\text{sat}(K_k^n, K_3) = \min\{2kn + n^2 - 4k - 1, 3kn - 3n - 6\}.$$

Faudree, Faudree, and Schmitt provide a survey of saturated graphs in [17].

Definition 1.3.2. *Given a host poset $\mathcal{Q} = (Q, \leq_{\mathcal{Q}})$ and target poset $\mathcal{P} = (P, \leq_{\mathcal{P}})$ and a family $\mathcal{F} \subseteq Q$, we say that $\mathcal{F}$ is* $\mathcal{P}$-saturated *in $\mathcal{Q}$ if the following two properties hold:*

- *$\mathcal{P}$ is not a subposet of the poset induced by $\mathcal{F}$ in $\mathcal{Q}$, and*
- *for any $S \in \mathcal{Q} - \mathcal{F}$, the poset $\mathcal{P}$ is a subposet of the poset induced by $\mathcal{F} \cup \{S\}$ in $\mathcal{Q}$.*

Given $n > 0$ and a poset $\mathcal{P}$, define $\text{sat}(n, \mathcal{P})$ as the minimum size of a $\mathcal{P}$-saturated family of $\mathcal{B}_n$. We say that $\text{sat}(n, \mathcal{P})$ is the *saturation number of* $\mathcal{P}$. It is worth noting that studying the maximum size of a $\mathcal{P}$-saturated family is defined in Section 1.2 as $\text{La}(n, \mathcal{P})$.

Saturation number in relation to posets were first introduced by Gerbner et al. [21]. They looked at the the path poset on k points, $\mathcal{P}_k$. The authors proved the following theorem using a construction technique and then bounding the size of $\text{sat}(n, \mathcal{P}_{k+1})$ by $2 \cdot \text{sat}(n-1, \mathcal{P}_k)$.

Theorem 1.3.3 (Gerbner et al. [21])**.** *For n sufficiently large,*

$$2^{k/2-1} \leq \text{sat}(n, \mathcal{P}_{k+1}) \leq 2^{k-1}.$$

Morrison, Noel, and Scott [41] improved the upper bound using a similar construction to Gerbner et al.

Theorem 1.3.4 (Morrison, Noel, Scott [41])**.** *There exists $\varepsilon > 0$ such that for all $k > 0$ and n sufficiently large,*

$$\text{sat}(n, \mathcal{P}_{k+1}) \leq 2^{(1-\varepsilon)k}.$$

We say $\mathcal{P}'$ is an *induced-subposet* of $\mathcal{P}$ if there exists an injective function $f : \mathcal{P}' \to \mathcal{P}$ such that $u \leq' v$ if and only if $f(u) \leq f(v)$. For any set $\mathcal{F} \subseteq \mathcal{P}$, the *poset induced by $\mathcal{F}$ in $\mathcal{P}$* is the induced subposet $(\mathcal{F}, \leq')$ of $\mathcal{P}$ where $\leq$ restricts $\leq'$ to $\mathcal{F}$.

Definition 1.3.5. *Given a poset host $\mathcal{Q} = (Q, \leq_{\mathcal{Q}})$ and target poset $\mathcal{P} = (P, \leq_{\mathcal{P}})$ and a family $\mathcal{F} \subseteq Q$, we say that $\mathcal{F}$ is* induced-$\mathcal{P}$-saturated *in $\mathcal{Q}$ if the following two properties hold:*

- *The poset induced by $\mathcal{F}$ in $\mathcal{Q}$ does not contain an induced copy of $\mathcal{P}$, and*
- *for every $S \in \mathcal{Q} - \mathcal{F}$, the poset induced by $\mathcal{F} \cup \{S\}$ in $\mathcal{Q}$ contains an induced copy of $\mathcal{P}$.*

Given $n > 0$ and a poset $\mathcal{P}$, define $\text{sat}^*(n, \mathcal{P})$ as the minimum size of a family $\mathcal{F}$ that is induced-$\mathcal{P}$-saturated in $\mathcal{B}_n$. We say $\text{sat}^*(n, \mathcal{P})$ is the *induced saturation number* of $\mathcal{P}$.

Ferrara et al. [19] introduced and studied the concept of $\text{sat}^*(n, \mathcal{P})$ for various posets $\mathcal{P}$. Our interest lies in studying induced saturation of the k-antichain $\mathcal{A}_k$. Let $\mathcal{A}_k$ denote the *k-antichain*, which is the poset of k elements in which no pair of elements is comparable. Ferrara et al. proved the following result regarding induced-$\mathcal{A}_{k+1}$-saturation.

Theorem 1.3.6 (Ferrara et al. [19])**.** *If $n > k > 1$, then*

$$2n \leq \text{sat}^*(n, \mathcal{A}_{k+1}) \leq (n-1)k - \left(\frac{1}{2} \log_2 k + \frac{1}{2} \log_2 \log_2 k + O(1) \right).$$

In particular, $\text{sat}^*(n, \mathcal{A}_{k+1}) = \Theta(n)$.

In Chapter 3, we make improvements to the lower bound of Theorem 1.3.6. The next theorem provides induced-$\mathcal{P}$-saturation results for various posets we defined in Section 1.2.

Theorem 1.3.7 (Ferrara et al. [19])**.** *For a given poset $\mathcal{P}$, the following are results of* $\text{sat}^*(n, \mathcal{P})$*:*

1. *If $n \geq 2$, then* $\text{sat}^*(n, \mathcal{V}_2) = n + 1$.
2. *If $n \geq 2$, then* $\lceil \log_2 n \rceil \leq \text{sat}^*(n, \mathcal{D}_2) \leq n + 1$.
3. *If $n \geq 3$, then* $\lceil \log_2 n \rceil \leq \text{sat}^*(n, \mathcal{O}_6) \leq \binom{n}{2} + 2n - 1$.

1.4 Vertex-identifying codes in graphs

Another problem that we give attention to is vertex-identifying codes in graphs. Vertex-identifying codes were first introduced in 1998 by Karpovsky, Chakrabarty, and Levitin [31]. For any nonnegative integer r, the *ball of radius r around vertex $v \in V$*, denoted $B_r(v)$, is the set of all vertices of distance at most r from v. When $r = 1$, $B_1(v) := N[v]$ where $N[v]$ is the closed neighborhood of a graph.

Definition 1.4.1. *For any positive integer r and finite graph G, an r-vertex-identifying code (r-VI code) in G is a subset $C \subseteq V(G)$, with the property that*

- $B_r(v) \cap C \neq \emptyset$, *for all $v \in V(G)$; and*
- $B_r(u) \cap C \neq B_r(v) \cap C$, *for all distinct $u, v \in V(G)$.*

If $r = 1$, we simply call C a VI code in G.

A nonempty set C of V is called a *code* and its elements are called *codewords*.

While vertex-identifying codes originated by Karpovsky et al. [31], origins of vertex identifying codes can be found in combinatorial search theory by Katona, see [32]. In combinatorial search theory, the primary goal is to find one or more defected elements from a set of n using the minimum number of queries possible from the set of queries allowed. This goal is similar to the goal of finding the minimum r-VI code. A basic lower bound for r-VI codes is the following:

Theorem 1.4.2 (Karpovsky et al. [31])**.** *Let G be a graph on n vertices such that $|B_r(v)| \leq \beta$ for all $v \in V(G)$ and let C be an r-VI code in G. Then,*

$$|C| \geq \max\left\{\lceil \log_2(n+1) \rceil, \left\lceil \frac{2n}{\beta+1} \right\rceil\right\}.$$

Theorem 1.4.3 gives a lower bound on the sizes of r-VI codes, provided the ball of radius r of every vertex is not too large. Entropy methods are used to establish the lower bound. The binary entropy function is $H_2(x) = -x\log_2 x - (1-x)\log_2(1-x)$.

Theorem 1.4.3 (Katona [32]). *Let G be a graph on n vertices such that $|B_r(v)| \leq \beta \leq n/2$ for all $v \in V(G)$ and let C be an r-VI code in G. Then,*

$$|C| \geq \frac{\log_2 n}{H_2(\beta/n)} \tag{1.1}$$
$$\geq \frac{n}{\beta} \cdot \frac{\log_2 n}{\log_2 (en/\beta)}. \tag{1.2}$$

Theorem 1.4.3 was discovered by us without knowledge of Katona's result. Also, Karpovsky et al. did not cite Katona's result in [31].

The result of Theorem 1.4.3 holds for $|B_r(v)| \geq \beta \geq n/2$. Note that bound (1.2) in Theorem 1.4.3 improves upon Karpovsky et al.'s bound $\frac{2n}{\beta+1}$ whenever $\beta \geq e\sqrt{n}$. Using the bound $H_2(p) \leq 1 - \frac{2}{\ln 2}\left(p - \frac{1}{2}\right)^2$, we can see that bound (1.1) improves upon Karpovsky et al.'s $\log_2(n+1)$ bound whenever $\beta \geq \frac{n}{2} - \sqrt{\frac{n \ln 2}{2}}$. Hence, bound (1.1) is at least as good as the bound in Theorem 1.4.2 for $\beta \in \left[e\sqrt{n}, \frac{n}{2} - \sqrt{\frac{n \ln 2}{2}}\right]$.

Moncel [40] determined all optimal graphs satisfying $|C| = \lceil \log_2(n+1) \rceil$. Charon, Hudry, and Lobstein [7] proved that minimizing the size of a vertex-identifying code is NP-hard. The authors also proved in [8] that the minimum size of a vertex identifying code is at most $n-1$. Kim et al. [34] studied a slight variant of vertex-identifying codes, in which they removed the restriction that the balls of radius r must cover all of the vertices of G.

One observation to make is that it may not be possible for an r-VI code to exist due to having two vertices u and v such that $B_r(u) = B_r(v)$. One simple example of a graph not having a vertex-identifying code is K_n, the complete graph on n vertices. See Lobstein [37] for a bibliography of results relating to vertex-identifying codes and other types of codes being studied.

1.5 Organization of dissertation

This dissertation is organized in the following manner. In Chapter 2, the $\mathcal{N}$ poset is defined and known upper and lower bounds for $\mathrm{La}(n, \mathcal{N})$ are given. We provide a potential improvement to the known lower bound result $\mathrm{La}(n, \mathcal{N})$ in Theorem 2.2.3. We also focus our attention on studying the size of $\mathcal{O}_6$-free families on two consecutive layers, which is the motivation behind the result of Theorem 2.3.1. In Chapter 3, we study $\text{sat}^*(n, \mathcal{A}_{k+1})$ and improve the result of Ferrara et al. (Theorem 1.3.6) in Theorem 3.1.2. In Chapter 4, two types of graphs are studied and lower bound results for r-VI codes are given for both. Theorem 4.2.1 gives a lower bound for a r-VI code in graphs with large symmetric differences. In particular, a lower bound for strongly-regular graphs are given in Corollary 4.2.3. Theorem 4.3.6 gives a lower bound for a 1-VI code for (p, β)-jumbled graphs. This work is concluded by providing a summary of results and future directions of research in Chapter 5.

CHAPTER 2. FORBIDDEN SUBPOSET RESULTS

Recall from Chapter 1.2 that for a poset $\mathcal{P}$, we define $\mathrm{La}(n,\mathcal{P})$ to be the size of a largest $\mathcal{P}$-free family in the n-dimensional Boolean Lattice, $\mathcal{B}_n$. In this chapter, we study the $\mathcal{N}$ poset and the $\mathcal{O}_6$ poset. We provide a lower bound result for $\mathrm{La}(n,\mathcal{N})$ in Theorem 2.2.3 in Section 2.2. The result for $\mathrm{La}(n,\mathcal{N})$ is work published in the paper "A note on the size of $\mathcal{N}$-free families" [39], published jointly with Ryan R. Martin in the *European Journal of Mathematics*. In studying the $\mathcal{O}_6$ poset on two consecutive levels of $\mathcal{B}_n$, we define a *triangle* as it relates to $\mathcal{O}_6$ and provide lower and upper bound results for the size of *triangle*-free families in Theorem 2.3.1 in Section 2.3. In an effort to prove Theorem 2.2.3 and Theorem 2.3.1 we use some coding theory.

