EXTREMAL GRAPH THEORY: SUPERSATURATION AND ENUMERATION

BY

HONG LIU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Professor Alexandr V. Kostochka, Chair
Professor József Balogh, Director of Research
Professor Bruce Reznick
J.L. Doob Research Assistant Professor Theo Molla

Abstract

In this thesis, we study supersaturation and enumeration problems in extremal combinatorics. In Chapter 2, with Balogh, we disprove a conjecture of Erdős and Tuza concerning the number of different ways one can create a copy of K_4, a complete graph on 4 vertices, in a K_4-free graph.

In Chapter 3, we extend a classical result of Kolaitis, Prömel and Rothschild on the typical structure of graphs forbidding a clique of fixed order as a subgraph, showing that the order of the forbidden clique can be as large as some polylogarithmic function of the order of the host graph. This is based on joint work with Balogh, Bushaw, Collares Neto, Morris and Sharifzadeh.

In Chapter 4 and Chapter 5, we study the number of maximal sum-free subsets of the set $[n] := \{1, 2, \ldots, n\}$. Together with Balogh, Sharifzadeh and Treglown, we show that, for each $1 \leq i \leq 4$, there are constants C_i such that the number of maximal sum-free subsets in $[n]$ is $(C_i + o(1))2^{n/4}$, where $i \equiv n \bmod 4$. This resolves a conjecture of Cameron and Erdős.

In Chapter 6, with Balogh and Sharifzadeh, we study the number of subsets of $[n]$ which does not contain an arithmetic progression of a fixed length. This addresses another question of Cameron and Erdős and provides an optimal bound for infinitely many n. As corollaries, we improve the known transference results on arithmetic progressions.

To my beloved parents, Yongquan Liu and Xinzhu Liu.

献给我的父亲，刘用铨，和我的母亲，刘爱芳。

Acknowledgments

I am profoundly grateful to my parents for their love and support. I would not have made it this far without their encouragement.

I am deeply indebted to my passionate and enthusiastic advisor, Józsi Balogh. He gives me advices in not just research and professional abilities, but also in life. None of this work would have been done without his guidance. I would like to thank Józsi for his hospitality when I was visiting him in Szeged, where part of the work in this thesis was carried out.

I would also like to thank my thesis committee members, Alexandr V. Kostochka, Bruce Reznick, and Theo Molla for their time and effort. In particular, many thanks to Professor Kostochka for many inspiring discussions.

Special thanks to Bernard Lidický and Maryam Sharifzadeh for their help on practically everything. Last but not the least, I want to thank my friends and colleagues during my study at the University of Illinois: including but not limited to Carlota Bullard, Jane Butterfield, Li-Hsin Chang, Ilkyoo Choi, Michelle Delcourt, Joe Grohens, Sahand Hariri, Eunice Kim, Jaehoon Kim, Saeed Maleki, Cory Palmer, Sarka Petříčková, Serhiy Potishuk, Maryam Sharifzadeh, Hannah Spinoza, Andrew Treglown, Grace Work.

Table of Contents

Chapter 1

Overview

I am interested in extremal combinatorics, one of the most central areas of modern combinatorics. The field has witnessed tremendous development in the past few decades, partially because of its close interplay with other disciplines such as number theory, probability, discrete geometry and theoretical computer science.

A typical problem in extremal combinatorics is to maximize or minimize some parameter f over a collection of discrete structures with certain properties $\mathcal{P}$. Those families satisfying $\mathcal{P}$ that realize the maximum or minimum of f are called *extremal*. I am interested in various problems in extremal combinatorics. In this thesis, we study the following two types of problems: (i) supersaturation problems which study the appearance of forbidden configurations in families that are somewhat larger (or smaller) than the extremal ones, and (ii) enumeration problems which ask for an estimate of the size of a family with certain constraints.

For the rest of this chapter, I will give an overview of the results that I have obtained with my coauthors.

1.1 Supersaturation, a conjecture of Erdős and Tuza

Consider an extremal problem of maximizing the size of a family of combinatorial objects while forbidding some substructure, a natural question to follow is to ask the number of forbidden configurations appearing in a family that has size over the extremal threshold. This type of question is called the *supersaturation* problem, as often going above the threshold forces not just one but many copies of forbidden substructure. For example, the balanced n-vertex complete bipartite graph $K_{\lceil n/2\rceil,\lfloor n/2\rfloor}$ has $\lfloor n^2/4\rfloor$ edges and yet does not contain a triangle. However, Rademacher observed that if an n-vertex graph has $\lfloor n^2/4\rfloor+1$ edges, then it must contain at least $\lfloor n/2\rfloor$ triangles. This was later generalized to larger cliques by Erdős [34].[1] Supersaturation problems have been studied in various other contexts. For example, Kleitman [59] determined the minimum number of 2-chains in a poset whose

[1]We refer readers to Chapter 2 for more references.

size is larger than its largest antichain[2], recently extended to k-chains by Das, Gan and Sudakov [29].

Let G be an n-vertex K_4-free graph, an edge in its complement is a K_4-*saturating* edge if the addition of this edge to G creates a copy of K_4. Many results related to clique-saturating edges have been established, some of which are recently phrased in the language of 'graph bootstrap percolation'. An example of Bollobás [10] shows that an n-vertex graph can have only $2n-3$ edges, and yet all edges in its complement are K_4-saturating edges. On the other extreme, $K_{\lceil n/2\rceil,\lfloor n/2\rfloor}$ shows that a graph could have up to $\lfloor n^2/4\rfloor$ edges with no K_4-saturating edge. Erdős and Tuza [35] conjectured in the 1980s that one additional edge would change the picture dramatically. This conjecture can be considered, as formulated above, a supersaturation problem concerning the number of K_4-saturating edges.

Conjecture (Erdős-Tuza [35])**.** *If an n-vertex K_4-free graph G has $\lfloor n^2/4\rfloor+1$ edges, then there are at least $(1+o(1))n^2/16$ K_4-saturating edges.*

In joint work with Balogh [11], we disprove this conjecture, giving a counterexample with only $\frac{2n^2}{33}$ K_4-saturating edges. Furthermore, we prove that $(1+o(1))\frac{2n^2}{33}$ is best possible:

Theorem (Balogh-Liu [11])**.** *There are at least $(1+o(1))\frac{2n^2}{33}$ K_4-saturating edges in an n-vertex K_4-free graph with $\lfloor n^2/4\rfloor+1$ edges.*

In fact, a stability result can be derived from our proof.

This work will be presented in Chapter 2.

1.2 Enumeration

Enumerating families of discrete objects with given properties and describing the typical structure[3] of these objects are fundamental problems in extremal combinatorics. In the context of graphs, this was initiated in 1976 by Erdős, Kleitman, and Rothschild [38], who studied the family of triangle-free graphs on n vertices. In extremal set theory, a celebrated result of Kleitman [60] in 1969 determined the number of antichains among subsets of an n-element set. These results have since inspired a great deal of research over the years. For example, Alon, Balogh, Keevash and Sudakov [3] studied the number of edge-colorings with a fixed number of colors that forbid monochromatic copy of a complete graph of prescribed size.

[2]A *k-chain* is a set system consisting of k sets $A_1\subseteq A_2\subseteq\ldots\subseteq A_k$. An *antichain* is a set system forbidding any pair A,B such that $A\subseteq B$.

[3]Loosely speaking, it is a structural statement that shows how most of these objects look like.

We will discuss one such problem in extremal graph theory and two conjectures of Cameron and Erdős on enumeration problems in additive number theory.

1.2.1 Graphs without a large clique

The above-mentioned seminal result of Erdős, Kleitman, and Rothschild [38] states that almost all triangle-free graphs on n vertices are bipartite; that is, the proportion of n-vertex triangle-free graphs that are not bipartite goes to zero as $n \to \infty$. Since then, various extensions of this theorem have been established. In particular, Kolaitis, Prömel and Rothschild [63] extended the result of Erdős-Kleitman-Rothschild [38] to larger forbidden cliques, but of fixed size. They showed that almost all K_{r+1}-free graphs are r-partite. In joint work with Balogh, Bushaw, Collares Neto, Morris and Sharifzadeh [16], we extend the result of Kolaitis, Prömel and Rothschild [63]. Our result allows the size of the forbidden clique grows with the host graph:

Theorem (Balogh-Bushaw-Collares Neto-Liu–Morris-Sharifzadeh [16]). *Almost all K_{r+1}-free graphs on n vertices are r-partite,*[4] *for any $r \leq (\log n)^{1/4}$.*

It is worth mentioning that the statement is no longer true if $r \geq (\log n)^{1+o(1)}$, thus our bound on r is not too far from optimal.[5] En route to the above theorem, we obtain a supersaturation result for cliques, which gives a new proof of the Erdős-Simonovits stability theorem [40] for arbitrary graphs.

We will present this work in Chapter 3.

1.2.2 Maximal sum-free sets

A triple x, y, z is a *Schur triple* if $x + y = z$ (note x, y and z may not necessarily be distinct). A set S is *sum-free* if S does not contain a Schur triple. Let $[n] := \{1, \ldots, n\}$. We say that $S \subseteq [n]$ is a *maximal sum-free subset of* $[n]$ if it is sum-free and it is not properly contained in another sum-free subset of $[n]$. Let $f(n)$ denote the number of sum-free subsets of $[n]$ and $f_{\max}(n)$ denote the number of maximal sum-free subsets of $[n]$. The study of sum-free sets of integers has a rich history. Indeed, Schur [83] in 1916 proved that if integers are finitely colored, then there is a monochromatic Schur triple. Clearly, any set of odd integers and any subset of $\{\lfloor n/2 \rfloor + 1, \ldots, n\}$ is a sum-free set, hence $f(n) \geq 2^{n/2}$. Cameron and Erdős [25] conjectured that $f(n) = O(2^{n/2})$. This conjecture was proven independently by Green [49]

[4] That is, the proportion of n-vertex K_{r+1}-free graphs that are not r-partite goes to zero as $n \to \infty$.
[5] See Chapter 3 for more details.

and Sapozhenko [79]. In fact, they showed that there are constants C_1 and C_2 such that $f(n) = (C_i + o(1))2^{n/2}$ for all $n \equiv i \mod 2$.

In a second paper, Cameron and Erdős [26] showed that $f_{\max}(n) \geq 2^{\lfloor n/4 \rfloor}$. Noting that all the sum-free subsets of $[n]$ described above lie in just two maximal sum-free sets, they asked whether there are significantly fewer sum-free sets that are maximal: $f_{\max}(n) = o(f(n))$ or even $f_{\max}(n) \leq f(n)/2^{\varepsilon n}$ for some constant $\varepsilon > 0$. Łuczak and Schoen [66] answered this question in the affirmative, showing that $f_{\max}(n) \leq 2^{n/2-2^{-28}n}$ for sufficiently large n. Later, Wolfovitz [90] proved that $f_{\max}(n) \leq 2^{3n/8+o(n)}$. This leaves a big gap in the exponent between the lower and upper bounds.

In a sequence of two papers [14, 15], we completely settle this question of Cameron and Erdős [26]. We first in [14] showed that the lower bound is essentially tight:

Theorem (Balogh-Liu-Sharifzadeh-Treglown [14])**.** *The number of maximal sum-free subsets in $[n]$ is $2^{n/4+o(n)}$.*

In [15], we give the following exact solution.

Theorem (Balogh-Liu-Sharifzadeh-Treglown [15])**.** *There are constants C_i, $1 \leq i \leq 4$, such that, given any $n \equiv i \mod 4$, $[n]$ contains $(C_i + o(1))2^{n/4}$ maximal sum-free sets.*

Our proof in fact provides the typical structure of a maximal sum-free set.

Sum-free sets in abelian groups have been extensively studied. For example, Green and Ruzsa [51] determined the number of sum-free sets in abelian groups. With the techniques we developed in [14, 15], a possible future research project is to enumerate maximal sum-free sets in abelian groups.

In Chapter 4, we will present the work in [14], and in Chapter 5, we will present the work in [15].

1.2.3 Sets with no arithmetic progression of prescribed length

A subset of $[n] := \{1, 2, \ldots, n\}$ is *k-AP-free* if it does not contain a k-term arithmetic progression. Denote by $r_k(n)$ the maximum size of a k-AP-free subset of $[n]$. Cameron and Erdős [26] asked for an enumeration for subsets of $[n]$ with no k-term arithmetic progression. In particular, they wondered if it is true that the number of k-AP-free subsets of $[n]$ is $2^{(1+o(1))r_k(n)}$.

Little progress had been made on bounding the number of k-AP-free subsets from above in the last 30 years. One of the reasons for this difficulty is our limited understanding of $r_k(n)$. Indeed, despite much effort, the gap between the current known lower and upper bounds on $r_3(n)$ is still rather large.

In a recent work with Balogh and Sharifzadeh [13], we are able to show the following weaker version.

Theorem (Balogh-Liu-Sharifzadeh [13])**.** *The number of k-AP-free subsets of $[n]$ is at most $2^{O(r_k(n))}$ for infinitely many values of n.*

We derive from the proof of the above result that this infinite set has upper density one. It is worth mentioning that two other natural conjectures of Erdős, regarding the number of C_6-free graphs and the number of Sidon sets[6], are false. It is not inconceivable that the answer to the question of Cameron and Erdős [25] is no.

For all values of n, we obtain a weaker estimate, which is nevertheless sufficient to transfer the current best upper bound on $r_k(n)$ to the sparse random setting. One of the novel ingredients in the heart of our proof is a supersaturation result on arithmetic progression, which states that sets of size $\Omega(r_k(n))$ already contain super linearly many k-term arithmetic progressions. As comparison, all previously known supersaturation results, e.g. [89], only provide meaningful bounds for sets of linear size.