2.1 Coding theory background

We get most of our basic definitions regarding coding theory from the *Handbook of Discrete and Combinatorial Mathematics, Chapter 14* [43]. A *binary word* is a $\{0,1\}$-vector of length n. A *binary code of length n*, say C, is a subset of all binary words of length n. An element of C is called a *codeword*. If $|C| = m$, then C is of *order m*. The *weight* of a codeword is the number of ones in the codeword. The *Hamming distance between two codewords* of equal length is the number of positions at which the corresponding entries differ. The *Hamming distance of a code* is the minimum Hamming distance over all pairs of codewords in that code.

Let $A(n,2\delta,k)$ denote the size of the largest family of $\{0,1\}$-vectors of length n such that each vector has exactly k ones and the Hamming distance between any pair of distinct vectors is at least 2δ. This is the same as the size of the largest family of subsets of $[n]$ such that each subset has size exactly k and the symmetric difference of any pair of distinct sets is at least 2δ.

The quantity $A(n,2\delta,k)$ is important in the field of error-correcting codes. In fact, $A(n,4,k)$ computes the size of a single-error-correcting code with constant weight k. Henceforth, we will use

"SEC code" as shorthand for "single-error-correcting code."

The first nontrivial value of δ for $A(n, 2\delta, k)$ is $\delta = 2$. Graham and Sloane [22] give a lower bound construction for $A(n, 4, k)$ and Johnson [30] gives an upper bound for $A(n, 4, k)$.

Theorem 2.1.1 (Graham and Sloane [22])**.** $A(n, 4, k) \geq \frac{1}{n}\binom{n}{k}$.

Theorem 2.1.2 (Johnson [30])**.** $A(n, 2\delta, k) \leq \frac{\binom{n}{k-\delta+1}}{\binom{k}{k-\delta+1}}$.

In particular, for $\delta = 2$ we have

$$A(n, 4, k) \leq \frac{\binom{n}{k-1}}{\binom{k}{k-1}} = \frac{1}{k}\binom{n}{k-1} = \frac{1}{n-k+1}\binom{n}{k}.$$

2.2 The size of $\mathcal{N}$-free families

The main result of this Section is Theorem 2.2.3, in which for some values of n, we potentially improve the bounds of $\mathrm{La}(n, \mathcal{N})$ in the second order term. The poset $\mathcal{N}$ consists of four distinct sets W, X, Y, Z such that $W \subset X$, $Y \subset X$, and $Y \subset Z$. However, W is not necessarily a subset of Z. For the Hasse diagram of the $\mathcal{N}$ poset, see Figure 2.1. Griggs and Katona [23] addressed $\mathcal{N}$-free families, obtaining Theorem 2.2.1 below.

Theorem 2.2.1 (Griggs and Katona [23])**.**

$$\binom{n}{\lfloor n/2 \rfloor}\left(1 + \frac{1}{n} + \Omega\left(\frac{1}{n^2}\right)\right) \leq \mathrm{La}(n, \mathcal{N}) \leq \binom{n}{\lfloor n/2 \rfloor}\left(1 + \frac{2}{n} + O\left(\frac{1}{n^2}\right)\right).$$

The construction for the lower bound of Theorem 2.2.1 comes directly from Theorem 1.2.1 since any $\mathcal{V}_2$-free family is also $\mathcal{N}$-free.

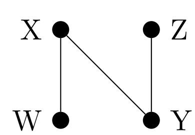


Figure 2.1 Hasse diagram of the $\mathcal{N}$ poset.

To establish the lower bound, Katona and Tarján used a constant-weight code construction due to Graham and Sloane [22] from 1980, which is referenced in Theorem 2.1.1. In the proof of

Theorem 2.2.3, we obtain a lower bound that appears to be larger than the current known bound. However, whether it is an improvement depends on the behavior of some functions well-known in coding theory.

2.2.1 $\mathcal{N}$-free lower bound result

Katona and Tarján [33] estimated the following lower bound for $\mathcal{N}$-free families.

Theorem 2.2.2 (Katona and Tarján [33]). *Let $k = \lfloor n/2 \rfloor$. Then,*

$$\mathrm{La}(n, \mathcal{N}) \geq \binom{n}{k} + A(n, 4, k+1).$$

The following theorem is our potential improvement of Katona and Tarján's result for the case when n is even.

Theorem 2.2.3 (Martin and -W. [39]). *Let n be even and let $k = n/2$. Then,*

$$\mathrm{La}(n, \mathcal{N}) \geq \binom{n}{k} + A(n, 4, k). \tag{2.1}$$

Remark 2.2.4. *We note that the same construction works for n odd and $k = (n-1)/2$. This gives a family of size $\binom{n}{k} + A(n, 4, k)$. However, since $A(n, 4, k) = A(n, 4, k+1)$ in the odd case, this does not provide an improvement to the known bounds.*

To prove Theorem 2.2.3, we begin with a constant weight code C and provide an upset and downset of C. We then prove a claim which details some important properties of the upset and downset of C.

Proof of Theorem 2.2.3.

Given $k = n/2$, let C be a constant weight SEC code of size $A(n, 4, k)$. Define $C_{\mathrm{up}} := \{c \cup \{i\} : c \in C,\, i \notin c\}$ and $C_{\mathrm{down}} := \{c - \{i\} : c \in C,\, i \in c\}$. Claim 2.2.5 gives some important properties of $C_{\mathrm{up}} \cup C_{\mathrm{down}}$.

Claim 2.2.5.

(i) Both C_{up} and C_{down} are SEC codes with constant weight $k+1$ and $k-1$, respectively.

(ii) If $c'' \in C_{\text{up}}$ and $c' \in C_{\text{down}}$, $c' \not\subseteq c''$.

Proof. *(i).* Let $c_1, c_2 \in C_{\text{up}}$. Then $|c_1 \triangle c_2| = |(c_1 - \{i\}) \triangle (c_2 - \{i\})| \geq 4$ since $(c_1 - \{i\}) \in C$, $(c_2 - \{i\}) \in C$, and their symmetric difference must be at least 4 in order for C to be a SEC code. Thus, C_{up} is a SEC code. By a similar argument, C_{down} is a SEC code.

(ii). Let $c'' \in C_{\text{up}}$, $c' \in C_{\text{down}}$, and $c' \subset c''$. Then, $(c' \cup \{i\}), (c'' - \{i\}) \in C$. So, $|(c'' - \{i\}) \triangle (c' \cup \{i\})| \geq 4$. This implies that there are two members of $[n]$ that are in $(c' \cup \{i\}) - (c'' - \{i\})$. One is i and the other is some $j \in c' - c''$, which contradicts the assumption that $c' \subset c''$. This concludes the proof of Claim 2.2.5. □

In order to finish the proof, we just need to show that the family $\mathcal{F} := \binom{[n]}{k} \cup C_{\text{up}} \cup C_{\text{down}}$ is $\mathcal{N}$-free.

To that end, suppose there is a subposet $\mathcal{N}$ with elements W, X, Y, Z where $W \subset X$, $Y \subset X$ and $Y \subset Z$ (see Figure 2.1). Where is the element X?

We know that $X \notin C_{\text{down}}$ because it has to have elements below it and the elements of C_{down} are all minimal in $\mathcal{F}$. We know that $X \notin \binom{[n]}{k}$ because that would force $W, Y \in C_{\text{down}}$ and, being subsets of X would require $|W \triangle Y| = 2$, a contradiction to C_{down} being a SEC code. Therefore, $X \in C_{\text{up}}$.

Now, where is Y? We know that $Y \notin C_{\text{up}}$ because $Y \subset X$. We know $Y \notin \binom{[n]}{k}$ because that would force $X, Z \in C_{\text{up}}$ and thus would force $|X \triangle Z| = 2$, this is a contradiction to the fact that C_{up} is a SEC code. Therefore, $Y \in C_{\text{down}}$.

In order for the copy of $\mathcal{N}$ to exist, $Y \subset X$, which implies $Y \subset X - \{i\}$ and so $|(Y \cup \{i\}) \triangle (X - \{i\})| = 2$. Recall, however, that $Y \cup \{i\}$ and $X - \{i\}$ are distinct members of C and so have symmetric difference at least 4, a contradiction. □

Remark 2.2.6. *We believe that, for $n \geq 6$, the quantity $A(n, 4, k)$ is strictly unimodal as a function of k as long as $3 \leq k \leq n - 3$. This strict unimodality has been established [5] for $6 \leq n \leq 12$ and known bounds suggest that it is the case for larger values of n as well. If unimodality holds, then $A(n, 4, k)$ would achieve its maximum uniquely at $k = \lfloor n/2 \rfloor$ or $k = \lceil n/2 \rceil$. Therefore, we expect*

(2.1) to also be a strict improvement over Theorem 2.2.2 in the case where n is even. However, to our knowledge, the unimodality of $A(n, 4, k)$ has never been established and seems to be a highly nontrivial problem.

2.3 Triangle-free and generalized r-tight cliques

In an effort to make progress on $\mathrm{La}(n, \mathcal{O}_6)$, we focused our study to $\mathcal{O}_6$-free families on two consecutive levels. By studying the structure of $\mathcal{O}_6$, we found that if two consecutive levels of $\mathcal{B}_n$ have a triangle, then it has $\mathcal{O}_6$ as a subposet. We define a *triangle* to be a collection of three sets of size k from $[n]$ such that the common intersection of the three k-sets is of size $k-2$ and the pairwise intersection of any two of the k-sets is of size $k-1$. Using this definition of a triangle we obtain the following result.

Theorem 2.3.1. *Let $\mathcal{T}$ be a k-uniform hypergraph on n vertices. Then, there exists a triangle-free family $\mathcal{T}$ of subsets of $[n]$ such that $\frac{2}{n}\binom{n}{k} \leq |\mathcal{T}|$, and for every triangle-free family $\mathcal{T}$ of subsets of $[n]$, $|\mathcal{T}| \leq \frac{2}{k+1}\binom{n}{k}$.*

The proof of the lower bound is nearly identical to Graham and Sloane's proof of Theorem 2.1.1 and we provide its proof for completeness.

Proof of Theorem 2.3.1.

Note that by representing each k-set of a triangle as a binary code of length n with constant weight k, the Hamming distance between any two members of a triangle is 2. Let $\mathbb{F}_k^n$ denote the set of $\binom{n}{k}$ binary vectors of length n and weight k and let $\mathbb{Z}_n$ be the residue class modulo n. Consider the map $T : \mathbb{F}_k^n \to \mathbb{Z}_n$ where

$$T(a) = \sum_{i=0}^{n-1} i a_i \pmod{n}$$

for $a = (a_0, a_1, \ldots, a_{n-1}) \in \mathbb{F}_k^n$ as defined in [22].

Let C_i be the constant weight code $T^{-1}(i)$ for $0 \leq i \leq n-1$. For any two constant weight codes C_i and C_j, we claim that there is no triangle in $\mathcal{T} := C_i \cup C_j$. Let c be any codeword in C_i

and a and b be any two codewords in C_j. We prove that the Hamming distance between a and b is at least four. Suppose that the Hamming distance is two. Having constant weight k, implies that both a and b agree in all positions except two. Without loss of generality, let these two positions be the r-th and s-th positions.

Claim 2.3.2. *Without loss of generality, the r-th position of a is one and the s-th position of b is one.*

Proof. The r-th and s-th positions of a cannot both be one, because then a must have constant weight $k+2$ since the r-th and s-th positions of b are both zero and b has constant weight k and the two differ in exactly two positions. Likewise, the r-th and s-th positions of b cannot both be one. Thus, we may assume that in the r-th position of a there is a one and in the s-th position of b there is a one. □

Note that $T(a) = T(b) = j \pmod{n}$ since a and b are codewords of C_j. By definition, we have $T(a) = x + r = j \pmod{n}$ and $T(b) = x + s = j \pmod{n}$ for some $x \in \mathbb{Z}_n$. This implies that $r \equiv s \pmod{n}$, which is not feasible. Therefore, a and b have Hamming distance at least four and C_j has a Hamming distance of at least four between any two of its codewords. Thus, there is no triangle in $\mathcal{T}$. Consider the two largest constant weight codes in cardinality, say C_i and C_j. Note that $|C_i| + |C_j| \geq \frac{2}{n}\binom{n}{k}$. Therefore, $\frac{2}{n}\binom{n}{k} \leq |\mathcal{T}|$ as desired. This completes the proof of the lower bound.