All these results in this subsection will be presented in Chapter 6.

1.3 Background

We provide here some basic concepts and notations. For the readers' convenience, uncommon terms and definitions will be introduced later in each corresponding chapter.

For an integer $n \geq 1$, denote by $[n]$ the set $\{1, 2, \dots, n\}$. Given two functions $f(n)$ and $g(n)$, write $f(n) = o(g(n))$ or equivalently $f(n) \ll g(n)$ if $\lim_{n\to\infty} \dfrac{f(n)}{g(n)} = 0$; write $f(n) = \omega(g(n))$ or equivalently $f(n) \gg g(n)$ if $\frac{f(n)}{g(n)} \to \infty$ as $n \to \infty$; write $f(n) = O(g(n))$ if there exists a constant $C > 0$ such that $f(n) \leq C \cdot g(n)$; write $f(n) = \Omega(g(n))$ if there exists a constant $c > 0$ such that $f(n) \geq c \cdot g(n)$; write $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.

A *graph* G consists of a vertex set $V(G)$ and an edge set $E(G)$. The edge set $E(G)$ is a set of unordered pairs of vertices in $V(G)$. Two vertices x and y are *adjacent*, denoted by $x \sim y$, if they form an edge $e \in E(G)$ and x, y are the *endpoints* of the edge e. The *order* of a graph G is the number of vertices in G and the *size* of G is $e(G) := |E(G)|$, the number of edges in G. Given two disjoint vertex subsets A and B of G, denote by $E(A, B)$ the set of all edges in G with one endpoint in A and the other in B and denote by $e(A, B) := |E(A, B)|$.

[6]A set $A \subseteq [n]$ is a *Sidon* set if there do not exist distinct $a, b, c, d \in A$ such that $a + b = c + d$.

Let $G[A]$ denote the subgraph induced on vertex set A, i.e. $E(G[A])$ consists of all edges in $E(G)$ with both endpoints in A.

Here are some useful graphs. A *path* on t vertices, denoted by P_t, is a graph whose vertices can be ordered as $v_1, v_2, \dots, v_t$ such that v_i is adjacent to v_j if and only if $|i-j| = 1$. A *cycle* on t vertices, denoted by C_t, is a graph whose vertices can be ordered as $v_1, v_2, \dots, v_t$ such that v_i is adjacent to v_j if and only if $i - j = 1$ or -1 modular t. A *complete graph* on t vertices, denoted by K_t, is a graph in which every pair of vertices forms an edge. A *complete bipartite graph* on vertex set $X \cup Y$, denoted by $K_{|X|,|Y|}$, is a graph in which two vertices form an edge if and only if one of them is in X and the other one is in Y.

Given a graph G, a set $I \subseteq V(G)$ is an independent set if I induces no edge in G. The *independence number* of G, denoted by $\alpha(G)$, is the size of the largest independent set in G. A graph $G = (V, E)$ is r-partite if the vertex set V can be partitioned into r disjoint sets $V_1, V_2, \dots, V_r$ such that each V_i, $1 \le i \le r$, is an independent set. Given a vertex $x \in G$, denote by $N_G(x)$ the *neighborhood* of x in G, i.e. the set of vertices adjacent to x in G. The *degree* of x, denoted by $d_G(x)$ or $\deg_G(x)$, is the number of edges incident to x, i.e. $|N_G(x)|$ (we drop the subscript G when it is clear from the contents). Denote by $\Delta(G) := \max\limits_{v \in V(G)} \deg(x)$ and $\delta(G) := \min\limits_{v \in V(G)} \deg(x)$ the *maximum degree* and *minimum degree* of G respectively. For any vertex subset $U \subseteq V(G)$, denote $N(U) := \bigcap_{v \in U} N(v)$.

Given an integer k, a *k-coloring* of a graph G is a function $\phi : V(G) \to [k]$. A coloring ϕ is *proper* if no two adjacent vertices receive the same color, i.e. $\phi(x) \neq \phi(y)$ if $x \sim y$. The *chromatic number* $\chi(G)$ of a graph G is the minimum integer k such that there exists a proper k-coloring of G.

Given two graphs H and G, we say that H is a *subgraph* of G, denoted by $H \subseteq G$, if there is an injection $f : V(H) \to V(G)$ that preserves adjacency, i.e. $f(x) \sim_G f(y)$ if $x \sim_H y$. We say that G is H-free if G does not contain H as a subgraph. The *extremal number* of H, denoted by $\mathrm{ex}(n, H)$, is the maximum number of edges an n-vertex H-free graph can have.

A *hypergraph* $\mathcal{H}$ consists of a vertex set $V(\mathcal{H})$ and an edge set $E(\mathcal{H})$. The edge set is a collection of subsets of the vertex set $V(\mathcal{H})$. A hypergraph $\mathcal{H}$ is *r-uniform* if every edge in $E(\mathcal{H})$ contains r vertices.

Denote by $G(n,p)$ the probability space, introduced by Erdős and Rényi [39], over the set of all graphs on vertex set $[n]$, in which each edge is presented with probability p independently of all others.

We will often use the following Chernoff bound (see e.g. [5]).

Lemma 1.3.1. *Let $X = \sum_i X_i$, where $X_1, \dots, X_n$ are independent 0-1 random variables* *with* $\mathbb{P}[X_i = 1] = p$*. Writing* $\mathbb{E}X = \mu$*, then for any* $0 < \delta < 1$*,*

$$\mathbb{P}[X > (1+\delta)\mu] \leq e^{-\frac{\delta^2\mu}{3}},$$

and

$$\mathbb{P}[X < (1-\delta)\mu] \leq e^{-\frac{\delta^2\mu}{2}}.$$

Chapter 2

Supersaturation

Let G be a K_4-free graph. An edge in its complement is a K_4-*saturating* edge if the addition of this edge to G creates a copy of K_4. Erdős and Tuza [35] conjectured that for any n-vertex K_4-free graph G with $\lfloor \frac{n^2}{4} \rfloor + 1$ edges, one can find at least $(1 + o(1))\frac{n^2}{16}$ K_4-saturating edges. We construct a graph with only $\frac{2n^2}{33}$ K_4-saturating edges. Furthermore, we prove that it is best possible, i.e., one can always find at least $(1 + o(1))\frac{2n^2}{33}$ K_4-saturating edges in an n-vertex K_4-free graph with $\lfloor \frac{n^2}{4} \rfloor + 1$ edges.

2.1 Introduction

Mantel [67] showed that the maximum number of edges in an n-vertex triangle-free graph is $\lfloor \frac{n^2}{4} \rfloor$. Rademacher in 1941 (unpublished) extended this result by showing that any n-vertex graph with $\lfloor \frac{n^2}{4} \rfloor + t$ edges contains at least $t\lfloor \frac{n}{2} \rfloor$ triangles, for $t = 1$. Lovász and Simonovits [65], improving a result of Erdős [33], proved this for every $t \leq \frac{n}{2}$. Erdős [34] showed analogue results for cliques, and Mubayi [71, 72] proved relevant results for color-critical graphs and some hypergraphs.

In general, *Erdős-Rademacher-type* problem is defined as follows: for any extremal question, what is the number of forbidden configurations appearing in a graph somewhat denser than the extremal graph? This type of problems have been studied in various contexts: A *book* of size q consists of q triangles sharing a common edge. Khadžiivanov and Nikiforov [58] showed that any n-vertex graph with $\lfloor \frac{n^2}{4} \rfloor + 1$ edges contains a book of size at least $\frac{n}{6}$. In the context of Sperner's Theorem [85], Kleitman [59] determined the minimum number of 2-chains in a poset whose size is larger than its largest anti-chain. Recently, this theorem was extended to k-chains by Das, Gan and Sudakov [29].

Let G be an n-vertex K_4-free graph. An edge in its complement $\overline{G}$ is a K_4-*saturating* edge if the addition of this edge to G creates a copy of K_4. Denote by $f(G)$ the number of K_4-saturating edges in $\overline{G}$ and by $f(n, e)$ the maximum integer ℓ such that every n-vertex K_4-free graph with e edges must have at least ℓ K_4-saturating edges. The first extremal result related to clique-saturating edges was by Bollobás [21], who proved that if every edge in $\overline{G}$ is a K_r-

saturating edge, then $e(G) \geq \binom{n}{2} - \binom{n-r+2}{2}$ and this bound is best possible. Later the result in [21] was extended by Alon [1], Frankl [43] and Kalai [57] using linear algebraic method. Saturation problems have been phrased in the language of 'graph bootstrap percolation', see [6] and [7] for recent developments.

In the case of K_4, Bollobás' example [21] is the following: let F be an n-vertex K_4-free graph obtained from adding an edge to the partite set of size two in $K_{2,n-2}$. This graph has only $2n-3$ edges, and yet all edges in $\overline{F}$ are K_4-saturating edges. To the other extreme, $K_{\lceil \frac{n}{2} \rceil, \lfloor \frac{n}{2} \rfloor}$ shows that a graph could have up to $\lfloor \frac{n^2}{4} \rfloor$ edges with no K_4-saturating edge, i.e. $f(n, \lfloor \frac{n^2}{4} \rfloor) = 0$. Erdős and Tuza [35] conjectured that if a K_4-free graph G has $\lfloor \frac{n^2}{4} \rfloor + 1$ edges, then suddenly there are at least $(1+o(1))\frac{n^2}{16}$ K_4-saturating edges. They also stated, without giving any specific example, that there is a graph with at most $(1+o(1))\frac{n^2}{16}$ K_4-saturating edges. Our guess is the following: add a new vertex and make it adjacent to roughly half of vertices in each partite set of $K_{\lceil \frac{n}{2} \rceil - 1, \lfloor \frac{n}{2} \rfloor}$. This conjecture can be considered, as formulated before, an Erdős-Rademacher-type problem concerning the number of K_4-saturating edges.

Conjecture 2.1.1 (Erdős-Tuza [35])**.**

$$f\left(n, \left\lfloor \frac{n^2}{4} \right\rfloor + 1\right) = (1+o(1))\frac{n^2}{16}.$$

We disprove this conjecture. We give a counterexample with only $\frac{2n^2}{33}$ K_4-saturating edges. Furthermore, we prove that $(1+o(1))\frac{2n^2}{33}$ is best possible, that is, one can always find at least $(1+o(1))\frac{2n^2}{33}$ K_4-saturating edges in an n-vertex K_4-free graph with $\lfloor \frac{n^2}{4} \rfloor + 1$ edges.

Theorem 2.1.2. *For $n \geq 73$,*

$$\frac{2n^2}{33} - \frac{3n}{11} \leq f\left(n, \left\lfloor \frac{n^2}{4} \right\rfloor + 1\right) \leq \frac{2n^2}{33} - \frac{7n}{33}.$$

We shall prove the following theorem, which implies the lower bound in Theorem 2.1.2.

Theorem 2.1.3. *Let G be an n-vertex K_4-free graph with $\lfloor \frac{n^2}{4} \rfloor$ edges, for $n \geq 73$. If G contains a triangle, then*

$$f(G) \geq \frac{2n^2}{33} - \frac{3n}{11}.$$

This is best possible when n is divisible by 66.

Proof of Theorem 2.1.2. The upper bound is by the construction described in Section 2.2. For the lower bound, let G be a K_4-free graph with $\lfloor \frac{n^2}{4} \rfloor + 1$ edges. By Mantel's theorem, it contains a triangle. Let G' be a subgraph obtained from G by removing an edge such that

G' contains a triangle. By Theorem 2.1.3, $f(G') \geq \frac{2n^2}{33} - \frac{3n}{11}$. The relation $f(G) \geq f(G')$ completes the proof. $\square$

Remark: (i) A slight modification of our proof gives the following stability result: Given any K_4-free graph G with $(1-o(1))\frac{n^2}{4}$ edges, if G contains a triangle, then $f(G) \geq (1-o(1))\frac{2n^2}{33}$.

(ii) Unlike the case about the number of triangles in [33] and [65], where each additional edge, up to $\frac{n}{2}$ edges, gurantees $\lfloor \frac{n}{2} \rfloor$ additional triangles, in our problem, even with linear many extra edges, the number of K_4-saturating edges is still at most $(1+o(1))\frac{2n^2}{33}$. In particular, $f\left(n, \lfloor \frac{n^2}{4} \rfloor + t\right) = \frac{2n^2}{33} + O(n)$ for $1 \leq t \leq \frac{n}{66}$.

(iii) One might define a K_r-saturating edge of a graph G, for $r \geq 5$, as we did for K_4. Denote by $\text{ex}(n, K_{r-1})$ the maximum size of an n-vertex K_{r-1}-free graph. We think that a similar phenomenon holds: if G is K_r-free and $e(G) = \text{ex}(n, K_{r-1}) + 1$, then the number of K_r-saturating edges is at least $\left(\frac{2(r-3)^2}{(r-1)(4r^2-19r+23)} + o(1)\right) n^2$. A generalization of our construction shows that if the conjecture is true, then it is best possible. (The construction is an appropriate blow-up of the following graph: take a new vertex and make it adjacent to exactly one vertex in each partite set of a $(r-2)$-partite complete graph $K_{2,\ldots,2}$.) Some of the ideas of our proof works for $r \geq 5$ as well, but some does not.

This chapter is organized as follows: We give a construction for the upper bound in Theorem 2.1.2 and an extremal example for Theorem 2.1.3 in Section 2.2. The proof for Theorem 2.1.3 is given in Section 2.3. We will omit floors and ceilings when it is not critical and we make no effort optimizing some of the constants.

2.2 Upper bound constructions

Fix an integer n divisible by 66. We present an n-vertex K_4-free graph H with $\frac{n^2}{4} + \frac{n}{66}$ edges and $f(H) = \frac{2n^2}{33} - \frac{7n}{33}$. Note that from this graph one can easily remove $\frac{n}{66} - 1$ edges without changing the number of K_4-saturating edges. We also give an extremal example showing the bound in Theorem 2.1.3 is best possible.

Construction for Theorem 2.1.2: To construct H, start with a C_5 on $\{v_1, v_2, v_3, v_4, v_5\}$ with a chord v_1v_3. Blow up each v_i to an independent set V_i of the following size: $|V_1| = |V_3| = \frac{16n}{66}$, $|V_2| = \frac{4n}{66} + 1$, $|V_4| = \frac{15n}{66}$ and $|V_5| = \frac{15n}{66} - 1$, see Figure 2.1. Then H is K_4-free with $\frac{n^2}{4} + \frac{n}{66}$ edges. The only K_4-saturating edges are those in V_1, V_2, V_3, which gives $f(H) = \frac{2n^2}{33} - \frac{7n}{33}$. The ratio of the blow-up is obtained by solving the following optimization problem. Set $x = \frac{|V_1|}{n} = \frac{|V_3|}{n}$, $y = \frac{|V_4|}{n} = \frac{|V_5|}{n}$ and $z = \frac{|V_2|}{n}$. Minimize $x^2 + \frac{z^2}{2}$, subject to (i) $2x + 2y + z = 1$, and (ii) $2xz + 2xy + x^2 + y^2 \geq \frac{1}{4}$.

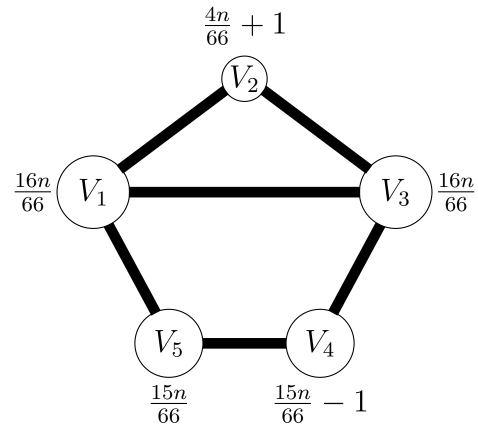

Figure 2.1: A K_4-free graph H with $e(H) = \frac{n^2}{4} + \frac{n}{66}$ and $f(H) = \frac{2n^2}{33} - \frac{7n}{33}$.

Construction for Theorem 2.1.3: Define H' the same way as H, except that $|V_2'| = \frac{4n}{66}$ and $|V_4'| = \frac{15n}{66}$. This graph is K_4-free with $\frac{n^2}{4}$ edges and $f(H') = \frac{2n^2}{33} - \frac{3n}{11}$.

2.3 Proof of Theorem 2.1.3

Let G be a K_4-free graph with $\frac{n^2}{4}$ edges and containing a triangle. Fix in G a maximum family of vertex-disjoint triangles, say $\mathcal{T} = \{T_1, T_2, ..., T_{tn}\}$, where $0 < t \leq \frac{1}{3}$. We write $V(\mathcal{T})$ for $\bigcup_{i=1}^{tn} V(T_i)$, $E(\mathcal{T})$ for $E(G[V(\mathcal{T})])$ and $e(\mathcal{T}) := |E(\mathcal{T})|$. Let $G' = G - V(\mathcal{T})$, since $\mathcal{T}$ is of maximum size, G' is a K_3-free graph with $e(G') \leq \frac{(1-3t)^2n^2}{4}$. Denote by r_1n^2 the number of K_4-saturating edges incident to $V(\mathcal{T})$, and by r_2n^2 the number of K_4-saturating edges in $V(G')$. Hence $f(G) = (r_1 + r_2)n^2$. First we give a lower bound on r_1.

Lemma 2.3.1.

$$r_1n^2 \geq \left(\frac{1}{4} - t + \frac{3t^2}{2}\right)n^2 - e(G') - \frac{3}{2}tn \geq \left(\frac{t}{2} - \frac{3t^2}{4}\right)n^2 - \frac{3}{2}tn.$$

Proof. Let $t_i = e(T_i, G \setminus \bigcup_{j=1}^{i} T_j)$, clearly $\sum_{i=1}^{tn} t_i = e(G) - e(G') - 3tn$. Since G is K_4-free, every vertex can have at most two neighbors in each triangle. Thus $t_i - (n - 3i)$ is a lower bound on the number of vertices in $G \setminus \bigcup_{j=1}^{i} T_j$ having degree 2 in T_i, each of which gives a K_4-saturating edge. Indeed, say $V(T_1) = \{x, y, z\}$, and $w \in N(x) \cap N(y)$, then wz is a

K_4-saturating edge. Thus,

$$\begin{aligned} r_1 n^2 &\geq \sum_{i=1}^{tn}(t_i-(n-3i)) = (e(G)-e(G')-3tn) - \left(tn^2 - 3\frac{tn(tn+1)}{2}\right) \\ &\geq \left(\frac{1}{4}-t+\frac{3t^2}{2}\right)n^2 - e(G') - \frac{3}{2}tn \geq \left(\frac{t}{2}-\frac{3t^2}{4}\right)n^2 - \frac{3}{2}tn, \end{aligned}$$

where the last inequality follows from $e(G') \leq \frac{(1-3t)^2n^2}{4}$. □

Let $T_i \in \mathcal{T}$ be a triangle in $\mathcal{T}$. Denote by $N_j(T_i) \subseteq V(G')$, for $0 \leq j \leq 3$, the set of vertices in G' that has exactly j neighbors in T_i. Since G is K_4-free, $N_3(T_i) = \emptyset$, for every T_i's. Further define $p_0(T_i) = \frac{|N_0(T_i)|}{n}$, $p_1(T_i) = \frac{|N_1(T_i)|}{n}$ and $p_2(T_i) = \frac{|N_2(T_i)|}{n}$. Thus by definition, $p_0(T_i) + p_1(T_i) + p_2(T_i) = 1 - 3t$.

The next lemma shows that there is a triangle $T \in \mathcal{T}$ with large $|N_2(T)|$.

Lemma 2.3.2. *There exists a triangle $T \in \mathcal{T}$, such that*
(i) $e(T, G') \geq \left(\frac{3}{2} - \frac{21t}{4}\right)n$, and
(ii) $p_2(T) \geq \frac{1}{2} - \frac{9t}{4} + p_0(T)$.

Proof. (i) The edge set of G can be partitioned into $E(G'), E(\mathcal{T}, G')$ and $E(\mathcal{T})$. Notice that since G is K_4-free, there are at most 6 edges between any pair of triangles in $\mathcal{T}$. Hence $e(\mathcal{T}) \leq 3tn + 6\binom{tn}{2} = 3t^2n^2$.

Thus we have $e(\mathcal{T}, G') = e(G) - e(G') - e(\mathcal{T}) \geq \frac{n^2}{4} - \frac{(1-3t)^2n^2}{4} - 3t^2n^2 \geq \left(\frac{3t}{2} - \frac{21t^2}{4}\right)n^2$. Therefore, there exists a triangle $T \in \mathcal{T}$ with $e(T, G') \geq e(\mathcal{T}, G')/(tn) \geq \left(\frac{3}{2} - \frac{21t}{4}\right)n$.
(ii) Let $T \in \mathcal{T}$ be a triangle satisfying (i). Note that $2p_2(T) + p_1(T) = \frac{e(T,G')}{n}$. Using $p_0(T) + p_1(T) + p_2(T) = 1 - 3t$, we have $p_2(T) - p_0(T) \geq \frac{3}{2} - \frac{21t}{4} - (1-3t) = \frac{1}{2} - \frac{9t}{4}$. □

From now on, we let $T = \{x, y, z\}$ be a triangle in $\mathcal{T}$ sending the most edges to G', hence it has the two properties of Lemma 2.3.2. For brevity we write $p_j = p_j(T)$ and $N_i = N_i(T)$ for $0 \leq j \leq 2$. Furthermore, define $A = N_{G'}(xy)$, $B = N_{G'}(yz)$ and $C = N_{G'}(xz)$. Note that A, B, C are pairwise disjoint independent sets, otherwise $T \cup A \cup B \cup C$ contains a copy of K_4. Define $N_x := N_{G'}(x), N_y := N_{G'}(y)$ and $N_z := N_{G'}(z)$. Let $a = \frac{|A|}{|N_2|}$, $b = \frac{|B|}{|N_2|}$ and $c = \frac{|C|}{|N_2|}$, thus $a + b + c = 1$. For $1 \leq k \leq 3$, we say that T spans a *k-joint-book*, if among A, B, C, exactly $3 - k$ of them are empty sets.

Lemma 2.3.3. *If T spans a 3-joint-book, then we have*

$$r_2 n^2 \geq \frac{1}{6}\left[\frac{3}{2} - \frac{21t}{4}\right]^2 n^2 - e(\overline{G'}) - (1-3t)n.$$

Proof. First notice that N_x, N_y and N_z are all independent sets. Indeed, suppose N_x contains an edge, then $T \cup N_x \cup B$ contains two vertex-disjoint triangles, contradicting the maximality of $\mathcal{T}$.

Note that $\binom{|N_x|}{2} + \binom{|N_y|}{2} + \binom{|N_z|}{2} \leq r_2 n^2 + e(\overline{G'})$. Indeed, every pair of vertices in N_x, N_y or N_z gives a non-edge in G' and those K_4-saturating edges in A, B, C are counted twice. Additionally, $|N_x| + |N_y| + |N_z| = e(T, G') \geq \left(\frac{3}{2} - \frac{21t}{4}\right) n$, and $e(T, G') \leq 2(1-3t)n$. Thus,

$$\begin{aligned} r_2 n^2 + e(\overline{G'}) &\geq \binom{|N_x|}{2} + \binom{|N_y|}{2} + \binom{|N_z|}{2} \geq 3\binom{e(T,G')/3}{2} \\ &= \frac{1}{6}(e(T,G'))^2 - \frac{1}{2}e(T,G') \geq \frac{n^2}{6}\left[\frac{3}{2} - \frac{21t}{4}\right]^2 - (1-3t)n. \end{aligned}$$

□

We first show that if T spans a 3-joint-book, then $f(G) \geq \frac{2n^2}{33} - \frac{3n}{11}$.

Lemma 2.3.4. *For $n \geq 73$, if T spans a 3-joint-book, then $f(G) \geq \frac{2n^2}{33} - \frac{3n}{11}$.*

Proof. Note that $e(G') + e(\overline{G'}) = \frac{(1-3t)^2 n^2}{2} - \frac{(1-3t)n}{2}$. By Lemmas 2.3.1 and 2.3.3, we have

$$\begin{aligned} f(G) &= (r_1 + r_2)n^2 \geq \left(\frac{1}{4} - t + \frac{3t^2}{2}\right)n^2 - e(G') - \frac{3}{2}tn \\ &+ \frac{1}{6}\left[\frac{3}{2} - \frac{21t}{4}\right]^2 n^2 - e(\overline{G'}) - (1-3t)n \\ &\geq \left(\frac{51t^2}{32} - \frac{5t}{8} + \frac{1}{8}\right)n^2 - \frac{n}{2} \geq \frac{13n^2}{204} - \frac{n}{2} \geq \frac{2n^2}{33} - \frac{3n}{11}, \end{aligned}$$

since $\frac{51t^2}{32} - \frac{5t}{8} + \frac{1}{8} \geq \frac{13}{204}$ when $0 < t \leq \frac{1}{3}$, and the last inequality holds for $n \geq 73$. □

Proof of Theorem 2.1.3. By Lemma 2.3.4, we may assume that T spans a k-joint-book with $k \leq 2$. Without loss of generality assume that $B = \emptyset$, i.e. $b = 0$. Then $a + c = 1$ and $|A| + |C| = p_2 n$. Notice that each pair of vertices in A and C is a K_4-saturating edge, hence

$$r_2 n^2 \geq \binom{|A|}{2} + \binom{|C|}{2} \geq 2\binom{p_2 n/2}{2} = \frac{p_2^2}{4}n^2 - \frac{p_2 n}{2}. \tag{2.1}$$

If $t \geq \frac{1}{5}$, then Lemma 2.3.1 implies $f(G) \geq r_1 n^2 \geq \left(\frac{t}{2} - \frac{3t^2}{4}\right)n^2 - \frac{n}{2} \geq \frac{2n^2}{33}$ for $n \geq 54$. Thus we may assume that $t < \frac{1}{5}$. The right hand side in (2.1) is minimized when p_2 is at its lower bound provided by Lemma 2.3.2, as $\frac{1}{2} - \frac{9t}{4} > \frac{1}{n}$ for $n \geq 20$. Hence

$$r_2 n^2 \geq \frac{1}{4}\left(\frac{1}{2} - \frac{9t}{4}\right)^2 n^2 - \frac{1}{2}\left(\frac{1}{2} - \frac{9t}{4}\right)n.$$

Therefore using Lemma 2.3.1, we have

$$\begin{aligned} f(G) &= (r_1+r_2)n^2 \geq \left(\left(\frac{t}{2}-\frac{3t^2}{4}\right)+\frac{1}{4}\left(\frac{1}{2}-\frac{9t}{4}\right)^2\right)n^2-\frac{1}{2}\left(3t+\frac{1}{2}-\frac{9t}{4}\right)n \\ &= \left(\frac{33t^2}{64}-\frac{t}{16}+\frac{1}{16}\right)n^2-\frac{1}{2}\left(\frac{3t}{4}+\frac{1}{2}\right)n \geq \frac{2n^2}{33}-\frac{3n}{11}-\frac{3}{44}, \end{aligned}$$

where the function on the right hand side is minimized at $t=\frac{2}{33}+\frac{4}{11n}$. Since both tn and $f(G)$ are integers, checking all n modulo 33, we have

$$f(G) \geq \frac{2n^2}{33}-\frac{3n}{11}.$$

We remark that the extremal example corresponds to the last case when $t<\frac{1}{5}$ and T spans a 2-joint book with $|A|=|C|$. $\square$

Chapter 3

Forbidding large cliques

In 1987, Kolaitis, Prömel and Rothschild [63] proved that, for every fixed $r \in \mathbb{N}$, almost every n-vertex K_{r+1}-free graph is r-partite. In this chapter we extend this result to all functions $r = r(n)$ with $r \leq (\log n)^{1/4}$. The proof combines a new (close to sharp) supersaturation version of the Erdős–Simonovits stability theorem [40], the hypergraph container method developed by Balogh, Morris and Samotij [17], and by Saxton and Thomason [81]; and a counting technique developed by Balogh, Bollobás and Simonovits [9].