To prove the upper bound, let $\mathcal{T} \subseteq \binom{[n]}{k}$ and $\mathcal{G} = \binom{[n]}{k+1}$. We prove $(n-k)|\mathcal{T}| \leq e(\mathcal{G}, \mathcal{T}) \leq 2\binom{n}{k+1}$ where $e(\mathcal{G}, \mathcal{T})$ is the maximum number of edges between $\mathcal{G}$ and $\mathcal{T}$ such that there is no triangle. To prove the lower bound, let $S \in \mathcal{T}$. It is clear that the number of sets that S is a subset of in $\mathcal{G}$ is $(n-k)$. Thus, $(n-k)|\mathcal{T}| \leq e(\mathcal{G}, \mathcal{T})$. For the upper bound, we claim that for all $R \in \mathcal{G}$ there exist at most two elements of $\mathcal{T}$ that are subsets of R. Without loss of generality, let $R := \{1, 2, \ldots, k, k+1\}$ and let the three elements of $\mathcal{T}$ be $S_1 := \{1, 2, \ldots, k\}$, $S_2 := \{1, 2, \ldots, k-1, k+1\}$, and $S_3 := \{1, 2, \ldots, k-2, k, k+1\}$. Then we have that $|S_1 \cap S_2 \cap S_3| = k-2$ and $|S_i \cap S_j| = k-1$ for $1 \leq i < j \leq 3$. The three sets S_1, S_2, and S_3 form a triangle, so there can be at most two elements of $\mathcal{T}$ that are subsets of R. Hence, $e(\mathcal{G},\mathcal{T}) \leq 2|\mathcal{G}| = 2\binom{n}{k+1}$. Therefore, $|\mathcal{T}| \leq \frac{2}{n-k}\binom{n}{k+1} = \frac{2}{k+1}\binom{n}{k}$ which proves the upper bound $|\mathcal{T}| \leq \frac{2}{k+1}\binom{n}{k}$. □

Remark 2.3.3. *When $k > n/2$, our upper bound is a sharper improvement of Johnson's [30] upper bound.*

2.3.1 Generalized r-tight cliques

In a natural way, we extend the results for triangle-free families to generalized r-tight clique families. We say that a *generalized r-tight clique* is defined by the sets $A_1, A_2, \ldots, A_{r+1}$ of size k such that if $S = \bigcap_{i=1}^{r+1} A_i$, then $A_1 - S, A_2 - S, \ldots, A_{r+1} - S$ is a K_{r+1}^r, where K_{r+1}^r is the complete r-graph on $r+1$ vertices. Using the definition of a generalized r-tight clique, we can now extend Theorem 2.3.1 to the following:

Theorem 2.3.4. *Let $\mathcal{F}$ be a family of k-uniform hypergraph on n vertices. There exists a generalized r-tight clique free family $\mathcal{F}$ such that $\frac{r}{n}\binom{n}{k} \leq |\mathcal{F}|$, and for every generalized r-tight clique free family, $\mathcal{F}$, $|\mathcal{F}| \leq \frac{r}{k+1}\binom{n}{k}$.*

For completeness, we provide the proof of Theorem 2.3.4, which follows straightforward from the proof of Theorem 2.3.1.

Proof of Theorem 2.3.4.

To prove the lower bound, represent each k-set of a generalized r-tight clique as a binary code of length n with constant weight k. The Hamming distance between any two members of a generalized r-tight clique is two. Let $\mathbb{F}_k^n$ denote the set of $\binom{n}{k}$ binary vectors of length n and weight k and let $\mathbb{Z}_n$ be the residue class modulo n. Consider the map $T : \mathbb{F}_k^n \to \mathbb{Z}_n$ where

$$T(a) = \sum_{i=0}^{n-1} ia_i \pmod{n}$$

for $a = (a_0, a_1, \ldots, a_{n-1}) \in \mathbb{F}_k^n$.

Let C_i be the constant weight code $T^{-1}(i)$ for $0 \leq i \leq n-1$. For any r constant weight codes, we claim that there is no generalized r-tight clique in $\mathcal{F} := \bigcup_{i=1}^{r} C_{\sigma_i}$ where $\sigma_i \in \{0, 1 \ldots, n-1\}$. To prove this, we show that for any code C_j, any two codewords in C_j must have Hamming distance of at least four. Let a and b be any two codewords in C_j. Suppose that the Hamming distance is two. Having constant weight k, implies that both a and b agree in all positions except two. Let these two positions be the r-th and s-th positions, where in the r-th position of a there is a one and in the s-th position of b there is a one (for reference see Claim 2.3.2). Note that $T(a) = T(b) = j \pmod{n}$ and by definition, we have $T(a) = x + r = j \pmod{n}$ and $T(b) = x + s = j \pmod{n}$ for some $x \in \mathbb{Z}_n$. This implies that $r \equiv s \pmod{n}$, which is not feasible. Therefore, a and b have Hamming distance at least four and C_j has Hamming distance of at least four between any two of its codewords. Thus, there is no generalized r-tight clique in $\mathcal{F}$. Consider the r largest constant weight codes in cardinality, without loss of generality, say $C_1, C_2, \ldots. C_r$. Note that $|C_1| + |C_2| + \cdots + |C_r| \geq \frac{r}{n}\binom{n}{k}$. Therefore, $\frac{r}{n}\binom{n}{k} \leq |\mathcal{F}|$ as desired.

For the upper bound, let $\mathcal{F} \subseteq \binom{[n]}{k}$ and $\mathcal{G} = \binom{[n]}{k+1}$. We prove $(n-k)|\mathcal{F}| \leq e(\mathcal{G}, \mathcal{F}) \leq r\binom{n}{k+1}$ where $e(\mathcal{G}, \mathcal{F})$ is the maximum number of edges between $\mathcal{G}$ and $\mathcal{F}$ such that there is no generalized r-tight clique. To prove the lower bound, let $S \in \mathcal{F}$. The number of sets for which S is a subset in $\mathcal{G}$ is $(n-k)$. Thus, $(n-k)|\mathcal{F}| \leq e(\mathcal{G}, \mathcal{F})$. For the upper bound, suppose without loss of generality, $R = \{1, 2, \ldots, k, k+1\}$. Suppose there exists $r+1$ elements of $\mathcal{F}$ that are subsets of R. Define these elements to be $A_i := S_i \cup \{r+2, r+3, \ldots, k+1\}$, where r fixed and k significantly larger than r, and $S_i := [r+1] \backslash \{i\}$ for $1 \leq i \leq r+1$. We claim that for all $R \in \mathcal{G}$ there exists at most r elements of the A_i's of $\mathcal{F}$ that are subsets of R. Then, these $r+1$ elements form a generalized r-tight clique since $M := \cap_{i=1}^{r+1} A_i = \{r+2, r+3, \ldots, k+1\}$ and $A_i - M = S_i$ and the set of S_i's form a K_{r+1}^{r}. Thus, there can be at most r elements of $\mathcal{F}$ of the form A_i for $1 \leq i \leq r+1$ that are subsets of R. Hence, $e(\mathcal{G}, \mathcal{F}) \leq r|\mathcal{G}| = r\binom{n}{k+1}$. Therefore, $|\mathcal{F}| \leq \frac{r}{n-k}\binom{n}{k+1} = \frac{r}{k+1}\binom{n}{k}$ as desired. $\square$

CHAPTER 3. INDUCED-$\mathcal{A}_{k+1}$-SATURATION THEOREM

In this chapter, we provide a lower bound result regarding induced-$\mathcal{A}_{k+1}$-saturated families, which improves the result given in Ferrara et al. [19]. Recall that $\text{sat}^*(n, \mathcal{A}_{k+1})$ is the size of the smallest induced-$\mathcal{A}_{k+1}$-saturated family in the n-dimensional Boolean lattice, $\mathcal{B}_n$, and Theorem 1.3.6 states that

$$2n \leq \text{sat}^*(n, \mathcal{A}_{k+1}) \leq (n-1)k - \left(\frac{1}{2}\log_2 k + \frac{1}{2}\log_2\log_2 k + O(1)\right).$$

Theorem 3.1.2 gives two lower bounds for $\text{sat}^*(n, \mathcal{A}_{k+1})$, both of which are $\Omega(kn/\log k)$. We then determine which of the two lower bounds is larger for each k and n sufficiently large in Proposition 3.2.1 in Section 3.2.

3.1 Induced-$\mathcal{A}_{k+1}$-saturated results

The following theorem by Dilworth is an important theorem relating posets and antichains and will be used to prove Theorem 3.1.2.

Theorem 3.1.1 (Dilworth [13])**.** *Let $\mathcal{P} = (P, \leq)$ be a (finite) poset. If k is the size of the largest antichain in $\mathcal{P}$, then there exists disjoint chains $C_1, C_2, \ldots, C_k$ such that $\mathcal{P} = C_1 \cup C_2 \cup \cdots \cup C_k$.*

Theorem 3.1.2. *Let $k \geq 3$ be an integer and let $\mathcal{A}_{k+1}$ be an antichain of size $k+1$. Then for all $n \geq k$,*

(a) $\text{sat}^*(n, \mathcal{A}_{k+1}) \geq k\left\lceil \frac{n}{\lfloor \log_2 k \rfloor + 1} \right\rceil - k + 2.$

(b) $\text{sat}^*(n, \mathcal{A}_{k+1}) \geq 2n + \sum_{j=3}^{k} \left\lceil \frac{n}{d^*(j)} \right\rceil - k + 2$ *where $d^*(j)$ is the largest d such that $\binom{d}{\lfloor d/2 \rfloor} \leq j - 1$.*

Proof of Theorem 3.1.2.

Let $\mathcal{F}$ be an induced-$\mathcal{A}_{k+1}$-saturated family in $\mathcal{B}_n$. It is clear that $\mathcal{F}$ contains an antichain of size k, otherwise it cannot be induced-$\mathcal{A}_{k+1}$-saturated. Both $\emptyset$ and $[n]$ lie in $\mathcal{F}$. By Dilworth's theorem (Theorem 3.1.1), $\mathcal{F}$ can be partitioned into k disjoint chains $C_1, C_2, \ldots, C_k$. Add $\emptyset$ and $[n]$ to each C_i. Color each member of $\mathcal{F} \backslash \{\emptyset, [n]\}$ according to the chain to which it belongs. Color the elements $\emptyset$ and $[n]$ with all colors. Let C_i be a member of the chain partition and $x, y, z \in \mathcal{F}$ with $x, y \in C_i$ and $z \notin C_i$ such that there is no member of C_i other than x and y in the open interval $(x, y) := \{z : x \subset z \subset y\}$. Then we call (x, y) the *gap* between x and y and define the *size* of the gap to be $|y - x|$. Let d be the maximum gap size between any two elements in a chain C_i. Then the number of elements in C_i is at least $\lceil n/d \rceil - 1$.

For part (a)**:** Consider a maximum gap (x, y) of size $\bar{d}$ in C_j where x and y are the bottom and top elements in the gap, respectively. Suppose there are three elements, $a \subset b \subset c$, consecutive in C_i, for $i \neq j$, in the gap (x, y). Let c_i be the color of the elements of C_i and c_j be the color of the elements of C_j, respectively. Let d_1 be the size of gap (b, y) and d_2 be the size of gap (x, b). By recoloring b to color c_j, all gaps will remain the same size or decrease with the exception that C_i now has a gap (a, c) of size $|c - a| \leq \bar{d} - 2$. Hence, we can decrease the size of the gap (x, y) in C_j by adding element b to C_j.

We show that this procedure terminates because there are a finite number of gaps. By operating on all gaps of size $\bar{d}$, we decrease the size of all such gaps by at least 2. The argument terminates by induction on the size of the largest gap.

Every gap must have at least one color. In any gap, there can be at most two elements with the same color. So, we have $2(k-1) \geq 2^{\bar{d}} - 2$, which implies $\log_2 k + 1 \geq \bar{d}$. Therefore, we have

$$\operatorname{sat}^*(n, \mathcal{A}_{k+1}) \geq k \left(\left\lceil \frac{n}{\bar{d}} \right\rceil - 1 \right) + 2 \geq k \left\lceil \frac{n}{\lfloor \log_2 k \rfloor + 1} \right\rceil - k + 2 \geq \frac{kn}{\log_2 k + 1} - k + 2.$$

which completes the proof of part (a).

For part (b)**:** To establish the lower bound, we recolor the chains sequentially such that for all $i < j$, chain C_i is colored before chain C_j. Let $d^*(j)$ be the function such that it is the largest d for which $\binom{d}{\lfloor d/2 \rfloor} \leq j - 1$, where d represents the size of a gap in a chain. Every element in the largest antichain must have a different color, so the size of the largest antichain in a gap has to be less than or equal to the number of colors in the gap. By Proposition 10 in [28], $\binom{d}{\lfloor d/2 \rfloor}$ is asymptotically $\log_2(j) + \frac{1}{2} \log_2 \log_2(j)$. Suppose that each chain does not contain $\emptyset$ or $[n]$. Then,

$$\text{sat}^*(n, \mathcal{A}_{k+1}) \geq \sum_{j=1}^{k} \left(\left\lceil \frac{n}{d^*(j)} \right\rceil - 1 \right) + 2 = 2n + \sum_{j=3}^{k} \left\lceil \frac{n}{d^*(j)} \right\rceil - k + 2$$

because $d^*(1) = d^*(2) = 1$ and we obtain our desired result. □

3.2 Comparing induced-$\mathcal{A}_{k+1}$-saturated results

Proposition 3.2.1 shows that for small values of k, Theorem 3.1.2(b) is a better bound than part (a). However, for k sufficiently large, Theorem 3.1.2(a) is the better of the two bounds.

Proposition 3.2.1. *For $k \leq 243$ and an infinite sequence of n values*

$$\text{sat}^*(n, \mathcal{A}_{k+1}) \geq 2n + \sum_{j=3}^{k} \left\lceil \frac{n}{d^*(j)} \right\rceil - k + 2 > k \left\lceil \frac{n}{\lfloor \log_2 k \rfloor + 1} \right\rceil - k + 2.$$

For k sufficiently large and $n \gg k$,

$$\text{sat}^*(n, \mathcal{A}_{k+1}) \geq k \left\lceil \frac{n}{\lfloor \log_2 k \rfloor + 1} \right\rceil - k + 2 > 2n + \sum_{j=3}^{k} \left\lceil \frac{n}{d^*(j)} \right\rceil - k + 2.$$

For $k \leq 243$, we compare the two bounds in Appendix **??**.

Proof of Proposition 3.2.1.

For k sufficiently large and $n \gg k$, we prove Proposition 3.2.1 using Claim 3.2.2 and Lemma 3.2.3.

Claim 3.2.2. *For all $j \geq 5$, we obtain $d^*(j) \geq \log_2 j + \frac{1}{2} \log_2 \log_2 j$ for all $j \geq 5$.*

Claim 3.2.2 gives a simple bound, but a more precise bound can be found in reference [28].

Proof of Claim 3.2.2. If d is even, then $\binom{d}{\lfloor d/2 \rfloor} \leq \sqrt{\frac{2}{\pi d}} \cdot 2^d$, see [44]. If d is odd, then

$$\begin{aligned}\binom{d}{\lfloor d/2 \rfloor} = \frac{1}{2}\binom{d+1}{\frac{d+1}{2}} &\leq \frac{1}{2}\sqrt{\frac{2}{\pi}}\frac{2^{d+1}}{\sqrt{d+1}} \\ &= \sqrt{\frac{2}{\pi(d+1)}} \cdot 2^d \\ &\leq \sqrt{\frac{2}{\pi d}} \cdot 2^d.\end{aligned}$$

Define $f(d) := \sqrt{\frac{2}{\pi}}\frac{2^d}{\sqrt{d}}$. Then,

$$\begin{aligned}f'(d) &= \sqrt{\frac{2}{\pi}} \cdot \frac{\sqrt{d} \cdot 2^d \ln 2 - \frac{1}{2} \cdot d^{-1/2} \cdot 2^d}{d} \\ &= \sqrt{\frac{2}{\pi}} \cdot 2^d \left(\frac{\ln 2}{\sqrt{d}} - \frac{1}{2 \cdot d^{3/2}}\right) \\ &= \sqrt{\frac{2}{\pi}} \cdot \frac{2^d}{d^{3/2}} \left(d \cdot \ln 2 - \frac{1}{2}\right) > 0,\end{aligned}$$

for all $d \geq 1$. Thus, $f(d)$ is increasing over $(1, \infty)$. Let us suppose that $j \geq 5$ and $d \leq \log_2 j + \frac{1}{2}\log_2 \log_2 j$. Then since $f(d)$ is increasing,

$$\begin{aligned}\binom{d}{\lfloor d/2 \rfloor} \leq f(d) &\leq f(\log_2 j + \frac{1}{2}\log_2 \log_2 j) \\ &\leq \sqrt{\frac{2}{\pi(\log_2 j + \frac{1}{2}\log_2 \log_2 j)}} \cdot 2^{\log_2 j + \frac{1}{2}\log_2 \log_2 j} \\ &= \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{\log_2 j + \frac{1}{2}\log_2 \log_2 j}} \cdot j\sqrt{\log_2 j} \\ &= j \cdot \sqrt{\frac{2}{\pi}} \cdot \sqrt{\frac{\log_2 j}{\log_2 j + \frac{1}{2}\log_2 \log_2}} \qquad (3.1) \\ &\leq j \cdot \sqrt{\frac{2}{\pi}} \ \text{ if } j \geq 2 \\ &\leq j - 1 \ \text{ if } j \geq 5 \text{ as desired.}\end{aligned}$$

□

Direct computation shows that the result of Claim 3.2.2 holds for $j = 2, 3, 4$. In fact, (3.1) holds for $j = 4$.

Lemma 3.2.3 proves that for k sufficiently large, Theorem 3.1.2(a) gives a better bound than Theorem 3.1.2(b).

Lemma 3.2.3. *If $k \geq k_0 \geq 2^{64}$, then*

$$\frac{k}{\log_2 k + 1} - 2 - \sum_{j=3}^{k} \frac{1}{d^*(j)} \geq 1.$$

If $n \geq k - 2$, then

$$k \left\lceil \frac{n}{\lfloor \log_2 k \rfloor + 1} \right\rceil - k + 2 \geq 2n + \sum_{j=3}^{k} \left\lceil \frac{n}{d^*(j)} \right\rceil - k + 2.$$

Proof. Let $k_0 = 2^{64}$ and $k \geq k_0$. Consider the integral

$$\int_{k_0}^{k} \left(\frac{1}{\log_2 x + 1} - \frac{1}{\log_2 x + \frac{1}{2}\log_2 \log_2 x} - \frac{1}{\ln 2(\log_2 x + 1)^2} \right) dx. \tag{3.2}$$

Note that

$$\frac{1}{\log_2 x + 1} - \frac{1}{\log_2 x + \frac{1}{2}\log_2 \log_2 x} - \frac{1}{\ln 2(\log_2 x + 1)^2}$$
$$= \frac{1}{\log_2 x + 1} \left(1 - \frac{1}{1 + \frac{\frac{1}{2}\log_2 \log_2 x - 1}{\log_2 x + 1}} - \frac{1}{\ln 2(\log_2 x + 1)} \right).$$

Let $\alpha = \frac{\frac{1}{2}\log_2 \log_2 x - 1}{\log_2 x + 1}$. Using the fact $\frac{65}{67} \cdot \alpha \leq 1 - \frac{1}{1+\alpha}$, we have

$$\frac{1}{\log_2 x + 1} \left(1 - \frac{1}{1 + \frac{\frac{1}{2}\log_2 \log_2 x - 1}{\log_2 x + 1}} - \frac{1}{\ln 2(\log_2 x + 1)} \right)$$
$$\geq \frac{1}{\log_2 x + 1} \left(\frac{65}{67} \cdot \frac{\frac{1}{2}\log_2 \log_2 x - 1}{\log_2 x + 1} - \frac{1}{(\ln 2)(\log_2 x + 1)} \right)$$
$$\geq \frac{1}{\log_2 x + 1} \left(\frac{1}{x \ln 2} \right)$$

since $\left(\frac{65}{67}\cdot\frac{\frac{1}{2}\log_2\log_2 x-1}{\log_2 x+1}-\frac{1}{(\ln 2)(\log_2 x+1)}\right)\geq\frac{1}{x\ln 2}$ holds for all $x\geq k_0=2^{64}$. Hence,

$$\int_{k_0}^{k}\left(\frac{1}{\log_2 x+1}-\frac{1}{\log_2 x+\frac{1}{2}\log_2\log_2 x}-\frac{1}{\ln 2(\log_2 x+1)^2}\right)dx\geq\int_{k_0}^{k}\frac{dx}{x\ln 2(\log_2 x+1)}$$
$$=[\ln(\log_2 x+1)]_{k_0}^{k}$$

Thus, with k_0 fixed and k sufficiently large,

$$\ln(\log_2 k+1)-\ln(\log_2 k_0+1)\geq 3+\sum_{j=3}^{k_0}\frac{1}{d^*(j)}-\frac{k_0}{\log_2 k_0+1}$$

Therefore, we obtain

$$\int_{k_0}^{k}\left(\frac{1}{\log_2 x+1}-\frac{1}{\ln 2(\log_2 x+1)^2}-\frac{1}{\log_2 x+\frac{1}{2}\log_2\log_2 x}\right)dx\geq 3+\sum_{j=3}^{k_0}\frac{1}{d^*(j)}-\frac{k_0}{\log_2 k_0+1}$$

Rearranging terms, we have

$$\int_{k_0}^{k}\left(\frac{1}{\log_2 x+1}-\frac{1}{\ln 2(\log_2 x+1)^2}\right)dx+\frac{k_0}{\log_2 k_0+1}-2-\int_{k_0}^{k}\frac{dx}{\log_2 x+\frac{1}{2}\log_2\log_2 x}\geq 1+\sum_{j=3}^{k_0}\frac{1}{d^*(j)}$$
$$\frac{k}{\log_2 k+1}-2-\int_{k_0}^{k}\frac{dx}{\log_2 x+\frac{1}{2}\log_2\log_2 x}\geq 1+\sum_{j=3}^{k}\frac{1}{d^*(j)}$$
$$\frac{k}{\log_2 k+1}-2-\sum_{j=k_0+1}^{k}\frac{1}{d^*(j)}\geq 1+\sum_{j=3}^{k_0}\frac{1}{d^*(j)}.$$

If $n\geq k-2$, then

$$\frac{k}{\log_2 k+1}-2-\sum_{j=3}^{k}\frac{1}{d^*(j)}\geq 1\geq\frac{k-2}{n}.$$

Direct calculation gives,

$$\frac{kn}{\log_2 k+1}-k+2\geq 2n+\sum_{j=3}^{k}\frac{1}{d^*(j)}$$
$$k\left\lceil\frac{n}{\lfloor\log_2 k\rfloor+1}\right\rceil-k+2\geq 2n+\sum_{j=3}^{k}\left\lceil\frac{n}{d^*(j)}\right\rceil-k+2$$

completing the proof. □

Therefore, by combining Claim 3.2.2 and Lemma 3.2.3, we have

$$k\left\lceil \frac{n}{\lfloor \log_2 k \rfloor + 1} \right\rceil \geq 2n + \sum_{j=3}^{k} \left\lceil \frac{n}{d^*(j)} \right\rceil$$

as desired. □

CHAPTER 4. VERTEX-IDENTIFYING CODES

We begin this chapter by introducing the entropy function and using entropy to prove Theorem 1.4.3 in Section 4.1. Recall that for a finite graph G, an r-vertex-identifying code (r-VI code) in G is a subset $C \subset V(G)$, with the property that $B_r(u) \cap C \neq B_r(v) \cap C$, for all distinct $u, v \in V(G)$ and $B_r(v) \cap C \neq \emptyset$, for all $v \in V(G)$, where $B_r(v)$ is the ball of radius r around vertex v.

In Section 4.2, Theorem 4.2.1 gives a lower bound for r-VI codes regarding graphs with large symmetric difference. In Corollary 4.2.3, we apply Theorem 4.2.1 to bound the size of a 1-VI code (VI code) in strongly-regular graphs. We conclude this chapter by defining (p, β)-jumbled graphs and providing a lower bound for 1-VI codes in (p, β)-jumbled graphs in Theorem 4.3.6, Section 4.3.

4.1 Entropy background

We introduce entropy and some basic facts using the notes of Galvin [20] and a survey of Radhakrishnan [42].