3.1 Introduction

Determining the extremal properties of graphs which avoid a clique of a given size is one of the oldest problems in combinatorics, going back to the early paper of Mantel [67] and the groundbreaking work of Ramsey [75], Erdős and Szekeres [42] and Turán [88] over 70 years ago. The study of the *typical* properties of such graphs was initiated by Erdős, Kleitman and Rothschild [38], who proved in 1976 that almost all triangle-free graphs on n vertices are bipartite.[1] This result was extended to K_{r+1}-free graphs, for every fixed $r \in \mathbb{N}$, ten years later by Kolaitis, Prömel and Rothschild [63], who showed that almost all such graphs are r-partite. Various extensions of this theorem have since been obtained, see for example [9, 74] for work on other forbidden subgraphs, and [18, 73] for a sparse analogue.

In this chapter we extend the result of Kolaitis, Prömel and Rothschild [63] in a different direction, to K_{r+1}-free graphs where $r = r(n)$ is a function which is allowed to grow with n. More precisely, we prove the following theorem.[2]

Theorem 3.1.1. *Let $r = r(n) \in \mathbb{N}_0$ be a function satisfying $r \leq (\log n)^{1/4}$ for every $n \in \mathbb{N}$. Then almost all K_{r+1}-free graphs on n vertices are r-partite.*

Note that if $r \geq 2\log_2 n$ then almost all graphs are K_{r+1}-free (and almost none are r-partite if $r \ll n/\log n$), so the bound on r in Theorem 3.1.1 is not far from being best

[1] That is, the proportion of n-vertex triangle-free graphs that are not bipartite goes to zero as $n \to \infty$.

[2] All logs are natural unless otherwise stated.

possible. It would be extremely interesting (and likely very difficult) to determine the largest $\alpha \in [1/4, 1]$ such that the theorem holds for some function $r = (\log n)^{\alpha+o(1)}$. It may well be the case that this supremum is equal to 1, though we are not prepared to state this as a conjecture.

Theorem 3.1.1 improves a recent result of Mousset, Nenadov and Steger [70], who showed that, for the same[3] family of functions $r = r(n)$, the number of n-vertex K_{r+1}-free graphs is

$$2^{t_r(n)+o(n^2/r)}, \tag{3.1}$$

where $t_r(n) = \mathbf{ex}(n, K_{r+1})$ denotes the number of edges of the Turán graph, the r-partite graph on n vertices with the maximum possible number of edges. A bound of this type for fixed $r \in \mathbb{N}$ was originally proved in [38], and extended to an arbitrary (fixed) forbidden graph H in [37]. The problem for H-free graphs with $v(H) \to \infty$ as $n \to \infty$ was first studied by Bollobás and Nikiforov [22], who proved bounds corresponding to (3.1) whenever $v(H) = o(\log n)$ and $\chi(H_n) = r + 1$ is fixed. For more precise bounds for a fixed forbidden graph H, see [8], and for similar bounds in the hereditary (i.e., induced-H-free) setting, see [2, 10, 23] and the references therein.

The proof of Theorem 3.1.1 has three main ingredients. The first is the so-called 'hypergraph container method', which was recently developed by Balogh, Morris and Samotij [17], and independently by Saxton and Thomason [81]. This method was used by Mousset, Nenadov and Steger [70] to prove Theorem 3.3.2, below, from which they deduced the bound (3.1) using a supersaturation theorem of Lovász and Simonovits [65].

In order to obtain the much more precise result stated in Theorem 3.1.1, we will use the method of Balogh, Bollobás and Simonovits [8, 9], who determined the structure of almost all H-free graphs for every fixed graph H. This powerful technique (see Sections 3.4 and 3.5) allows one to compare the number of K_{r+1}-free graphs that are 'close' to being r-partite, with the total number of K_{r+1}-free graphs.

The missing ingredient is the main new contribution of this work. In order to deduce from Theorem 3.3.2 a bound on the number of K_{r+1}-free graphs that are 'far' from being r-partite, we will need an analogue of the Lovász–Simonovits [65] supersaturation result, mentioned above, for the well-known stability theorem of Erdős and Simonovits [40]. Although a weak such analogue can easily be obtained via the regularity lemma, this gives bounds which are far from sufficient for our purposes. Instead we will adapt an argument due to Füredi [44] in order to prove the following close-to-best-possible such result. We say that a graph G is

[3] In fact, a very slightly weaker theorem was stated in [70], but a little additional case analysis easily gives the result for all $r \leq (\log n)^{1/4}$.

t-far from being r-partite[4] if $\chi(G') > r$ for every subgraph $G' \subset G$ with $e(G') > e(G) - t$.

Theorem 3.1.2. *For every $n, r, t \in \mathbb{N}$, the following holds. Every graph G on n vertices which is t-far from being r-partite contains at least*

$$\frac{n^{r-1}}{e^{2r} \cdot r!} \left(e(G) + t - \left(1 - \frac{1}{r}\right) \frac{n^2}{2} \right)$$

copies of K_{r+1}.

Note that the graph obtained by adding t edges to the Turán graph $T_r(n)$ is t-far from being r-partite and has roughly $t \cdot (n/r)^{r-1}$ copies of K_{r+1}, so Theorem 3.1.2 is sharp to within a factor of roughly e^r. We remark also that the Erdős–Simonovits stability theorem [40] for an arbitrary graph H follows from Theorem 3.1.2 together with the well-known result of Erdős [36] that the Turán density of any k-partite k-uniform hypergraph is zero. Indeed, given $c > 0$, every graph G with n vertices and $e(G) \geq t_r(n) - cn^2$ edges that is $2cn^2$-far from being r-partite contains at least εn^{r+1} copies of K_{r+1} for some $\varepsilon = \varepsilon(c, r) > 0$. By the result of Erdős [36], it follows that G contains a copy of $K_{r+1}(s)$, the s-blow-up of K_{r+1}, as long as $n \geq n_0(c, r, s)$ is sufficiently large. We would like to thank Wojciech Samotij for pointing out to us this consequence of Theorem 3.1.2.

We will prove Theorem 3.1.2 in Section 3.2, and use it in Section 3.3 to count the K_{r+1}-free graphs that are n^{2-1/r^2}-far from being r-partite. We prove various simple properties of almost all K_{r+1}-free graphs in Section 3.4, and finally, in Section 3.5, we use the Balogh–Bollobás–Simonovits method to deduce Theorem 3.1.1.

3.2 A supersaturated Erdős-Simonovits stability theorem

In this section, we prove our 'supersaturated stability theorem' for K_{r+1}-free graphs. As noted in the Introduction, we do so by adapting a proof of Füredi [44].

Given a graph G, a vertex $v \in V(G)$ and an integer $m \in \mathbb{N}$, let us write $K_m(G)$ for the number of m-cliques in G, and $K_m(v)$ for the number of such m-cliques containing v.

Proof of Theorem 3.1.2. We will prove by induction on r that

$$K_{r+1}(G) \geq \frac{n^{r-1}}{c(r)} \left(e(G) + t - \left(1 - \frac{1}{r}\right) \frac{n^2}{2} \right), \tag{3.2}$$

[4]Similarly, we say that G is *t-close to being r-partite* if it is not t-far from being r-partite.

where $c(r) := 2(r+1)^{r-1}r^{r-1}/r!$, for every graph G on n vertices that is t-far from being r-partite. Since $c(r) \le e^{2r}r!$, the theorem follows from (3.2).

Note first that the theorem holds in the case $r = 1$, since a graph is t-far from being 1-partite if and only if $e(G) \ge t$, and hence G has at least $\frac{e(G)+t}{2}$ copies of K_2, as required. So let $r \ge 2$ and assume that the result holds for $r-1$. Let $n, t \in \mathbb{N}$, and let G be a graph that is t-far from being r-partite.

First, for each $v \in V(G)$, set $B_v = N(v)$ (the set of neighbours of v in G) and $A_v = V(G) \setminus B_v$, and observe that

$$\sum_{u \in A_v} d(u) = e(G) + e(A_v) - e(B_v), \tag{3.3}$$

where $e(X)$ denotes the number of edges in the graph $G[X]$. Now, the graph $G[B_v]$ is $\big(t - e(A_v)\big)$-far from being $(r-1)$-partite, and so, by the induction hypothesis,

$$K_{r+1}(v) \ \ge \ \frac{|B_v|^{r-2}}{c(r-1)} \left(e(B_v) + t - e(A_v) - \left(1 - \frac{1}{r-1}\right) \frac{|B_v|^2}{2} \right), \tag{3.4}$$

since each copy of K_r in $G[B_v]$ corresponds to a copy of K_{r+1} in G that contains v.

Combining (3.3) and (3.4), noting that $|B_v| = d(v)$, and summing over v, it follows that

$$(r+1) \cdot K_{r+1}(G) \ \ge \sum_{v \in V(G)} \frac{d(v)^{r-2}}{c(r-1)} \left(e(G) + t - \sum_{u \in A_v} d(u) - \left(1 - \frac{1}{r-1}\right) \frac{d(v)^2}{2} \right). \tag{3.5}$$

We claim that

$$\sum_{v \in V(G)} \sum_{u \in A_v} d(u) d(v)^{r-2} \le \sum_{v \in V(G)} \sum_{u \in A_v} d(v)^{r-1} = \sum_{v \in V(G)} d(v)^{r-1}\big(n - d(v)\big). \tag{3.6}$$

Indeed, let $X = \big\{(v,u) : v \in V(G),\ u \in A_v\big\}$ denote the set of ordered pairs in the sum above, and note that $(v,u) \in X$ if and only if $uv \notin E(G)$. Since X is symmetric, the inequality in (3.6) is in fact an equality for $r = 2$, and for $r = 3$ we apply the Cauchy-Schwarz inequality to obtain

$$\sum_{(v,u) \in X} d(u)d(v) \le \bigg(\sum_{(v,u) \in X} d(u)^2 \bigg)^{1/2} \bigg(\sum_{(v,u) \in X} d(v)^2 \bigg)^{1/2}.$$

For $r \geq 4$, applying Hölder's inequality[5] with $p = r-2$ and $q = (r-2)/(r-3)$ gives

$$\sum_{(v,u)\in X} d(u)d(v)^{r-2} \leq \bigg(\sum_{(v,u)\in X} d(u)^{r-2}d(v) \bigg)^{1/p} \bigg(\sum_{(v,u)\in X} d(v)^{r-1} \bigg)^{1/q},$$

since $\big(r-2-\frac{1}{r-2}\big)\frac{r-2}{r-3} = \frac{r^2-4r+3}{r-3} = r-1$. Once again using the symmetry of X, and noting that $1 - 1/p = 1/q$, the claimed inequality (3.6) follows.

Combining the inequalities (3.5) and (3.6), we obtain

$$(r+1)\cdot K_{r+1}(G) \;\geq \sum_{v\in V(G)} \frac{d(v)^{r-2}}{c(r-1)}\bigg(e(G) + t - d(v)n + \bigg(1 + \frac{1}{r-1}\bigg)\frac{d(v)^2}{2}\bigg).$$

Since the factor in parentheses is minimized when $d(v) = \frac{r-1}{r}\cdot n$, it follows that

$$(r+1)\cdot K_{r+1}(G) \;\geq \sum_{v\in V(G)} \frac{d(v)^{r-2}}{c(r-1)}\bigg(e(G) + t - \bigg(1 - \frac{1}{r}\bigg)\frac{n^2}{2}\bigg).$$

Finally, note that every graph G is $\big(e(G)/r\big)$-close to being r-partite (take a random partition), and hence we may assume that $\big(1+\frac{1}{r}\big)e(G) \geq \big(1-\frac{1}{r}\big)\frac{n^2}{2}$, since otherwise the theorem is trivial. Thus, by the convexity of x^{r-2},

$$\sum_{v\in V(G)} d(v)^{r-2} \geq n\cdot\bigg(\frac{2e(G)}{n}\bigg)^{r-2} \geq \bigg(\frac{r-1}{r+1}\bigg)^{r-2} n^{r-1},$$

and so, since $c(r-1)\cdot(r+1)^{r-1} = c(r)\cdot(r-1)^{r-2}$, it follows that

$$K_{r+1}(G) \;\geq\; \frac{n^{r-1}}{c(r)}\bigg(e(G) + t - \bigg(1 - \frac{1}{r}\bigg)\frac{n^2}{2}\bigg),$$

as claimed. □

3.3 An approximate structural result

In this section we will prove the following approximate version of Theorem 3.1.1.

Theorem 3.3.1. *Let $r = r(n) \in \mathbb{N}$ be a function satisfying $r \leq (\log n)^{1/4}$ for each $n \in \mathbb{N}$. Then almost all K_{r+1}-free graphs on n vertices are n^{2-1/r^2}-close to being r-partite.*

---

[5]The discrete version of Hölder's inequality states that $\sum_{i=1}^n |x_iy_i| \leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}\left(\sum_{i=1}^n |y_i|^q\right)^{1/q}$ for $n\in\mathbb{N}$, $x,y\in\mathbb{R}^n$ and $p,q>1$ satisfying $1/p+1/q=1$.