Definition 4.1.1. *Let X be a discrete random variable. The (binary) entropy $H_2(X)$ of X is given by*

$$H_2(X) = \sum_x -\mathbb{P}(\{X = x\}) \log_2 \mathbb{P}(\{X = x\}),$$

where x varies over the range of X.

In binary entropy the random variable is a Bernoulli random variable with probability p. Other entropy functions can be defined by changing the base of the logarithm. One property of the binary entropy function is that subadditivity holds. Note that a vector $(X_1, X_2, \ldots, X_n)$ of random variables is a random variable itself and so we have the following property.

Property 4.1.2. *(Subadditivity) For a random vector $(X_1, X_2, \ldots, X_n)$,*

$$H_2(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} H_2(X_i).$$

The following theorem of Galvin is a straightforward application of adding up binomial coefficients and we use it to prove Lemma 4.3.7 in Section 4.3.2.

Theorem 4.1.3 (Galvin [20])**.** *Fix* $\alpha \le 1/2$*. For all* n*,*

$$\sum_{i \le \alpha n} \binom{n}{i} \le 2^{H_2(\alpha) n}.$$

For completeness, we restate Theorem 1.4.3 and provide a proof.

Theorem 1.4.3. *Let* G *be a graph on* n *vertices such that* $|B_r(v)| \le \beta \le n/2$ *for all* $v \in V(G)$ *and let* C *be an* r*-VI code in* G*. Then,*

$$|C| \ge \frac{\log_2 n}{H_2(\beta/n)} \tag{4.1}$$

$$\ge \frac{n}{\beta} \cdot \frac{\log_2 n}{\log_2 (en/\beta)}. \tag{4.2}$$

Proof. Let $C = \{c_1, \ldots, c_k\}$ be an r-VI code in G of cardinality k. Choose a vertex $v \in V(G)$ uniformly at random. For each $c \in C$, let $X_c(v)$ be the Boolean random variable that is 1 if $c \in B_r(v)$ and 0 otherwise. Consider the joint distribution $(X_{c_1}(v), \ldots, X_{c_k}(v))$. It is a k-length string of zeroes and ones that achieves exactly n values. Each value it achieves is equally likely.

Thus, the binary entropy of the joint distribution is $H_2(X_{c_1}(v), \ldots, X_{c_k}(v)) = \log_2 n$. Hence,

$$\begin{aligned} \log_2 n &= H_2(X_{c_1}(v), \ldots, X_{c_k}(v)) \\ &\le \sum_{i=1}^{k} H_2(X_{c_i}(v)) \ \text{(by subadditivity)} \\ &= \sum_{i=1}^{k} H_2(|B_r(c)|/n) \\ &\le k H_2(\beta/n), \end{aligned} \tag{4.3}$$

because $H_2(x)$ is increasing for $x \in (0, 1/2)$ and decreasing for $x \in (1/2, 1)$. The result $k \ge \frac{\log_2 n}{H_2(\beta/n)}$ in (4.1) follows.

As for (4.2), we simply need to produce bounds for $H_2(x)$. For $x \in (0,1)$, we have

$$H_2(x) = -x \log_2 x - (1-x)\log_2(1-x) \le -x \log_2 x + \frac{x}{\ln 2}.$$

Consequently,

$$\begin{aligned} k &\geq \frac{\log_2 n}{H_2(\beta/n)} \\ &\geq \frac{\log_2 n}{-\frac{\beta}{n}\log_2\left(\frac{\beta}{n}\right) + \frac{\beta/n}{\ln 2}} \\ &= \frac{\ln n}{-\frac{\beta}{n}\ln\left(\frac{\beta}{n}\right) + \frac{\beta}{n}} \\ &= \frac{n}{\beta} \cdot \frac{\log_2 n}{\log_2 n\left(\frac{en}{\beta}\right)}. \end{aligned}$$

□

4.2 Strongly-regular graphs results

In this section, we find upper bounds for r-VI codes in strongly-regular graphs. To do so, we show that if a graph G has the property that the symmetric differences $B_r(u) \triangle B_r(v)$ for distinct $u, v \in V(G)$ are all large, then G is guaranteed to have a small r-VI code.

Theorem 4.2.1. *Let G be a graph on $n \geq 4$ vertices such that $|B_r(u) \triangle B_r(v)| \geq \sigma > 2\ln n$ for all distinct $u, v \in V(G)$. Then, G has an r-VI code of size at most*

$$\left(1 + \sqrt{\frac{1}{\ln n}}\right) \frac{2n \ln n}{\sigma}.$$

It should be noted that the probabilistic method used to prove Theorem 4.2.1 appears in Kim et al [34].

Proof. Choose each vertex to be in C with probability p. The pair of distinct vertices $\{u, v\}$ is a *bad pair* if no vertex in $B_r(u) \triangle B_r(v)$ was chosen. By a simple union bound,

$$\Pr(\exists \text{ a bad pair}) \leq \sum_{u,v} (1-p)^{|B_r(u) \triangle B_r(v)|} \leq \binom{n}{2} (1-p)^{\sigma} < \frac{1}{2} \exp\{2\ln n - p\sigma\}.$$

Set $p := 2\ln n/\sigma$. Thus, with this value of p, the probability is greater than $1/2$ that there exists no bad pair.

Let $X = |C|$ and observe that X is a binomial random variable with mean pn and variance $np(1-p)$. A basic Chernoff bound (see Appendix A of Alon and Spencer [1]) gives that

$$\Pr(X - pn > \alpha) < \exp\left\{-\frac{\alpha^2}{4np(1-p)}\right\}.$$

By setting $\alpha = \sqrt{3np(1-p)}$, we see that $|C| \leq pn + \alpha = pn + \sqrt{3np(1-p)}$, with probability at least $1 - \exp\left\{-\frac{3}{4}\right\} > \frac{1}{2}$.

So, with strictly positive probability,

$$\begin{aligned} |C| &\leq pn + \sqrt{3np(1-p)} \\ &\leq \frac{2n\ln n}{\sigma} + \sqrt{\frac{6n\ln n}{\sigma}} \\ &= \frac{2n\ln n}{\sigma}\left(1 + \sqrt{\frac{3\sigma}{2n\ln n}}\right). \end{aligned} \tag{4.4}$$

What remains is to control the error term of $\sqrt{\frac{3\sigma}{2n\ln n}}$, to do so, we must bound σ. The average value of $|B_r(u)|$ over all vertices $u \in V(G)$ is determined by the matrix $\mathbf{R}$, where the (u,v) entry is 1 if the distance between vertices u and v is at most r and is zero otherwise. Hence, with $\mathbf{J}$ being the $n \times n$ matrix whose entries are all one and $\mathbf{1}$ the $1 \times n$ row vector with all ones,

$$\begin{aligned} \sigma &\leq \binom{n}{2}^{-1} \mathbf{1}^T(\mathbf{J} - \mathbf{R})\mathbf{R}\mathbf{1} \\ &= \binom{n}{2}^{-1} \left(n\mathbf{1}^T\mathbf{R}\mathbf{1} - \mathbf{1}^T\mathbf{R}^2\mathbf{1}\right) \\ &= \binom{n}{2}^{-1} \left(n\sum_{u\in V(G)} |B_r(u)| - \sum_{u\in V(G)} |B_r(u)|^2\right) \\ &\leq \binom{n}{2}^{-1} \left(n\sum_{u\in V(G)} |B_r(u)| - n\left(\frac{1}{n}\sum_{u\in V(G)} |B_r(u)|\right)^2\right), \end{aligned} \tag{4.5}$$

where the last inequality of (4.5) holds by Jensen's inequality. The maximum of (4.5) occurs for $\sum_u |B_r(u)| = n^2/2$. So,

$$\sigma \leq \binom{n}{2}^{-1} \left(n \cdot \frac{n^2}{2} - n\left(\frac{1}{n} \cdot \frac{n^2}{2}\right)^2\right) = \frac{n^2}{2(n-1)}.$$

We return to (4.4) and, for $n \geq 4$,

$$
\begin{aligned}
|C| &\leq \frac{2n\ln n}{\sigma}\left(1+\sqrt{\frac{3\sigma}{2n\ln n}}\right) \\
&\leq \frac{2n\ln n}{\sigma}\left(1+\sqrt{\frac{3}{2n\ln n}\cdot\frac{n^2}{2(n-1)}}\right) \\
&\leq \frac{2n\ln n}{\sigma}\left(1+\frac{1}{\sqrt{\ln n}}\sqrt{\frac{3n}{4(n-1)}}\right) \\
&\leq \frac{2n\ln n}{\sigma}\left(1+\frac{1}{\sqrt{\ln n}}\right).
\end{aligned}
$$

This concludes the proof of Theorem 4.2.1. □

Random graphs were first introduced by Erdős and Rényi [16] in 1960. A *random graph* $G(n,p)$ is a graph on n vertices such that each edge is chosen independently with probability $0 \leq p \leq 1$. Random graphs are best known for providing extremal graphs for several extremal problems and they offer examples of graphs with certain properties. For any fixed constant p, the random graph $G(n,p)$ has large symmetric differences with high probability. An explicitly-defined class of graphs which have large symmetric differences $B_r(u) \triangle B_r(v)$ are strongly-regular graphs. Strongly regular graphs were introduced by Bose [4] in 1963.

Definition 4.2.2. *Let $N(v)$ denote the neighborhood of a vertex v and let n be a positive integer and d,a,c be nonnegative integers. An (n,d,a,c)*-strongly-regular graph *G is a graph on n vertices that is d-regular such that, for all distinct u and v, $|N(u) \cap N(v)| = a$ if $u \sim v$ and c otherwise.*

So we may now apply Theorem 4.2.1 to strongly-regular graphs to bound the size of a 1-VI code in a strongly-regular graph:

Corollary 4.2.3. *If G is a (n,d,a,c)-strongly-regular graph with $n \geq 4$, $d \leq n-2$ and $c \geq 1$, then G has a VI code of size at most*

$$\left(1+\frac{1}{\sqrt{\ln n}}\right)\frac{n\ln n}{\min\{d-a,d-c+1\}}.$$

Moreover, if $n \geq 4$, $d \leq n-2$, then a $(n, d, a, a+1)$-strongly-regular graph has a VI code of size at most

$$\left(1+\frac{1}{\sqrt{\ln n}}\right)\frac{n(n-1)\ln n}{(d+1)(n-1)-d^2}.$$

Proof. Every pair of distinct vertices $u, v \in V(G)$ has the property that $|B_1(u) \triangle B_1(v)|$ is either $2d-2a$ or $2d-2c+2$, depending on whether or not u and v are adjacent. So $\sigma = 2\min\{d-a, d-c+1\}$ and the bound $\left(1+\frac{1}{\sqrt{\ln n}}\right)\frac{n\ln n}{\min\{d-a,d-c+1\}}$ follows for general strongly-regular graphs.

It is well known that the parameters of an (n, d, a, c)-strongly-regular graph obey the equation $(n-d-1)c = d(d-a-1)$. So, for $c = a+1$, we have $(a+1)(n-1) = d^2$. Thus, $\min\{d-a, d-c+1\} = d+1-\frac{d^2}{n-1}$ and the bound $\left(1+\frac{1}{\sqrt{\ln n}}\right)\frac{n(n-1)\ln n}{(d+1)(n-1)-d^2}$ follows for strongly-regular graphs with $c = a+1$. □

Remark 4.2.4. *A Paley graph is a strongly-regular graph with $c = a+1$ and $d = (n-1)/2$ and so it has a VI code of size at most $\left(1+\frac{1}{\sqrt{\ln n}}\right)\frac{4n\ln n}{n+3}$, which follows from Corollary 4.2.3.*

4.3 Pseudo-random graphs result

4.3.1 Pseudo-random graphs background

In 1987, Thomason [47] provided a class of graphs called (p, β)-jumbled graphs which behaved in many ways like random graphs with edge probability p. A (p, β)-jumbled graph is defined as follows:

Definition 4.3.1. *A graph G on a vertex set V is (p, β)-jumbled if, for all vertex subsets $X, Y \subseteq V(G)$,*

$$|e(X,Y) - p|X||Y|| \leq \beta\sqrt{|X||Y|}.$$

This definition for a (p, β)-jumbled graph suggests that (p, β)-jumbled graphs are "pseudo-random" because of their behavior in which edges are chosen with probability p. There are many important preliminary facts to note about (p, β)-jumbled graphs. If a graph G is (p, β)-jumbled, then every induced subgraph is (p, β)-jumbled and the complement of G is $(1-p, \beta)$-jumbled. It is also natural that every graph on n vertices is $(p, n/2)$-jumbled, so β small is of more interest.

We provide a few examples of (p, β)-jumbled graphs for β small to show how (p, β)-jumbled graphs behave like random graphs. These examples and others can be found in [47].

Example 4.3.2. *Let G be a graph whose edges are chosen at random with probability p. Then G is almost surely $(p, 2(pn)^{1/2})$-jumbled under the provision that $pn \to \infty$ and $(1-p)n \to \infty$.*

Example 4.3.3. *Let G be a graph in $G(n, p)$ and let X be a subset of vertices of G with $|X| = \lfloor (pn)^{1/2} \rfloor$. Join each pair of vertices in X. Then G is almost surely $(p, (pn)^{1/2})$-jumbled.*

The next example gives a strongly-regular graph that is (p, β)-jumbled.

Example 4.3.4. *A* Paley graph *on n vertices is an $\left(n, \frac{n-1}{2}, \frac{n-5}{4}, \frac{n-1}{4}\right)$-strongly-regular graph and is $(\frac{1}{2}, n^{1/2})$-jumbled.*

The following theorem provides a local approach for testing the jumbledness of a graph. It uses a degree condition to test the jumbledness of a graph.

Theorem 4.3.5 (Thomason [47])**.** *Let n be an integer, and let $0 < p < 1$ and $\mu \geq 0$ be real numbers. Let G be a graph of order n with minimum degree pn in which no two vertices have more than $p^2 n + \mu$ common neighbors. Then G is $(p, ((pn + (n-1)\mu)^{\frac{1}{2}} + p)/2)$-jumbled.*

For more information about testing the jumbledness of a graph or properties of jumbled graphs see [47].

4.3.2 (p, β)-jumbled graph results

Theorem 4.3.6 below gives a lower bound for (p, β)-jumbled graphs for p fixed and β small.

Theorem 4.3.6. *Let n be an integer, $0 < p < 1$ where p is fixed, and let $\beta = o(\sqrt{n \log_2 n})$. There exists an $\varepsilon = o(1)$ such that if G is a (p, β)-jumbled simple graph on n vertices, then every VI code in G has cardinality at least $\frac{(1-\varepsilon)\log_2 n}{H_2(p)}$.*

To prove Theorem 4.3.6, we first provide the following two necessary and supporting lemmas.

Lemma 4.3.7. *Let G be a (p,β)-jumbled graph. Let C be a VI code of size k. Let $\delta = \delta(n) > 0$, $p = p(n) > 0$, and $(p+\delta) \leq \frac{1}{2}$. Then,*

$$|\{v \in V : |N[v] \cap C| \leq (p+\delta)k\}| \leq 2^{H_2(p)k + \delta k \log_2\left(\frac{1-p}{p}\right)}.$$

Lemma 4.3.7 says that if p is small, then the number of vertices having small intersection with code C is also small because of Theorem 4.1.3.

Lemma 4.3.8. *Let G be a (p,β)-jumbled simple graph. Let C be a VI code of size k. Let $\delta = \delta(n) > 0$ and suppose $(p+\delta) \leq \frac{1}{2}$. Then,*

$$|\{v \in V : |N[v] \cap C| > (p+\delta)k\}| \leq \frac{\beta^2 k}{(\delta k - 1)^2} \leq \frac{2\beta^2}{\delta^2 k}$$

as long as $\delta k \geq 4$.

Lemma 4.3.8 says that if p is small, then the number of vertices having large intersection with code C is small because of jumbledness.

Now that we have established the above lemmas, we will prove Theorem 4.3.6. For completeness, we prove Lemmas 4.3.7 and 4.3.8 in Section 4.3.3.

Proof of Theorem 4.3.6. We prove this theorem by contradiction. We first suppose that we have a code of a particular size and then we show that the size of the code must be bigger.

Let G be a (p,β)-jumbled graph on n vertices where $\beta \geq \sqrt{\frac{8nH_2(p)}{\log_2 n}}$ and $\delta k \geq 4$. Suppose that G has a VI code C of size k, where $k = \max\left\{\frac{(1-\varepsilon)\log_2 n}{H_2(p)}, \log_2 n\right\}$. We may assume that $\frac{1-\varepsilon}{H_2(p)} > 1$ or else the result is trivial by Karpovsky et al. [31]. Let $\delta = \sqrt{\frac{2\beta^2}{(1-\varepsilon)kn}}$ and $\varepsilon = \frac{2\delta k \log_2\left(\frac{1-p}{p}\right)}{\log_2 n}$. Consider the set $S_- := \{v \in V : |N[v] \cap C| \leq (p+\delta)k\}$. By Lemma 4.3.7, $|S_-| \leq 2^{H_2(p)k + \delta k \log_2\left(\frac{1-p}{p}\right)}$. Now consider the set $S_+ := \{v \in V : |N[v] \cap C| \geq (p+\delta)k + 1\}$. By Lemma 4.3.8, $|S_+| \leq \frac{\beta^2 k}{(\delta k - 1)^2} \leq \frac{2\beta^2}{\delta^2 k}$. Note that it is enough to show that $|S_-| + |S_+| < n$ because there needs to be at least n sets $N[v] \cap C$ for C to be a code since G is a graph on n vertices.

Note that

$$\frac{2\beta^2}{\delta^2 k} = \frac{2\beta^2}{\frac{2\beta^2}{(1-\varepsilon)kn} \cdot k} = (1-\varepsilon)n.$$

On the other hand,

$$H_2(p)k + \delta k \log_2\left(\frac{1-p}{p}\right) = H_2(p)\frac{(1-\varepsilon)\log_2 n}{H_2(p)} + \delta k \log_2\left(\frac{1-p}{p}\right)$$
$$= \left(1 - \frac{\varepsilon}{2}\right)\log_2 n$$

since $\varepsilon = \frac{2\delta k \log_2\left(\frac{1-p}{p}\right)}{\log_2 n}$. Furthermore, for n sufficiently large,

$$\left(1 - \frac{\varepsilon}{2}\right)\log_2 n \quad \leq \log_2\left(\frac{\varepsilon}{2}n\right).$$

Therefore,

$$|S_-| + |S_+| \leq 2^{H_2(p)k + \delta k \log_2\left(\frac{1-p}{p}\right)} + \frac{\beta^2 k}{(\delta k - 1)^2}$$
$$\leq \frac{\varepsilon}{2}n + (1-\varepsilon)n$$
$$= \left(1 - \frac{\varepsilon}{2}\right)n$$
$$< n$$

as desired. □

Remark 4.3.9. *For $p = \frac{1}{2}$ and β small, Theorem 4.3.6 says there exists an $\varepsilon = o(1)$ such that if G is $(\frac{1}{2}, \beta)$-jumbled, every VI code of G has cardinality $\frac{(1-\varepsilon)\log_2 n}{-\log_2(1/2)}$. On the other hand, the size of the smallest VI code in a random graph $G(n, \frac{1}{2})$ is asymptotically $\frac{2\log_2 n}{-\log_2(1/2)}$.*

4.3.3 Proof of Lemma 4.3.7 and Lemma 4.3.8

Prior to proving Lemma 4.3.7, we provide a supporting proposition.

Proposition 4.3.10. *Let $0 < p < 1$ and $-p \leq \delta \leq 1 - p$. Then $H_2(p+\delta) \leq H_2(p) + \delta \log_2\left(\frac{1-p}{p}\right)$.*

Proof. We prove the following: Let $0 \leq x \leq 1$ and $p \in (0, 1)$. Then $H_2(x) \leq H_2(p) + (x - p)\log_2\left(\frac{1-p}{p}\right)$. Taking derivatives, we obtain that $H_2'(x) = \log_2\left(\frac{1-x}{x}\right)$ and $H_2''(x) = \frac{-1}{x(1-x)\ln 2} < 0$ for $0 \leq x \leq 1$. Therefore, $H_2(x)$ is concave down and any tangent line of $H_2(x)$ lies above $H_2(x)$.

Note that the tangent line of $H_2(x)$ at $x = p$ is $H_2(p) + (x - p)\log_2\left(\frac{1-p}{p}\right)$. Thus, by replacing x with $p + \delta$, we have $H_2(p + \delta) \leq H_2(p) + \delta \log_2\left(\frac{1-p}{p}\right)$.

□

The importance of Proposition 4.3.10 is to provide an upper bound to Theorem 4.1.3, which will be used in Lemma 4.3.7. The proof of Lemma 4.3.7 follows from Theorem 4.1.3 and Proposition 4.3.10.

Proof of Lemma 4.3.7. Note that the size of $\{v \in V : |N[v] \cap C| \leq (p + \delta)k\}$ is at most $\sum_{i \leq (p+\delta)k} \binom{k}{i}$. Since $(p + \delta) \leq \frac{1}{2}$, then it follows that

$$\begin{aligned} |\{v \in V : |N[v] \cap C| \leq (p + \delta)k\}| &\leq \sum_{i \leq (p+\delta)k} \binom{k}{i} \\ &\leq 2^{H_2(p+\delta)k} && \text{(by Theorem 4.1.3)} \\ &\leq 2^{H_2(p)k + \delta k \log_2\left(\frac{1-p}{p}\right)} && \text{(by Proposition 4.3.10).} \end{aligned}$$

□

Proof of Lemma 4.3.8. Let G be a simple graph, which has no looped edges. Recall that $S_+ := \{v \in V : |N[v] \cap C| \geq (p + \delta)k + 1\}$. Let $e(S_+, C)$ be the number of edges between S_+ and the code C. Note that $e(S_+, C)$ is at least $|S_+|(\lceil(p + \delta)k\rceil + 1) - |S_+ \cap C|$ since each vertex $v \in S_+$ is adjacent to at least $(p + \delta)k + 1$ vertices of C, and since for any vertex $v_c \in S_+$ where $v_c \in C$, an edge from v_c to v_c is counted and G is simple. Since G is (p, β)-jumbled, then by definition $e(S_+, C) \leq p|S_+|k + \beta\sqrt{|S_+|k}$. Thus,

$$|S_+|(\lceil(p + \delta)k\rceil + 1) - |S_+ \cap C| \leq e(S_+, C) \leq p|S_+|k + \beta\sqrt{|S_+|k}.$$

Since $|S_+| - |S_+ \cap C| \geq 0$, we have

$$\begin{aligned}
|S_+|\lceil (p+\delta)k \rceil &\leq p|S_+|k + \beta\sqrt{|S_+|k} && \text{(since } |S_+| - |S_+ \cap C| \geq 0\text{)} \\
\sqrt{|S_+|}(\lceil (p+\delta)k \rceil - pk) &\leq \beta\sqrt{k} \\
\sqrt{|S_+|}(\delta k - 1) &\leq \beta\sqrt{k} \\
|S_+| &\leq \frac{\beta^2 k}{(\delta k - 1)^2} \\
&\leq \frac{2\beta^2}{\delta^2 k}
\end{aligned}$$

because $\left(1 - \frac{1}{\delta k}\right) \geq \frac{1}{2}$. □

CHAPTER 5. SUMMARY AND FUTURE WORKS

5.1 Forbidden subposet problems conclusion

Let $\mathcal{P}$ be a poset. Recall that $\mathrm{La}(n, \mathcal{P})$ denotes the size of a largest $\mathcal{P}$-free family in the n dimensional Boolean lattice, $\mathcal{B}_n$. In Chapter 2, we discussed the $\mathcal{N}$ poset and proved in Theorem 2.2.3 that $\mathrm{La}(n, \mathcal{N}) \geq \binom{n}{k} + A(n, 4, k)$, where $k = \lfloor n/2 \rfloor$ and $A(n, 4, k)$ is the size of a single-error-correcting code with constant weight k. Recall that Theorem 2.2.3 was only a potential improvement of Katona and Tarján's lower bound result given in Theorem 2.2.2 for an infinite family of values of n. In either case, it can be seen by the result of Griggs and Katona, Theorem 2.2.1, that the lower and upper bounds differ by a factor of 2 in their second order term. We believe that the correct second order term should be 2. Our result in Theorem 2.2.3 improves the lower bound on the second order for n even and $k = \lfloor n/2 \rfloor$. In an effort to improve the lower bound for $\mathcal{N}$-free families, one could improve upon the second order term for all values of n.