Theorem 3.3.1 is a straightforward consequence of Theorem 3.1.2 and the following version of 'container' theorem in [70] using the hypergraph container method of Balogh, Morris and Samotij [17] and Saxton and Thomason [81]. The following theorem is slightly stronger than the result stated in [70], but follows easily from essentially the same proof. We remark that the deduction of this theorem from the main results of [17, 81] is the only point in the proof of Theorem 3.1.1 where we use our assumption that $r \le (\log n)^{1/4}$, although in Sections 3.4 and 3.5 we will require a similar (but somewhat weaker) upper bound on r.

Theorem 3.3.2. *Let $r = r(n) \in \mathbb{N}$ be a function satisfying $r \le (\log n)^{1/4}$. Then there exists a collection $\mathcal{C}$ of graphs such that the following hold for each sufficiently large $n \in \mathbb{N}$:*

(a) every K_{r+1}-free graph on n vertices is a subgraph of some $G \in \mathcal{C}_n$,

(b) $K_{r+1}(G) \le n^{r+1-2/r^2}$ for every $G \in \mathcal{C}_n$, and

(c) $|\mathcal{C}_n| \le \exp\left(n^{2-2/r^2}\right)$,

where $\mathcal{C}_n = \big\{G \in \mathcal{C} : v(G) = n\big\}$.

Deducing Theorem 3.3.1 from Theorems 3.1.2 and 3.3.2 is straightforward.

Proof of Theorem 3.3.1. For each $t \in \mathbb{N}$, set

$$\mathcal{F}_t = \left\{ G \,:\, e(G) \ge \left(1 - \frac{1}{r}\right) \frac{v(G)^2}{2} - \frac{t}{2} \text{ and } G \text{ is } t\text{-far from being } r\text{-partite} \right\},$$

and observe that if $G \in \mathcal{F}_t$, then

$$K_{r+1}(G) \ge \frac{v(G)^{r-1} \cdot t}{e^{2r+1} \cdot r!},$$

by Theorem 3.1.2. Therefore, letting $\mathcal{C}$ be the collection of graphs given by Theorem 3.3.2, and setting $t = n^{2-1/r^2}$, it follows from property (b) and the bound $r \le (\log n)^{1/4}$ that we have $\mathcal{C}_n \cap \mathcal{F}_t = \emptyset$ for all sufficiently large $n \in \mathbb{N}$.

Now, for each K_{r+1}-free graph G on n vertices that is n^{2-1/r^2}-far from being r-partite, we have $G \subset C$ for some $C \in \mathcal{C}_n$, and by the observations above and the definition of $\mathcal{F}_t$, it follows that

$$e(C) \le \left(1 - \frac{1}{r}\right) \frac{n^2}{2} - \frac{t}{2}.$$

Therefore, summing over all members of $\mathcal{C}_n$, the number of such graphs is at most

$$\exp\left(n^{2-2/r^2}\right) \cdot 2^{t_r(n) - t/2} \ll 2^{t_r(n) - t/4},$$

which is clearly smaller than the number of K_{r+1}-free graphs on n vertices, as required. $\square$

3.4 Some properties of a typical K_{r+1}-free graph

In this section we will prove some useful structural properties of almost all K_{r+1}-free graphs. These structural properties will allow us (in Section 3.5) to count the K_{r+1}-free graphs that are close to being r-partite, and hence to complete the proof of Theorem 3.1.1. We emphasize that the lemmas in this section were all proved for fixed $r \in \mathbb{N}$ in [8], and no extra ideas are required in order to extend their proofs to our more general setting.

Let us fix throughout this section a function $2 \le r = r(n) \le (\log n)^{1/4}$, and let us denote by $\mathcal{G}$ the collection of K_{r+1}-free graphs on n vertices that are n^{2-1/r^2}-close to being r-partite, where n is assumed to be sufficiently large. We begin with two simple definitions.

Definition 3.4.1 (Optimal partitions). An r-partition $(U_1, \ldots, U_r)$ of the vertex set of a graph G is called *optimal* if the number of interior edges, $\sum_{i=1}^{r} e(U_i)$, is minimized.

Definition 3.4.2 (Uniformly dense graphs). We say that a graph G is *uniformly dense* if for every optimal r-partition $(U_1, \ldots, U_r)$ and every $i, j \in [r]$ with $i \neq j$, we have

$$e(A, B) > \frac{|A||B|}{32} \tag{3.7}$$

for every $A \subset U_i$ and $B \subset U_j$ with $|A| = |B| \ge 2^{-10r} n$.

Lemma 3.4.3. *The number of graphs in $\mathcal{G}$ that are not uniformly dense is at most*

$$2^{t_r(n) - 2^{-22r} n^2},$$

and therefore almost all K_{r+1}-free graphs are uniformly dense.

Proof. In order to count such graphs, we first choose the optimal partition $\mathcal{U} = (U_1, \ldots, U_r)$, the pair $\{i, j\} \subset [r]$, and the sets $A \subset U_i$ and $B \subset U_j$ for which (3.7) fails. We then choose the edges between A and B, and finally the remaining edges. Note first that we have at most r^n choices for $\mathcal{U}$, at most r^2 choices for $\{i, j\}$, and at most 2^{2n} choices for the pair (A, B).

Now, the number of choices for the edges between A and B is at most

$$\sum_{k=0}^{|A||B|/32} \binom{|A||B|}{k} \le n^2 (32e)^{|A||B|/32} \le 2^{|A||B|/4},$$

and the number of choices for the remaining edges is at most

$$2^{t_r(n)-|A||B|}\binom{n^2}{n^{2-1/r^2}} \;\le\; 2^{t_r(n)-|A||B|}\exp\Big(n^{2-1/r^2}\log n\Big) \;\le\; 2^{t_r(n)-|A||B|/2},$$

since $\mathcal{U}$ is optimal, $|A||B| \ge 2^{-20r}n^2$, and each $G \in \mathcal{G}$ is n^{2-1/r^2}-close to being r-partite.

It follows that the number of graphs in $\mathcal{G}$ that are not uniformly dense is at most

$$r^{n+2} \cdot 2^{2n} \cdot 2^{t_r(n)-|A||B|/4} \;\le\; 2^{t_r(n)-2^{-22r}n^2},$$

as claimed. □

Our next definition controls the maximum degree inside the parts of an optimal partition.

Definition 3.4.4 (Internally sparse graphs)**.** A graph G is said to be *internally sparse* if, for every optimal partition $\mathcal{U} = (U_1, \ldots, U_r)$ of G, we have

$$\Delta\big(G[U_i]\big) \;\le\; 2^{-5r}n \tag{3.8}$$

for every $1 \le i \le r$. Otherwise we say that G is *internally dense.*

Lemma 3.4.5. *If $G \in \mathcal{G}$ is internally dense then it is not uniformly dense.*

We will prove Lemma 3.4.5 using the following embedding lemma[6] from [3].

Lemma 3.4.6. *Let $0 < \alpha < 1$, G be a graph, and $W_1, \ldots, W_r \subset V(G)$ be disjoint sets of vertices. Suppose that for every pair $\{i,j\} \subset [r]$ and every pair of sets $A \subset W_i$ and $B \subset W_j$ with $|A| \ge \alpha^r|W_i|$ and $|B| \ge \alpha^r|W_j|$, we have $e(A,B) > \alpha|A||B|$.*

Then G contains a copy of K_r with one vertex in each set W_j.

Proof of Lemma 3.4.5. Suppose for a contradiction that $G \in \mathcal{G}$ is both internally dense and uniformly dense. Let $\mathcal{U} = (U_1, \ldots, U_r)$ be the optimal partition given by Definition 3.4.4, and suppose that $v \in U_1$ has degree at least $2^{-5r}n$ in $G[U_1]$. For each $i \in [r]$, let $W_i = N(v) \cap U_i$, and observe that $|W_i| \ge 2^{-5r}n$, since $\mathcal{U}$ is optimal.

Observe that $W_1, \ldots, W_r$ satisfy the conditions of Lemma 3.4.6 with $\alpha = 1/32$, since G is uniformly dense, so $e(A,B) > |A||B|/32$ for every pair $\{i,j\} \subset [r]$, and every $A \subset U_i$ and $B \subset U_j$ with $|A| = |B| \ge 2^{-10r}n$. Thus, by Lemma 3.4.6, there exists a copy of K_r in the neighborhood of v, which (including v) gives a copy of K_{r+1} in G, but this is a contradiction,

[6] In fact, the version stated here is slightly more general than [3, Lemma 3.1], but follows from exactly the same proof.

since our graph is K_{r+1}-free. Thus, every internally dense graph $G \in \mathcal{G}$ is not uniformly dense, as claimed. $\square$

Our final definition controls the sizes of the parts in an optimal partition.

Definition 3.4.7 (Balanced graphs)**.** A graph G is said to be *balanced* if, for every optimal partition $\mathcal{U} = (U_1, \ldots, U_r)$ of G, we have

$$\frac{n}{r} - 2^{-5r} n \,\leq\, |U_i| \,\leq\, \frac{n}{r} + 2^{-5r} n \tag{3.9}$$

for every $1 \leq i \leq r$. Otherwise we say that G is *unbalanced*.

Lemma 3.4.8. *The number of unbalanced graphs in $\mathcal{G}$ is at most*

$$2^{t_r(n) - 2^{-12r} n^2},$$

and therefore almost all K_{r+1}-free graphs are balanced.

Proof. Let $G \in \mathcal{G}$ be an unbalanced graph, and let $\mathcal{U} = (U_1, \ldots, U_r)$ be an optimal partition of G for which (3.9) fails. Note that

$$\sum_{i=1}^{r-1} \sum_{j=i+1}^{r} |U_i||U_j| \,\leq\, t_r(n) - 2^{-11r} n^2,$$

since moving a vertex from a set of size at least $n/r + a$ to one of size $n/r - b$ creates at least $a + b - 1$ new potential cross edges. The number of such graphs $G \in \mathcal{G}$ is therefore at most

$$r^n \cdot 2^{t_r(n) - 2^{-11r} n^2} \cdot \binom{n^2}{n^{2-1/r^2}} \,\leq\, 2^{t_r(n) - 2^{-12r} n^2},$$

as claimed. $\square$

3.5 Proof of Theorem 3.1.1

In this section we will deduce Theorem 3.1.1 from Theorem 3.3.1, using the method of Balogh, Bollobás and Simonovits [8, 9]. Recall from the previous section that almost all K_{r+1}-free graphs are uniformly dense, internally sparse and balanced.

Let us fix throughout this section a function $2 \leq r = r(n) \leq (\log n)^{1/4}$, and assume that n is sufficiently large.

Definition 3.5.1. Let $\mathcal{Q}(n,r)$ denote the collection of K_{r+1}-free graphs on n vertices that are not r-partite, but are n^{2-1/r^2}-close to being r-partite, and are moreover uniformly dense, internally sparse and balanced.

Let $\mathcal{K}(n,r)$ denote the collection of K_{r+1}-free graphs on n vertices. We will prove the following proposition, which completes the proof of Theorem 3.1.1.

Proposition 3.5.2. *For every sufficiently large $n \in \mathbb{N}$,*

$$|\mathcal{Q}(n,r)| \,\le\, 2^{-2^{-6r}n} \cdot |\mathcal{K}(n,r)|.$$

The idea of the proof is as follows. We will define a collection of bipartite graphs F_m (see Definition 3.5.8) with parts $\mathcal{Q}(n,r,m)$ and $\mathcal{K}(n,r)$, where the sets $\mathcal{Q}(n,r,m)$ form a partition of $\mathcal{Q}(n,r)$ (see Definitions 3.5.4 and 3.5.5). These bipartite graphs will have the following property: the degree in F_m of each $G \in \mathcal{Q}(n,r,m)$ will be significantly larger than the degree of each $G \in \mathcal{K}(n,r)$ (see Lemmas 3.5.10 and 3.5.12). The result will then follow by double counting the edges of each F_m and summing over m.

In order to define $\mathcal{Q}(n,r,m)$ and F_m, we will need the following simple concept.

Definition 3.5.3 (Bad sets)**.** Let G be a graph and let $U \subset V(G)$. A set of r vertices $R \subset V(G) \setminus U$ is said to be *bad* towards U if it has no common neighbor in U.

In the following definition we may choose the partition $\mathcal{U}$ and the sets $X^{(1)}, \dots, X^{(r)}$ arbitrarily, subject to the given conditions.

Definition 3.5.4. For each $G \in \mathcal{Q}(n,r)$, fix an optimal partition $\mathcal{U} = (U_1, \dots, U_r)$ of $V(G)$, and for each $j \in [r]$ choose a maximal collection of vertex-disjoint sets $X^{(j)} = \big\{R_1^{(j)}, \dots, R_{\ell(j)}^{(j)}\big\}$ such that $R_i^{(j)}$ is bad towards U_j for each $i \in [\ell(j)]$. We define

$$m(G) \,:=\, \max\big\{\ell(j) \,:\, j \in [r]\big\},$$

let $j(G)$ denote the smallest j for which this maximum is attained, and set

$$X(G) \,:=\, R_1^{(j(G))} \cup \cdots \cup R_{\ell(j(G))}^{(j(G))}.$$

With this definition in place, it is natural to partition $\mathcal{Q}(n,r)$ by the size of $m(G)$.

Definition 3.5.5. For each $m \in \mathbb{N}$, we define

$$\mathcal{Q}(n,r,m) \,=\, \big\{G \in \mathcal{Q}(n,r) \,:\, m(G) = m\big\}.$$

Before continuing, let us note a simple but key fact.

Lemma 3.5.6. *For every $G \in \mathcal{Q}(n,r)$, $m(G) \geq 1$.*

Proof. This follows from the fact that G is not r-partite. Indeed, suppose that $m(G) = 0$ and let $x_0 x_1 \in E(G[U_1])$ be an 'interior' edge of G with respect to $\mathcal{U}$. Since there are no bad r-sets towards U_j for any $j \in [r]$, we can recursively choose vertices $x_j \in U_j$ such that $\{x_0, \ldots, x_j\}$ forms a clique. But this is a contradiction, since G is K_{r+1}-free. □

In order to establish an upper bound on those m which we need to consider, we count those graphs in $\mathcal{Q}(n,r)$ for which $m(G)$ is large.

Lemma 3.5.7. *If $m \geq 2^{-8r} n$, then*

$$|\mathcal{Q}(n,r,m)| \leq 2^{t_r(n) - mn/2^{3r}}.$$

Proof. Let $m \geq 2^{-8r} n$, and consider the number of ways of constructing a graph $G \in \mathcal{Q}(n,r,m)$. We have at most r^n choices for the partition $\mathcal{U}$, at most $\binom{n}{r}^m$ choices for the set $X(G)$, and r choices for $j = j(G)$. Moreover, since $2^r - 1 \leq 2^r e^{-1/2^r}$, we have at most

$$2^{t_r(n) - |U_j||X(G)|} \big(2^r - 1\big)^{|U_j||X(G)|/r} \leq 2^{t_r(n) - mn/2^{2r}}$$

choices for the edges between different parts of $\mathcal{U}$, since $X(G)$ is composed of r-sets that are bad towards U_j, and G is balanced. Finally, we have at most $n^{O(n^{2-1/r^2})}$ choices for the edges inside parts of $\mathcal{U}$, since G is n^{2-1/r^2}-close to being r-partite.

It follows that

$$|\mathcal{Q}(n,r,m)| \leq r^n \cdot \binom{n}{r}^m \cdot r \cdot n^{O(n^{2-1/r^2})} \cdot 2^{t_r(n) - mn/2^{2r}} \leq 2^{t_r(n) - mn/2^{3r}}$$

as required, since $m \geq 2^{-8r} n$, so $n^{2-1/r^2} \log n \ll 2^{-3r} mn$. □

From now on, let us fix a function $1 \leq m = m(n) \leq 2^{-8r} n$. We are ready to define the bipartite graph F_m.

Definition 3.5.8. Define a map $\Phi_m \colon \mathcal{Q}(n,r,m) \to 2^{\mathcal{K}(n,r)}$ by placing $H \in \Phi_m(G)$ if and only if H can be constructed from G by first removing all edges of G that are incident to $X(G)$, and then adding an arbitrary subset of the edges between $X(G)$ and $V(G) \setminus \big(X(G) \cup U_{j(G)}\big)$.

Let F_m be the bipartite graph with edge set $\{(G,H) : H \in \Phi_m(G)\}$. Moreover, for each $H \in \mathcal{K}(n,r)$, let us write $\Phi_m^{-1}(H) = \{G \in \mathcal{Q}(n,r,m) : H \in \Phi_m(G)\}$.

We first observe that the map Φ_m is well-defined.

Lemma 3.5.9. *If $G \in \mathcal{Q}(n,r,m)$ and $H \in \Phi_m(G)$, then H is K_{r+1}-free.*

Proof. This follows easily from the fact that G is K_{r+1}-free, and the maximality of $X(G)$. Indeed, if there exists a copy of K_{r+1} in H, then it must contain a vertex of $X(G)$, and therefore it must contain no other vertices of $X(G) \cup U_{j(G)}$. Hence it contains exactly r vertices of $V(G) \setminus \big(X(G) \cup U_{j(G)}\big)$, and by the maximality of $X(G)$ these have a common neighbor in $U_{j(G)}$. But this contradicts our assumption that G is K_{r+1}-free, as required. □

We are now ready to prove our first bound on the degrees in F_m.

Lemma 3.5.10. *For every $G \in \mathcal{Q}(n,r,m)$,*

$$\log_2 |\Phi_m(G)| \,\ge\, \left(1 - \frac{1}{r} - \frac{1}{2^{5r}} - \frac{mr}{n}\right) mnr.$$

Proof. This follows immediately from the fact that G is balanced. Indeed, we have two choices for each of the

$$|X(G)| \cdot \big|V(G) \setminus \big(X(G) \cup U_{j(G)}\big)\big| \,\ge\, mr \cdot \left(1 - \frac{1}{r} - \frac{1}{2^{5r}} - \frac{mr}{n}\right) n \qquad (3.10)$$

potential edges between $X(G)$ and $V(G) \setminus \big(X(G) \cup U_{j(G)}\big)$. □

In order to bound the degrees in F_m of vertices in $\mathcal{K}(n,r)$, we will need the following lemma, which counts the optimal partitions in the neighborhood of such a vertex. We note that here, the upper bound on m from Lemma 3.5.7 is crucial.

Lemma 3.5.11. *For each $H \in \mathcal{K}(n,r)$, there are at most $2^{n/2^{3r}}$ distinct partitions $\mathcal{U}$ of $V(H)$ such that $\mathcal{U}$ is an optimal partition of some graph $G \in \Phi_m^{-1}(H)$.*

Proof. We will use the fact that each $G \in \Phi_m^{-1}(H)$ is uniformly dense and n^{2-1/r^2}-close to being r-partite to show that the optimal partitions in question must be 'close' to one another.

To be precise, let $G_1, G_2 \in \Phi_m^{-1}(H)$, and let $\mathcal{U} = (U_1, \dots, U_r)$ be an optimal partition of G_1 and $\mathcal{V} = (V_1, \dots, V_r)$ be an optimal partition of G_2. We claim that

$$\big|\big\{ j \in [r] \,:\, |U_i \cap V_j| > 2^{-8r} n + 2mr \big\}\big| \,\le\, 1$$

for every $i \in [r]$. Indeed, suppose that

$$\big|U_i \cap V_j\big| > 2^{-8r} n + 2mr \qquad \text{and} \qquad \big|U_i \cap V_{j'}\big| > 2^{-8r} n + 2mr,$$

set $A = \big(U_i \cap V_j\big) \setminus \big(X(G_1) \cup X(G_2)\big)$ and $B = (U_i \cap V_{j'}) \setminus \big(X(G_1) \cup X(G_2)\big)$, and note that, since G_2 is uniformly dense, we have $e_{G_2}(A,B) > |A||B|/32 > 2^{-16r-5}n^2$. But these edges are all contained in U_i, so this contradicts the fact that G_1 is n^{2-1/r^2}-close to being r-partite, as required.

It follows that (by renumbering the parts if necessary) we have

$$\big|U_i \setminus V_i\big| \,\le\, r \cdot \big(2^{-8r}n + 2mr\big) \,\le\, 2^{-6r}n$$

for every $i \in [r]$, where second inequality follows since $m \le 2^{-8r}n$. Set $D_i = U_i \setminus V_i$, and observe that the partition $\mathcal{V}$ and the collection $(D_1, \ldots, D_r)$ together determine $\mathcal{U}$. It follows that the number of optimal partitions is at most

$$\left(\sum_{k=0}^{2^{-6r}n} \binom{n}{k} \right)^r \le n^r \cdot \binom{n}{2^{-6r}n}^r \le 2^{r \log n} \cdot \big(e2^{6r}\big)^{r2^{-6r}n} \le 2^{n/2^{3r}}, \qquad (3.11)$$

as required. □

We can now bound the degrees on the right hand side of (3.11). Recall that in Definition 3.5.4 we chose a 'canonical' optimal partition for each graph $G \in \mathcal{Q}(n,r)$.

Lemma 3.5.12. *We have*

$$\log_2 \big|\Phi_m^{-1}(H)\big| \,\le\, \left(1 - \frac{1}{r} - \frac{1}{2^{4r}}\right) mnr$$

for every $H \in \mathcal{K}(n,r)$.

Proof. Let us fix a partition $\mathcal{U} = (U_1, \ldots, U_r)$, and count the number of graphs $G \in \mathcal{Q}(n,r,m)$ with $H \in \Phi_m(G)$ whose optimal partition is $\mathcal{U}$. To do so, first note that we have $\binom{n}{r}^m \le n^{mr}$ choices for $X(G)$, and at most r choices for $j = j(G)$. Now, since G is balanced, i.e., $\big||U_i| - n/r\big| \le n/2^{5r}$ for each $i \in [r]$, there are at most $(1 - 2/r + 2/2^{5r})n$ possible neighbours for each $v \in X(G)$ not in its own part of $\mathcal{U}$ or in U_j. Moreover, since G is internally sparse, each vertex $v \in X(G)$ has at most $2^{-5r}n$ neighbours in its own part of $\mathcal{U}$. Thus we have at most

$$2^{(1-2/r+2/2^{5r})n} \sum_{k=0}^{2^{-5r}n} \binom{n}{k} \,\le\, 2^{(1-2/r+1/2^{3r})n}$$

choices for the edges between each vertex $v \in X(G)$ and $V(G) \setminus U_j$, by bounding as in (3.11).

Finally, by the definition of bad sets, and since G is balanced, we have at most

$$(2^r - 1)^{(1/r+1/2^{5r})mn} \le 2^{(1/r+1/2^{5r})mnr} e^{-mn/r2^r} \le 2^{(1/r-3/2^{3r})mnr}$$

choices for the edges between $X(G)$ and U_j.

Since, by Lemma 3.5.11, we have at most $2^{n/2^{3r}}$ choices for the partition $\mathcal{U}$, it follows that

$$\begin{aligned} \log_2 \left|\Phi_m^{-1}(H)\right| &\le mr\log n + \log r + \left(1 - \frac{2}{r} + \frac{1}{2^{3r}} + \frac{1}{r} - \frac{3}{2^{3r}} + \frac{1}{2^{3r}}\right)mnr \\ &\le \left(1 - \frac{1}{r} - \frac{1}{2^{4r}}\right)mnr, \end{aligned}$$

as claimed. □

Finally we put the pieces together and prove Proposition 3.5.2.

Proof of Proposition 3.5.2. We claim first that

$$|\mathcal{Q}(n,r,m)| \le 2^{-2^{-5r}mnr} \cdot |\mathcal{K}(n,r)| \tag{3.12}$$

for every $m \le 2^{-8r}n$. To prove this, we simply double count the edges of F_m, using Lemmas 3.5.10 and 3.5.12. Indeed, we have

$$\log_2\left(\frac{|\mathcal{Q}(n,r,m)|}{|\mathcal{K}(n,r)|}\right) \le \left(1 - \frac{1}{r} - \frac{1}{2^{4r}}\right)mnr - \left(1 - \frac{1}{r} - \frac{1}{2^{5r}} - \frac{mr}{n}\right)mnr,$$

which implies (3.12) since $m \le 2^{-8r}n$.

Summing (3.12) over m, and recalling that G is n^{2-1/r^2}-close to being r-partite, we obtain

$$|\mathcal{Q}(n,r)| \le \sum_{m=1}^{2^{-8r}n} 2^{-2^{-5r}mnr} \cdot |\mathcal{K}(n,r)| + \sum_{m=2^{-8r}n}^{n} 2^{t_r(n) - mn/2^{3r}} \le 2^{-2^{-6r}n} \cdot |\mathcal{K}(n,r)|,$$

by Lemmas 3.5.6 and 3.5.7 (since $|\mathcal{K}(n,r)| \ge 2^{t_r(n)}$), as required. □

Finally, let us deduce Theorem 3.1.1.

Proof of Theorem 3.1.1. By Theorem 3.3.1, almost all K_{r+1}-free graphs on n vertices are n^{2-1/r^2}-close to r-partite. We further showed in Lemmas 3.4.3, 3.4.5, and 3.4.8 that almost all of these graphs are either r-partite, or in $\mathcal{Q}(n,r)$. Since by Proposition 3.5.2, for sufficiently large n, the size of $\mathcal{Q}(n, r)$ is (almost) exponentially small compared to $\mathcal{K}(n, r)$, it follows that almost all K_{r+1}-free graphs are r-partite, as required. $\square$

Chapter 4

Number of maximal sum-free sets, part I

Cameron and Erdős [26] raised the question of how many *maximal* sum-free sets there are in $\{1, \dots, n\}$, giving a lower bound of $2^{\lfloor n/4 \rfloor}$. In this chapter we prove that there are in fact at most $2^{(1/4+o(1))n}$ maximal sum-free sets in $\{1, \dots, n\}$. Our proof makes use of container and removal lemmas of Green [49, 48] as well as a result of Deshouillers, Freiman, Sós and Temkin [30] on the structure of sum-free sets.

4.1 Introduction

A fundamental notion in combinatorial number theory is that of a sum-free set: A set S of integers is *sum-free* if $x + y \notin S$ for every $x, y \in S$ (note x and y are not necessarily distinct here). The topic of sum-free sets of integers has a long history. Indeed, in 1916 Schur [83] proved that, if n is sufficiently large, then any r-colouring of $[n] := \{1, \dots, n\}$ yields a monochromatic triple x, y, z such that $x + y = z$.

Note that both the set of odd numbers in $[n]$ and the set $\{\lfloor n/2 \rfloor + 1, \dots, n\}$ are maximal sum-free sets. (A sum-free subset of $[n]$ is *maximal* if it is not properly contained in another sum-free subset of $[n]$.) By considering all possible subsets of one of these maximal sum-free sets, we see that $[n]$ contains at least $2^{\lceil n/2 \rceil}$ sum-free sets. Cameron and Erdős [25] conjectured that in fact $[n]$ contains only $O(2^{n/2})$ sum-free sets. The conjecture was proven independently by Green [49] and Sapozhenko [79]. Recently, a refinement of the Cameron–Erdős conjecture was proven in [4], giving an upper bound on the number of sum-free sets in $[n]$ of size m (for each $1 \leq m \leq \lceil n/2 \rceil$).