We also studied $\mathcal{O}_6$-free families in $\mathcal{B}_n$. Our motivation for resolving the size of the largest $\mathcal{O}_6$-free family on two consecutive layers of $\mathcal{B}_n$ led to the Theorem 2.3.1, which provided a lower and upper bound result for *triangle*-free families. We give a universal upper bound for the size of a *triangle*-free family and an existence result for the lower bound of such a family. Other results for $\mathrm{La}(n, \mathcal{O}_6)$ on two consecutive layers of $\mathcal{B}_n$ were proven by Boehnlein [3] and Kramer [35]. Boehnlein proved that $\mathrm{La}(n, \mathcal{O}_6) \leq 1.56\binom{n}{\lfloor n/2 \rfloor}$ using flag algebras and Kramer proved that $\mathrm{La}(n, \mathcal{O}_6) \leq 1.464\binom{n}{\lfloor n/2 \rfloor}$ by using a counting argument and Cauchy-Schwarz inequality. We believe that $\mathrm{La}(n, \mathcal{O}_6)$ on two consecutive layers can be improved further. One could also study $\mathrm{La}(n, \mathcal{O}_{10})$ on two consecutive layers of $\mathcal{B}_n$. In general, the crown problem remains open for the size of the largest $\{6, 10\}$-crown-free families. Griggs and Lu [26] showed that the best known bound is at most $\left(1 + \frac{1}{\sqrt{2}} + o(1)\right)\binom{n}{n/2}$. An improvement would be to show that the size of the $\{6, 10\}$-crown-free families is at most $(1 + o(1))\binom{n}{n/2}$.

5.2 Induced-$\mathcal{P}$-saturation conclusion

In Chapter 3, we studied the size of induced-$\mathcal{A}_{k+1}$-saturated families, where $\mathcal{A}_{k+1}$ is the antichain of size $k+1$. Recall that $\text{sat}^*(n, \mathcal{P})$ is the induced saturation number of a poset $\mathcal{P}$. In Theorem 3.1.2, we proved that for all k, $\text{sat}^*(n, \mathcal{A}_{k+1}) \geq k \left\lceil \frac{n}{\lfloor \log_2 k \rfloor + 1} \right\rceil - k + 2$. We also proved $\text{sat}^*(n, \mathcal{A}_{k+1}) \geq 2n + \sum_{j=3}^{k} \left\lceil \frac{n}{d^*(j)} \right\rceil - k + 2$, where $d^*(j)$ to is the largest d such that $\binom{d}{\lfloor d/2 \rfloor} \leq j - 1$. Later, we compared the two lower bound results for $\text{sat}^*(n, \mathcal{A}_{k+1})$ in Proposition 3.2.1. In continuing to study $\text{sat}^*(n, \mathcal{A}_{k+1})$, improvements can be made on the lower bound.

Recall Theorem 1.3.6, where bounds for $\text{sat}^*(n, \mathcal{P})$ were given for posets $\mathcal{P} = \mathcal{V}_2, \mathcal{D}_2, \mathcal{O}_4$. Ferrara et al. [19] also proved that for the poset $\mathcal{N}$, we have $\lceil \log_2 n \rceil \leq \text{sat}^*(n, \mathcal{N}) \leq 2n$. In their study of induced-$\mathcal{P}$-saturation, the authors conjectured the following:

Conjecture 5.2.1 (Ferrara et al. [19])**.** *For a given poset $\mathcal{P}$, the following are asymptotic values of* $\text{sat}^*(n, \mathcal{P})$.

1. *For $n > k \geq 3$, $\text{sat}^*(n, \mathcal{A}_{k+1}) = kn(1 + o(1))$.*
2. *For $n \geq 2$, $\text{sat}^*(n, \mathcal{D}_2) = \Theta(n)$.*
3. *For $n \geq 3$, $\text{sat}^*(n, \mathcal{O}_4) = \Theta(n^2)$.*

Theorem 3.1.2 is a step in the right direction for proving Conjecture 5.2.1(1), but it still does not prove the conjecture. Other problems of interest to study would be to determine for which posets $\mathcal{P}$ is $\text{sat}^*(n, \mathcal{P})$ unbounded, or to determine for which posets $\mathcal{P}$ does $\lim_{n\to\infty} \frac{\text{sat}^*(n,\mathcal{P})}{n}$ exist. Notice for the last problem, the limit resembles Griggs and Lu's conjecture (Conjecture 1.2.3).

5.3 Vertex-identifying codes in graphs conclusion

In Chapter 4, we proved in Theorem 4.2.1 that for a graph G with $n \geq 4$ vertices such that the symmetric differences between any two distinct vertices is at least $\sigma > 2 \ln n$, then G has an r-VI code of size at most $\left(1 + \sqrt{\frac{1}{\ln n}}\right) \frac{2n \ln n}{\sigma}$. We then used this result to bound the size of a VI code in a strongly-regular graph. VI codes for (p, β)-jumbled graphs were also studied. We showed in

Theorem 4.3.6 that if G is a (p, β)-jumbled graph where p is fixed and β is small, then every VI code in G has cardinality at least $\frac{(1-\varepsilon)\log_2 n}{H_2(p)}$. Future directions for studying VI codes in strongly-regular graphs and (p, β)-jumbled graphs would be to improve on the lower bounds of such graphs. Other interests would be to study the upper bounds for VI codes in graphs. While we only studied two types of graphs and the sizes of r-VI codes in them, there are other graphs in which VI-codes have been studied. One may also wish to study codes more generally, which was defined by Karpovsky et al. [31] as (ℓ, r)-identifying codes.

BIBLIOGRAPHY

[1] N. Alon and J. H. Spencer. *The Probabilistic Method.* John Wiley & Sons, 2004.

[2] M. Axenovich, J. Manske, and R. Martin. Q_2-free families in the boolean lattice. *Order*, 29(1):177–191, 2012.

[3] E. L. Boehnlein. *On Crown-free Set Families, Diffusion State Difference, and Non-uniform Hypergraphs.* PhD thesis, University of South Carolina, 2016.

[4] R. Bose. Strongly regular graphs, partial geometries and partially balanced designs. *Pacific Journal of Mathematics*, 13(2):389–419, 1963.

[5] A. E. Brouwer, J. B. Shearer, N. J. A. Sloane, and W. D. Smith. A New Table of Constant Weight Codes. In *IEEE Trans Inform Theory.* Citeseer, 1990.

[6] B. Bukh. Set families with a forbidden subposet. *The Electronic Journal of Combinatorics*, 16(1):R142, 2009.

[7] I. Charon, O. Hudry, and A. Lobstein. Minimizing the size of an identifying or locating-dominating code in a graph is NP-hard. *Theoretical Computer Science*, 290(3):2109–2120, 2003.

[8] I. Charon, O. Hudry, and A. Lobstein. Extremal cardinalities for identifying and locating-dominating codes in graphs. *Discrete Mathematics*, 307(35):356–366, 2007.

[9] S.-Y. Choi and P. Guan. Minimum critical squarefree subgraph of a hypercube. In *Proceedings of the Thirty-Ninth Southeastern International Conference on Combinatorics, Graph Theory and Computing*, volume 189, pages 57–64, 2008.

[10] D. Conlon. An extremal theorem in the hypercube. *The Electronic Journal of Combinatorics*, 17(1):R111, 2010.

[11] A. De Bonis and G. O. Katona. Largest families without an r-fork. *Order*, 24(3):181–191, 2007.

[12] A. De Bonis, G. O. Katona, and K. J. Swanepoel. Largest family without $A \cup B \subseteq C \cap D$. *Journal of Combinatorial Theory, Series A*, 111(2):331–336, 2005.

[13] R. P. Dilworth. A decomposition theorem for partially ordered sets. *Annals of Mathematics*, 51(1):161–166, 1950.

[14] P. Erdös. On a lemma of Littlewood and Offord. *Bulletin of the American Mathematical Society*, 51(12):898–902, 1945.

[15] P. Erdös, A. Hajnal, and J. W. Moon. A problem in graph theory. *The American Mathematical Monthly*, 71(10):1107–1110, 1964.

[16] P. Erdös and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

[17] J. R. Faudree, R. J. Faudree, and J. R. Schmitt. A survey of minimum saturated graphs. *The Electronic Journal of Combinatorics*, 1000:19–29, 2011.

[18] M. Ferrara, M. S. Jacobson, F. Pfender, and P. S. Wenger. Graph saturation in multipartite graphs. *Journal of Combinatorics*, 7(1):1–19, 2016.

[19] M. Ferrara, B. Kay, L. Kramer, R. R. Martin, B. Reiniger, H. C. Smith, and E. Sullivan. The saturation number of induced subposets of the boolean lattice. *Discrete Mathematics*, 340(10):2479–2487, 2017.

[20] D. Galvin. Three tutorial lectures on entropy and counting. *arXiv preprint arXiv:1406.7872*, 2014.

[21] D. Gerbner, B. Keszegh, N. Lemons, C. Palmer, D. Pálvölgyi, and B. Patkós. Saturating Sperner Families. *Graphs and Combinatorics*, 29(5):1355–1364, 2013.

[22] R. Graham and N. Sloane. Lower Bounds for Constant Weight Codes. *IEEE Transactions on Information Theory*, 26(1):37–43, 1980.

[23] J. R. Griggs and G. O. Katona. No four subsets forming an N. *Journal of Combinatorial Theory, Series A*, 115(4):677–685, 2008.

[24] J. R. Griggs and W.-T. Li. Progress on poset-free families of subsets. In *Recent Trends in Combinatorics*, pages 317–338. Springer, 2016.

[25] J. R. Griggs, W.-T. Li, and L. Lu. Diamond-free families. *Journal of Combinatorial Theory, Series A*, 119(2):310–322, 2012.

[26] J. R. Griggs and L. Lu. On families of subsets with a forbidden subposet. *Combinatorics, Probability and Computing*, 18(5):731–748, 2009.

[27] D. Grósz, A. Methuku, and C. Tompkins. An upper bound on the size of diamond-free families of sets. *Journal of Combinatorial Theory, Series A*, 156:164–194, 2018.

[28] M. Habib, L. Nourine, O. Raynaud, and E. Thierry. Computational aspects of the 2-dimension of partially ordered sets. *Theoretical Computer Science*, 312(2):401–431, 2004.

[29] K. A. Hogenson. *Random and deterministic versions of extremal poset problems.* PhD thesis, Iowa State University, 2016.

[30] S. Johnson. A New Upper Bound for Error-Correcting Codes. *IRE Transactions on Information Theory*, 8(3):203–207, 1962.

[31] M. G. Karpovsky, K. Chakrabarty, and L. B. Levitin. On a New Class of Codes for Identifying Vertices in Graphs. *IEEE Trans. Inf. Theor.*, 44(2):599–611, 1998.

[32] G. Katona. On separating systems of a finite set. *Journal of Combinatorial Theory*, 1(2):174–194, 1966.

[33] G. Katona and T. Tarján. Extremal problems with excluded subgraphs in the n-cube. *Graph Theory*, pages 84–93, 1983.

[34] Y. Kim, M. Kumbhat, Z. L. Nagy, B. Patks, A. Pokrovskiy, and M. Vizer. Identifying codes and searching with balls in graphs. *Discrete Applied Mathematics*, 193:39–47, 2015.

[35] L. Kramer. *On diamond-free subposets of the Boolean lattice: An application of flag algebras.* PhD thesis, Iowa State University, 2014.

[36] L. Kramer, R. R. Martin, and M. Young. On diamond-free subposets of the boolean lattice. *Journal of Combinatorial Theory, Series A*, 120(3):545–560, 2013.

[37] A. Lobstein. Watching systems, identifying, locating-dominating and discriminating codes in graphs, 2018. Available at https://www.lri.fr/~lobstein/debutBIBidetlocdom.pdf.

[38] L. Lu. On crown-free families of subsets. *Journal of Combinatorial Theory, Series A*, 126:216–231, 2014.

[39] R. R. Martin and S. Walker. A note on the size of $\mathcal{N}$-free families. *European Journal of Mathematics*, 3(2):429–432, 2017.

[40] J. Moncel. On graphs on n vertices having an identifying code of cardinality. *Discrete Applied Mathematics*, 154(14):2032–2039, 2006.