Let $f(n)$ denote the number of sum-free subsets of $[n]$ and $f_{\max}(n)$ denote the number of maximal sum-free subsets of $[n]$. Recall that the sum-free subsets of $[n]$ described above lie in just two maximal sum-free sets. This led Cameron and Erdős [26] to ask whether the number of maximal sum-free subsets of $[n]$ is "substantially smaller" than the total number of sum-free sets. In particular, they asked whether $f_{\max}(n) = o(f(n))$ or even $f_{\max}(n) \leq f(n)/2^{\varepsilon n}$ for some constant $\varepsilon > 0$. Łuczak and Schoen [66] answered this question, showing that $f_{\max}(n) \leq 2^{n/2-2^{-28}n}$ for sufficiently large n. More recently, Wolfovitz [90] proved that

$f_{\max}(n) \leq 2^{3n/8+o(n)}$.

In the other direction, Cameron and Erdős [26] observed that $f_{\max}(n) \geq 2^{\lfloor n/4 \rfloor}$. Indeed, let $m = n$ or $m = n-1$, whichever is even. Let S consist of m together with precisely one number from each pair $\{x, m-x\}$ for odd $x < m/2$. Then S is sum-free. Moreover, although S may not be maximal, no further odd numbers less than m can be added, so distinct S lie in distinct maximal sum-free subsets of $[n]$.

We prove that this lower bound is in fact, 'asymptotically', the correct bound on $f_{\max}(n)$.

Theorem 4.1.1. *There are at most $2^{(1/4+o(1))n}$ maximal sum-free sets in $[n]$. That is,*

$$f_{\max}(n) = 2^{(1/4+o(1))n}.$$

The proof of Theorem 4.1.1 makes use of 'container' and 'removal' lemmas of Green [48, 49] as well as a result of Deshouillers, Freiman, Sós and Temkin [30] on the structure of sum-free sets (see Section 4.2 for an overview of the proof).

Next we provide another collection of maximal sum-free sets in $[n]$. Suppose that $4|n$ and set $I_1 := \{n/2+1, \dots, 3n/4\}$ and $I_2 := \{3n/4+1, \dots, n\}$. First choose the element $n/4$ and a set $S \subseteq I_2$. Then for every $x \in I_2 \setminus S$, choose $x - n/4 \in I_1$. The resulting set is sum-free but may not be maximal. However, no further element in I_2 can be added, thus distinct S lie in distinct maximal sum-free sets in $[n]$. There are $2^{|I_2|} = 2^{n/4}$ ways to choose S.

Notation: Given a set $A \subseteq [n]$, denote by $f_{\max}(A)$ the number of maximal sum-free subsets of $[n]$ that lie in A and by $\min(A)$ the minimum element of A. Let $1 \leq p < q \leq n$ be integers, denote $[p, q] := \{p, p+1, \dots, q\}$. Denote by E the set of all even numbers in $[n]$ and by O the set of all odd numbers in $[n]$. A triple $x, y, z \in [n]$ is called a *Schur triple* if $x + y = z$ (here $x = y$ is allowed).

Throughout, all graphs considered are simple unless stated otherwise. We say that a graph G is a *graph possibly with loops* if G can be obtained from a simple graph by adding at most one loop at each vertex. Given a vertex x in G, we write $\deg_G(x)$ for the *degree of x in G*. Note that a loop at x contributes two to the degree of x. We write $\delta(G)$ for the *minimum degree of* G and $\Delta(G)$ for the *maximum degree of* G. Given a graph G, denote by $\mathrm{MIS}(G)$ the number of maximal independent sets in G. Given $T \subseteq V(G)$, denote by $\Gamma(T)$ the external neighbourhood of T, i.e. $\Gamma(T) := \{v \in V(G) \setminus T : \exists u \in T, uv \in E(G)\}$. Denote by $G[T]$ the induced subgraph of G on the vertex set T and let $G \setminus T$ denote the induced subgraph of G on the vertex set $V(G) \setminus T$. Denote by $E(T)$ the set of edges in G spanned by T and by $E(T, V(G) \setminus T)$ the set of edges in G with exactly one vertex in T.

4.2 Overview of the proof and preliminary results

4.2.1 Proof overview

We prove Theorem 4.1.1 in Section 4.3. A key tool in the proof is the following container lemma of Green [49] for sum-free sets. The first container-type result in the area (for counting sum-free subsets of $\mathbb{Z}_p$) was given by Green and Ruzsa [51].

Lemma 4.2.1 (Proposition 6 in [49]). *There exists a family $\mathcal{F}$ of subsets of $[n]$ with the following properties.*

(i) Every member of $\mathcal{F}$ has at most $o(n^2)$ Schur triples.

(ii) If $S \subseteq [n]$ is sum-free, then S is contained in some member of $\mathcal{F}$.

(iii) $|\mathcal{F}| = 2^{o(n)}$.

(iv) Every member of $\mathcal{F}$ has size at most $(1/2 + o(1))n$.

We refer to the elements of $\mathcal{F}$ from Lemma 5.2.1 as *containers*. In [49], condition (iv) was not stated explicitly. However, it follows immediately from (i) by, for example, applying Theorem 4.2.2 and Lemma 4.2.3 below. Lemma 5.2.1 can also be derived from a general theorem of Balogh, Morris and Samotij [17], and independently Saxton and Thomason [81] with better bounds in (i) and (iii).

Note that conditions (ii) and (iii) in Lemma 5.2.1 imply that, to prove Theorem 4.1.1, it suffices to show that every member of $\mathcal{F}$ contains at most $2^{n/4+o(n)}$ maximal sum-free subsets of $[n]$. For this purpose, we need to get a handle on the structure of the containers; this is made precise in Lemma 5.2.2 below. The following theorem of Deshouillers, Freiman, Sós and Temkin [30] provides a structural characterisation of the sum-free sets in $[n]$.

Theorem 4.2.2. *Every sum-free set S in $[n]$ satisfies at least one of the following conditions:*

(i) $|S| \leq 2n/5 + 1$;

(ii) S consists of odd numbers;

(iii) $|S| \leq \min(S)$.

We also need the following removal lemma of Green [48] for sum-free sets. (A simpler proof of Lemma 4.2.3 was later given by Král', Serra and Vena [64].)

Lemma 4.2.3 (Corollary 1.6 in [48]). *Suppose that $A \subseteq [n]$ is a set containing $o(n^2)$ Schur triples. Then, there exist B and C such that $A = B \cup C$ where B is sum-free and $|C| = o(n)$.*

Together, Theorem 4.2.2 and Lemma 4.2.3 yield the following structural result on containers of size close to $n/2$.

Lemma 4.2.4. *If $A \subseteq [n]$ has $o(n^2)$ Schur triples and $|A| = (\frac{1}{2} - \gamma)n$ with $\gamma = \gamma(n) \leq 1/11$, then one of the following conditions holds.*

(a) All but $o(n)$ elements of A are contained in the interval $[(1/2 - \gamma)n, n]$.

(b) Almost all elements of A are odd, i.e. $|A \setminus O| = o(n)$.

Proof. Applying Lemma 4.2.3 to A, we have $A = B \cup C$ with B sum-free and $|C| = o(n)$. The conclusion then follows from applying Theorem 4.2.2 to B. Indeed, alternative (i) is impossible, since $|B| \geq (1 - o(1))|A| > 2n/5 + 1$. If alternative (ii) occurs, then we have $|A \setminus O| \leq |C| = o(n)$. If alternative (iii) occurs, then $\min(B) \geq |B| \geq (1/2 - \gamma - o(1))n$. So all but except $o(n)$ elements of A are contained in $[(1/2 - \gamma)n, n]$. □

We remark that Lemma 5.2.2 was already essentially proven in [49] (without applying Lemma 4.2.3). Note that γ could be negative in Lemma 5.2.2. The upper bound $1/11$ on γ here can be relaxed to any constant smaller than $1/10$ (but not to a constant bigger than $1/10$). Roughly speaking, Lemma 5.2.2 implies that every container $A \in \mathcal{F}$ is such that (a) most elements of A lie in $[n/2, n]$, (b) most elements of A are odd or (c) $|A|$ is significantly smaller than $n/2$. Thus, the proof of Theorem 4.1.1 splits into three cases depending on the structure of our container. In each case, we give an upper bound on the number of maximal sum-free sets in a container by counting the number of maximal independent sets in various auxiliary graphs. (Similar techniques were used in [90], and in the graph setting in [19].) In the following subsection we collect together a number of results that are useful for this.

4.2.2 Maximal independent sets in graphs

Given a graph G, denote by MIS(G) the number of maximal independent sets in G. Moon and Moser [68] showed that for any graph G, $\text{MIS}(G) \leq 3^{|G|/3}$. We will need a looped version of this statement. Since any vertex with a loop cannot be in an independent set, the following statement is an immediate consequence of Moon and Moser's result.

Proposition 4.2.5. *Let G be a graph possibly with loops. Then*

$$\text{MIS}(G) \leq 3^{|G|/3}.$$

When a graph is triangle-free, the bound in Proposition 4.2.5 can be improved significantly. A result of Hujter and Tuza [55] states that for any triangle-free graph G,

$$\text{MIS}(G) \leq 2^{|G|/2}. \tag{4.1}$$

The following lemma is a slight modification of this result for graphs with 'few' triangles.

Lemma 4.2.6. *Let G be a graph possibly with loops. If there exists a set T such that $G \setminus T$ is triangle-free, then*

$$\mathrm{MIS}(G) \leq 2^{|G|/2+|T|/2}.$$

Proof. Every maximal independent set in G can be obtained in the following two steps:

(1) Choose an independent set $S \subseteq T$.

(2) Extend S in $V(G) \setminus T$, i.e. choose a set $R \subseteq V(G) \setminus T$ such that $R \cup S$ is a maximal independent set in G.

Note that although every maximal independent set in G can be obtained in this way, it is not necessarily the case that given an arbitrary independent set $S \subseteq T$, there exists a set $R \subseteq V(G) \setminus T$ such that $R \cup S$ is a maximal independent set in G. Notice that if $R \cup S$ is maximal, R is also a maximal independent set in $G \setminus \{T \cup \Gamma(S)\}$. The number of choices for S in (1) is at most $2^{|T|}$. Since $G \setminus \{T \cup \Gamma(S)\}$ is triangle-free, by the Hujter–Tuza bound, the number of extensions in (2) is at most $2^{(|G|-|T|)/2}$. Thus, we have $\mathrm{MIS}(G) \leq 2^{|T|} \cdot 2^{(|G|-|T|)/2} = 2^{|G|/2+|T|/2}$. □

The following lemma gives an improvement on Proposition 4.2.5 for graphs that are 'not too sparse and almost regular'. The proof uses an elegant and simple idea of Sapozhenko [80]; see [56] for a closely-related result.

Lemma 4.2.7. *Let $k \geq 1$ and let G be a graph on n vertices possibly with loops. Suppose that $\Delta(G) \leq k\delta(G)$ where $\delta(G) \geq f(n)$ for some real valued function f with $f(n) \to \infty$ as $n \to \infty$. Then*

$$\mathrm{MIS}(G) \leq 3^{\left(\frac{k}{k+1}\right)\frac{n}{3}+o(n)}.$$

Proof. Fix a maximal independent set I in G and set $b := \delta(G)^{1/2}$. We will repeat the following process as many times as possible. Let $V_1 := V(G)$. At the i-th step, for $i \geq 1$, choose $v_i \in V_i \cap I$ such that $\deg_{G[V_i]}(v_i) \geq b$ and set $V_{i+1} := V_i \setminus (\{v_i\} \cup \Gamma(v_i))$. This process is repeated $j \leq n/b$ times. Let $U := V_{j+1}$ be the resulting set. Define $Z := \{v \in U : \deg_{G[U]}(v) < b\}$. Notice that $\deg_{G[U]}(v) < b$ for all $v \in I \cap U$, hence $I \cap U \subseteq Z$. We have

$$\delta(G) \cdot |Z| \leq \sum_{v \in Z} \deg(v) = 2|E(Z)| + |E(Z, V \setminus Z)| \leq b|Z| + \Delta(G) \cdot (n - |Z|).$$

Hence,

$$|Z| \leq \frac{\Delta(G) \cdot n}{\delta(G) + \Delta(G) - b} \leq \frac{k}{k+1}n + \frac{2n}{b}. \tag{4.2}$$

By construction of U, no vertex in $I \setminus U$ has a neighbour in U. So as $Z \subseteq U$, no vertex in Z is adjacent to $I \setminus U$. Together with the fact that I is maximal, this implies that $I \cap U$ is a maximal independent set in $G[Z]$. By the above process, every maximal independent set I in G is determined by a set $I \setminus U$ of at most n/b vertices and a maximal independent set in $G[Z]$. Note that $n/b = o(n)$. Thus, Proposition 4.2.5 and (4.2) imply that

$$\mathrm{MIS}(G) \leq \sum_{0 \leq i \leq n/b} \binom{n}{i} 3^{\left(\frac{k}{k+1}\right)\frac{n}{3} + \frac{2n}{3b}} \leq 3^{\left(\frac{k}{k+1}\right)\frac{n}{3} + o(n)}. \tag{4.3}$$

□

Note that one could relax the minimum degree condition in Lemma 5.3.4 to (for example) a large constant, at the expense of a worse upper bound on MIS(G). However, Lemma 5.3.4 in its current form suffices for our applications.