[41] N. Morrison, J. A. Noel, and A. Scott. On saturated k-Sperner systems. *The Electronic Journal of Combinatorics*, 21(3):P3–22, 2014.

[42] J. Radhakrishnan. Entropy and counting. *Computational Mathematics, Modelling and Algorithms*, 146, 2003.

[43] K. H. Rosen. *Handbook of Discrete and Combinatorial Mathematics.* CRC press, 1999.

[44] J. Spencer and L. Florescu. Asymptopia, volume 71 of *Student Mathematical Library. American Mathematical Society, Providence, RI*, 2014.

[45] E. Sperner. Ein satz über untermengen einer endlichen menge. *Mathematische Zeitschrift*, 27(1):544–548, 1928.

[46] H. T. Thanh. An extremal problem with excluded subposet in the boolean lattice. *Order*, 15(1):51–57, 1998.

[47] A. Thomason. Pseudo-random graphs. In *Random graphs*, volume 85, pages 307–331. Elsevier, 1987.

[48] A. A. Zykov. On some properties of linear complexes. *Matematicheskii Sbornik*, 66(2):163–188, 1949.

APPENDIX. COMPARING LOWER BOUNDS OF PROPOSITION 3.2.1

Recall that $d^*(k)$ is the largest d for which $\binom{d}{\lfloor d/2 \rfloor} \leq k-1$. Below we calculate the lower bounds of Proposition 3.2.1 for all $k \leq 300$.

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
1	1	1.0	1.0
2	1	1.0	1.0
3	2	1.5	2.5
4	3	**1.3333**	2.83333
5	4	1.66667	3.08333
6	4	2.0	3.3333
7	4	2.33333	3.58333
8	5	**2.0**	3.78333
9	5	2.25	3.98333
10	5	2.5	4.18333
11	5	2.75	4.38333
12	6	3.0	4.55
13	6	3.25	4.71667
14	6	3.5	4.88333
15	6	3.75	5.05
16	6	**3.2**	5.21667
17	6	3.4	5.38333
18	6	3.6	5.55
19	6	3.8	5.71667
20	6	4.0	5.88333

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
21	6	4.2	6.05
22	7	4.4	6.19286
23	7	4.6	6.33571
24	7	4.8	6.47857
25	7	5.	6.62143
26	7	5.2	6.76429
27	7	5.4	6.90714
28	7	5.6	7.05
29	7	5.8	7.19286
30	7	6.0	7.33571
31	7	6.2	7.47857
32	7	**5.33333**	7.62143
33	7	5.5	7.76429
34	7	5.66667	7.90714
35	7	5.83333	8.05
36	7	6.0	8.19286
37	8	6.16667	8.31786
38	8	6.33333	8.44286
39	8	6.5	8.56786
40	8	6.66667	8.69286

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
41	8	6.83333	8.81786
42	8	7.0	8.94286
43	8	7.16667	9.06786
44	8	7.33333	9.19286
45	8	7.5	9.31786
46	8	7.66667	9.44286
47	8	7.83333	9.56786
48	8	8.0	9.69286
49	8	8.16667	9.81786
50	8	8.33333	9.94286
51	8	8.5	10.0679
52	8	8.66667	10.1929
53	8	8.83333	10.3179
54	8	9.0	10.4429
55	8	9.16667	10.5679
56	8	9.33333	10.6929
57	8	9.5	10.8179
58	8	9.66667	10.9429
59	8	9.83333	11.0679
60	8	10.	11.1929
61	8	10.1667	11.3179
62	8	10.3333	11.4429
63	8	10.5	11.5679
64	8	**9.14286**	11.6929
65	8	9.28571	11.8179

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
66	8	9.42857	11.9429
67	8	9.57143	12.0679
68	8	9.71429	12.1929
69	8	9.85714	12.3179
70	8	10.0	12.4429
71	8	10.1429	12.5679
72	9	10.2857	12.679
73	9	10.4286	12.7901
74	9	10.5714	12.9012
75	9	10.7143	13.0123
76	9	10.8571	13.1234
77	9	11.0	13.2345
78	9	11.1429	13.3456
79	9	11.2857	13.4567
80	9	11.4286	13.5679
81	9	11.5714	13.679
82	9	11.7143	13.7901
83	9	11.8571	13.9012
84	9	12.0	14.0123
85	9	12.1429	14.1234
86	9	12.2857	14.2345
87	9	12.4286	14.3456
88	9	12.5714	14.4567
89	9	12.7143	14.5679
90	9	12.8571	14.679

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
91	9	13.0	14.7901
92	9	13.1429	14.9012
93	9	13.2857	15.0123
94	9	13.4286	15.1234
95	9	13.5714	15.2345
96	9	13.7143	15.3456
97	9	13.8571	15.4567
98	9	14.0	15.5679
99	9	14.1429	15.679
100	9	14.2857	15.7901
101	9	14.4286	15.9012
102	9	14.5714	16.0123
103	9	14.7143	16.1234
104	9	14.8571	16.2345
105	9	15.0	16.3456
106	9	15.1429	16.4567
107	9	15.2857	16.5679
108	9	15.4286	16.679
109	9	15.5714	16.7901
110	9	15.7143	16.9012
111	9	15.8571	17.0123
112	9	16.0	17.1234
113	9	16.1429	17.2345
114	9	16.2857	17.3456
115	9	16.4286	17.4567

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
116	9	16.5714	17.5679
117	9	16.7143	17.679
118	9	16.8571	17.7901
119	9	17.0	17.9012
120	9	17.1429	18.0123
121	9	17.2857	18.1234
122	9	17.4286	18.2345
123	9	17.5714	18.3456
124	9	17.7143	18.4567
125	9	17.8571	18.5679
126	9	18.0	18.679
127	9	18.1429	18.7901
128	10	**16.0**	18.8901
129	10	16.125	18.9901
130	10	16.25	19.0901
131	10	16.375	19.1901
132	10	16.5	19.2901
133	10	16.625	19.3901
134	10	16.75	19.4901
135	10	16.875	19.5901
136	10	17.0	19.6901
137	10	17.125	19.7901
138	10	17.25	19.8901
139	10	17.375	19.9901
140	10	17.5	20.0901

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
141	10	17.625	20.1901
142	10	17.75	20.2901
143	10	17.875	20.3901
144	10	18.0	20.4901
145	10	18.125	20.5901
146	10	18.25	20.6901
147	10	18.375	20.7901
148	10	18.5	20.8901
149	10	18.625	20.9901
150	10	18.75	21.0901
151	10	18.875	21.1901
152	10	19.0	21.2901
153	10	19.125	21.3901
154	10	19.25	21.4901
155	10	19.375	21.5901
156	10	19.5	21.6901
157	10	19.625	21.7901
158	10	19.75	21.8901
159	10	19.875	21.9901
160	10	20.0	22.0901
161	10	20.125	22.1901
162	10	20.25	22.2901
163	10	20.375	22.3901
164	10	20.5	22.4901
165	10	20.625	22.5901

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
166	10	20.75	22.6901
167	10	20.875	22.7901
168	10	21.0	22.8901
169	10	21.125	22.9901
170	10	21.25	23.0901
171	10	21.375	23.1901
172	10	21.5	23.2901
173	10	21.625	23.3901
174	10	21.75	23.4901
175	10	21.875	23.5901
176	10	22.0	23.6901
177	10	22.125	23.7901
178	10	22.25	23.8901
179	10	22.375	23.9901
180	10	22.5	24.0901
181	10	22.625	24.1901
182	10	22.75	24.2901
183	10	22.875	24.3901
184	10	23.0	24.4901
185	10	23.125	24.5901
186	10	23.25	24.6901
187	10	23.375	24.7901
188	10	23.5	24.8901
189	10	23.625	24.9901
190	10	23.75	25.0901

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
191	10	23.875	25.1901
192	10	24.0	25.2901
193	10	24.125	25.3901
194	10	24.25	25.4901
195	10	24.375	25.5901
196	10	24.5	25.6901
197	10	24.625	25.7901
198	10	24.75	25.8901
199	10	24.875	25.9901
200	10	25.0	26.0901
201	10	25.125	26.1901
202	10	25.25	26.2901
203	10	25.375	26.3901
204	10	25.5	26.4901
205	10	25.625	26.5901
206	10	25.75	26.6901
207	10	25.875	26.7901
208	10	26.0	26.8901
209	10	26.125	26.9901
210	10	26.25	27.0901
211	10	26.375	27.1901
212	10	26.5	27.2901
213	10	26.625	27.3901
214	10	26.75	27.4901
215	10	26.875	27.5901

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
216	10	27.0	27.6901
217	10	27.125	27.7901
218	10	27.25	27.8901
219	10	27.375	27.9901
220	10	27.5	28.0901
221	10	27.625	28.1901
222	10	27.75	28.2901
223	10	27.875	28.3901
224	10	28.0	28.4901
225	10	28.125	28.5901
226	10	28.25	28.6901
227	10	28.375	28.7901
228	10	28.5	28.8901
229	10	28.625	28.9901
230	10	28.75	29.0901
231	10	28.875	29.1901
232	10	29.0	29.2901
233	10	29.125	29.3901
234	10	29.25	29.4901
235	10	29.375	29.5901
236	10	29.5	29.6901
237	10	29.625	29.7901
238	10	29.75	29.8901
239	10	29.875	29.9901
240	10	30.0	30.0901

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
241	10	30.125	30.1901
242	10	30.25	30.2901
243	10	30.375	30.3901
244	10	30.5	30.4901
245	10	30.625	30.5901
246	10	30.75	30.6901
247	10	30.875	30.7901
248	10	31.0	30.8901
249	10	31.125	30.9901
250	10	31.25	31.0901
251	10	31.375	31.1901
252	10	31.5	31.2901
253	10	31.625	31.3901
254	11	31.75	31.481
255	11	31.875	31.5719
256	11	**28.4444**	31.66281
257	11	28.5556	31.7537
258	11	28.6667	31.8446
259	11	28.7778	31.9355
260	11	28.8889	32.0264
261	11	29.0	32.1174
262	11	29.1111	32.2083
263	11	29.2222	32.2992
264	11	29.3333	32.3901
265	11	29.4444	32.481

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
266	11	29.5556	32.5719
267	11	29.6667	32.6628
268	11	29.7778	32.7537
269	11	29.8889	32.8446
270	11	30.0	32.9355
271	11	30.1111	33.0264
272	11	30.2222	33.1174
273	11	30.3333	33.2083
274	11	30.4444	33.2992
275	11	30.5556	33.3901
276	11	30.6667	33.481
277	11	30.7778	33.5719
278	11	30.8889	33.6628
279	11	31.0	33.7537
280	11	31.1111	33.8446
281	11	31.2222	33.9355
282	11	31.3333	34.0264
283	11	31.4444	34.1174
284	11	31.5556	34.2083
285	11	31.6667	34.2992
286	11	31.7778	34.3901
287	11	31.8889	34.481
288	11	32.0	34.5719
289	11	32.1111	34.6628
290	11	32.2222	34.7537

k	$d^*(k)$	$\frac{k}{\lfloor \log_2 k \rfloor + 1}$	$\sum_{j=1}^{k} \frac{1}{d^*(j)}$
291	11	32.3333	34.8446
292	11	32.4444	34.9355
293	11	32.5556	35.0264
294	11	32.6667	35.1174
295	11	32.7778	35.2083
296	11	32.8889	35.2992
297	11	33.0	35.3901
298	11	33.1111	35.481
299	11	33.2222	35.5719
300	11	33.3333	35.6628

Note: If $k \leq 243$, then $\frac{k}{\lfloor \log_2 k \rfloor + 1} \leq \sum_{j=1}^{k} \frac{1}{d^*(j)}$. However, when $k > 243$ and $d^*(k) < 10$, then $\frac{k}{\lfloor \log_2 k \rfloor + 1} > \sum_{j=1}^{k} \frac{1}{d^*(j)}$. At $k = 256$, $d^*(k) = 11$ and we see that $\frac{k}{\lfloor \log_2 k \rfloor + 1} \leq \sum_{j=1}^{k} \frac{1}{d^*(j)}$. It is worth mentioning the behavior of the bound $\frac{k}{\lfloor \log_2 k \rfloor + 1}$. For $k = 2^m$ for $m \geq 1$ and an integer, the bound $\frac{k}{\lfloor \log_2 k \rfloor + 1}$ drops and then begins to increase as k increases.