4.3 Proof of Theorem 4.1.1

Let $\mathcal{F}$ be the family of containers obtained from Lemma 5.2.1. Recall that given a set $A \subseteq [n]$, $f_{\max}(A)$ denotes the number of maximal sum-free subsets of $[n]$ that lie in A. Since every sum-free subset of $[n]$ is contained in some member of $\mathcal{F}$ and $|\mathcal{F}| = 2^{o(n)}$, it suffices to show that $f_{\max}(A) \leq 2^{(1/4+o(1))n}$ for every container $A \in \mathcal{F}$.

Lemmas 5.2.1 and 5.2.2 imply that every container $A \in \mathcal{F}$ satisfies at least one of the following conditions:

(a) $|A| \leq (1/2 - 1/11)n \leq 0.45n$,

or one of the following holds for some $-o(1) \leq \gamma = \gamma(n) \leq 1/11$:

(b) $|A| = \left(\frac{1}{2} - \gamma\right) n$ and $|A \cap [(1/2-\gamma)n]| = o(n)$;

(c) $|A| = \left(\frac{1}{2} - \gamma\right) n$ and $|A \setminus O| = o(n)$.

We deal with each of the three cases separately.

For any subsets $B, S \subseteq [n]$, let $L_S[B]$ be the *link graph of S on B* defined as follows. The vertex set of $L_S[B]$ is B. The edge set of $L_S[B]$ consists of the following two types of edges:

(i) Two vertices x and y are adjacent if there exists an element $z \in S$ such that $\{x, y, z\}$ forms a Schur triple;

(ii) There is a loop at a vertex x if $\{x, x, z\}$ forms a Schur triple for some $z \in S$ or if $\{x, z, z'\}$ forms a Schur triple for some $z, z' \in S$.

The following simple result will be applied in all three cases of our proof.

Lemma 4.3.1. *Suppose that B, S are both sum-free subsets of $[n]$. If $I \subseteq B$ is such that $S \cup I$ is a maximal sum-free subset of $[n]$, then I is a maximal independent set in $G := L_S[B]$.*

Proof. First notice that I is an independent set in G, since otherwise $S \cup I$ is not sum-free. Suppose to the contrary that there exists a vertex $v \notin I$ such that $I' := I \cup \{v\}$ is still an independent set in G. Then since $I' \subseteq B$ is sum-free, the definition of G implies that $S \cup I'$ is a sum-free set in $[n]$ containing $S \cup I$, a contradiction to the maximality of $S \cup I$. □

4.3.1 Small containers

The following lemma deals with containers of 'small' size.

Lemma 4.3.2. *If $A \in \mathcal{F}$ has size at most $0.45n$, then $f_{\max}(A) = o(2^{n/4})$.*

Proof. Lemma 5.2.1 (i) implies that we can apply Lemma 4.2.3 to A to obtain that $A = B \cup C$ where B is sum-free and $|C| = o(n)$. Notice crucially that every maximal sum-free subset of $[n]$ in A can be built in the following two steps:

(1) Choose a sum-free set S in C;

(2) Extend S in B to a maximal one.

(As in Lemma 4.2.6, note that it is not necessarily the case that given an arbitrary sum-free set $S \subseteq C$, there exists a set $R \subseteq B$ such that $R \cup S$ is a maximal sum-free set in $[n]$.)

The number of choices for S is at most $2^{|C|} = 2^{o(n)}$. For a fixed S, denote by $N(S, B)$ the number of extensions of S in B in Step (2). It suffices to show that for any given sum-free set $S \subseteq C$, $N(S, B) \leq 2^{0.249n}$. Let $G := L_S[B]$ be the link graph of S on B. Since $|A| \leq 0.45n$ and S and B are sum-free, Lemma 4.3.1 and Proposition 4.2.5 imply that

$$N(S, B) \leq \mathrm{MIS}(G) \leq 3^{|B|/3} \leq 3^{|A|/3} \leq 3^{0.45n/3} \ll 2^{0.249n}.$$

□

4.3.2 Large containers

We now turn our attention to containers of relatively large size.

Lemma 4.3.3. *Let $-o(1) \leq \gamma = \gamma(n) \leq 1/11$. If $A \subseteq [n]$ has $o(n^2)$ Schur triples, $|A| = \left(\frac{1}{2} - \gamma\right) n$ and $|A \cap [(1/2 - \gamma)n]| = o(n)$, then*

$$f_{\max}(A) \leq 2^{(1/4+o(1))n}.$$

Proof. Let $A \in \mathcal{F}$ be as in the statement of the lemma. Let $A_1 := A \cap [\lfloor n/2 \rfloor]$ and $A_2 := A \setminus A_1$. Since $|A \cap [(1/2 - \gamma)n]| = o(n)$, we have that $|A_1| \leq (\gamma + o(1))n$. Every maximal sum-free subset of $[n]$ in A can be built from choosing a sum-free set $S \subseteq A_1$ and extending S in A_2. The number of choices for S is at most $2^{|A_1|}$.

Let $G := L_S[A_2]$ be the link graph of S on vertex set A_2. Since S and A_2 are sum-free, Lemma 4.3.1 implies that $N(S, A_2) \leq \mathrm{MIS}(G)$. Notice that G is triangle-free. Indeed, suppose to the contrary that $z > y > x > n/2$ form a triangle in G. Then there exists $a, b, c \in S$ such that $z - y = a, y - x = b$ and $z - x = c$, which implies $a + b = c$ with $a, b, c \in S$. This is a contradiction to S being sum-free. Thus by (5.1) we have $N(S, A_2) \leq \mathrm{MIS}(G) \leq 2^{|A_2|/2}$. Then we have

$$f_{\max}(A) \leq 2^{|A_1|+|A_2|/2} = 2^{|A_1|+((1/2-\gamma)n-|A_1|)/2} = 2^{n/4+(|A_1|-\gamma n)/2} \leq 2^{n/4+o(n)},$$

where the last inequality follows since $|A_1| \leq (\gamma + o(1))n$. □

Lemma 4.3.4. *If $A \in \mathcal{F}$ such that $|A \setminus O| = o(n)$, then*

$$f_{\max}(A) \leq 2^{(1/4+o(1))n}.$$

Proof. Let $A \in \mathcal{F}$ be as in the statement of the lemma. Notice that if $S \subseteq T \subseteq [n]$ then $f_{\max}(S) \leq f_{\max}(T)$. Using this fact, we may assume that $A = O \cup C$ with $C \subseteq E$ and $|C| = o(n)$. Similarly as before, every maximal sum-free subset of $[n]$ in A can be built from choosing a sum-free set $S \subseteq C$ (at most $2^{|C|} = 2^{o(n)}$ choices) and extending S in O to a maximal one. Fix an arbitrary sum-free set S in C and let $G := L_S[O]$ be the link graph of S on vertex set O. Since O is sum-free, by Lemma 4.3.1 we have that $N(S, O) \leq \mathrm{MIS}(G)$. It suffices to show that $\mathrm{MIS}(G) \leq 2^{n/4+o(n)}$. We will achieve this in two cases depending on the size of S.

Case 1: $|S| \geq n^{1/4}$.

In this case, we will show that G is 'not too sparse and almost regular'. Then we apply Lemma 5.3.4.

We first show that $\delta(G) \geq |S|/2$ and $\Delta(G) \leq 2|S| + 2$, thus $\Delta(G) \leq 6\delta(G)$. Let x be any vertex in O. If $s \in S$ such that $s < \max\{x, n - x\}$ then at least one of $x - s$ and $x + s$ is adjacent to x in G. If $s \in S$ such that $s \geq \max\{x, n - x\}$ then $s - x$ is adjacent to x in G. By considering all $s \in S$ this implies that $\deg_G(x) \geq |S|/2$ (we divide by 2 here as an edge xy may arise from two different elements of S). For the upper bound consider $x \in O$. If $xy \in E(G)$ then $y = x + s$, $x - s$ or $s - x$ for some $s \in S$ and only two of these terms are positive. Further, there may be a loop at x in G (contributing 2 to the degree of x in G).

Thus, $\deg_G(x) \le 2|S| + 2$, as desired.

Since $\delta(G) \ge |S|/2 \ge n^{1/4}/2$ we can apply Lemma 5.3.4 to G with $k = 6$. Hence,

$$\mathrm{MIS}(G) \le 3^{\left(\frac{6}{7}\right)\frac{n/2}{3}+o(n)} \ll 2^{0.24n+o(n)} = o(2^{n/4}).$$

Case 2: $|S| \le n^{1/4}$.

In this case, it suffices to show that G has very few, $o(n)$, triangles, since then by applying Lemma 4.2.6 with T being the vertex set of all triangles in G, we have $|T| = o(n)$ and then $\mathrm{MIS}(G) \le 2^{n/4+o(n)}$. Recall that for each edge xy in G, at least one of the evens $x + y$ and $|x - y|$ is in S. We call xy a BLUE edge if $|x - y|$ is in S and a RED edge if $|x - y| \notin S$ and $x + y \in S$.

Claim 4.3.5. Each triangle in G contains either 0 or 2 BLUE edges.

Proof. Let xyz be a triangle in G with $x < y < z$. Suppose that xyz has only one BLUE edge xz. Then $s_1 := z - x, s_2 := x + y$ and $s_3 := y + z$ are elements of S and $s_1 + s_2 = s_3$, a contradiction to S being sum-free. All other cases, including when all the edges are BLUE, are similar, we omit the proof here. □

Consider an arbitrary triple $\{s_1, s_2, s_3\}$ in S (where s_1, s_2 and s_3 are not necessarily distinct). We say that $\{s_1, s_2, s_3\}$ *forces a triangle* $\mathcal{T}$ *in* G if the vertex set $\{x, y, z\}$ of $\mathcal{T}$ is such that s_1, x, y; s_2, y, z and; s_3, x, z form Schur triples. Note that by definition of G, every triangle in G is forced by some triple in S.

Fix an arbitrary triple $\{s_1, s_2, s_3\}$ in S. We will show that $\{s_1, s_2, s_3\}$ forces at most 24 triangles in G. This then implies that G has at most $24|S|^3 = o(n)$ triangles as desired.

By Claim 4.3.5, a triangle xyz with $x < y < z$ can only be one of the following four types: (1) all edges are RED; (2) xy is the only RED edge; (3) yz is the only RED edge; (4) xz is the only RED edge.

It suffices to show that $\{s_1, s_2, s_3\}$ can force at most 6 triangles of each type. We show it only for Type (1), the other types are similar. Suppose that xyz is a Type (1) triangle forced by $\{s_1, s_2, s_3\}$. Set $M := \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$. Then $\mathbf{u} = (x, y, z)^T$ is a solution to $M \cdot \mathbf{u} = \mathbf{s}$ for some $\mathbf{s}$ whose entries are precisely the elements of $\{s_1, s_2, s_3\}$.

Since $\det(M) = 2 \ne 0$, if a solution $\mathbf{u}$ exists to $M \cdot \mathbf{u} = \mathbf{s}$, it should be unique. The number of choices for $\mathbf{s}$, for fixed $\{s_1, s_2, s_3\}$, is $3! = 6$. Thus in total there are at most 6 triangles of Type (1) forced by $\{s_1, s_2, s_3\}$.

This completes the proof of Lemma 5.4.17. □

Chapter 5

Number of maximal sum-free sets, part II

[1]Cameron and Erdős [26] asked whether the number of *maximal* sum-free sets in $\{1, \dots, n\}$ is much smaller than the number of sum-free sets. In the same paper they gave a lower bound of $2^{\lfloor n/4 \rfloor}$ for the number of maximal sum-free sets. In this chapter, we prove the following: For each $1 \le i \le 4$, there is a constant C_i such that, given any $n \equiv i \mod 4$, $\{1, \dots, n\}$ contains $(C_i + o(1))2^{n/4}$ maximal sum-free sets. Our proof makes use of container and removal lemmas of Green [49, 48], a structural result of Deshouillers, Freiman, Sós and Temkin [30] and a recent bound on the number of subsets of integers with small sumset by Green and Morris [50]. We also discuss related results and open problems on the number of maximal sum-free subsets of abelian groups.

5.1 Introduction

A triple x, y, z is a *Schur triple* if $x + y = z$ (note x, y and z may not necessarily be distinct). A set S is *sum-free* if S does not contain a Schur triple. Let $[n] := \{1, \dots, n\}$. We say that $S \subseteq [n]$ is a *maximal sum-free subset of* $[n]$ if it is sum-free and it is not properly contained in another sum-free subset of $[n]$. Let $f(n)$ denote the number of sum-free subsets of $[n]$ and $f_{\max}(n)$ denote the number of maximal sum-free subsets of $[n]$. The study of sum-free sets of integers has a rich history. Clearly, any set of odd integers and any subset of $\{\lfloor n/2 \rfloor + 1, \dots, n\}$ is a sum-free set, hence $f(n) \ge 2^{n/2}$. Cameron and Erdős [25] conjectured that $f(n) = O(2^{n/2})$. This conjecture was proven independently by Green [49] and Sapozhenko [79]. In fact, they showed that there are constants C_1 and C_2 such that $f(n) = (C_i + o(1))2^{n/2}$ for all $n \equiv i \mod 2$.

In a second paper, Cameron and Erdős [26] showed that $f_{\max}(n) \ge 2^{\lfloor n/4 \rfloor}$. Noting that all the sum-free subsets of $[n]$ described above lie in just two maximal sum-free sets, they asked whether $f_{\max}(n) = o(f(n))$ or even $f_{\max}(n) \le f(n)/2^{\varepsilon n}$ for some constant $\varepsilon > 0$. Łuczak and Schoen [66] answered this question in the affirmative, showing that $f_{\max}(n) \le 2^{n/2 - 2^{-28} n}$ for

[1]For the convenience of the readers and to make this chapter self-contained, we repeat some background from Chapter 4.

sufficiently large n. Later, Wolfovitz [90] proved that $f_{\max}(n) \leq 2^{3n/8+o(n)}$. More recently, the authors [14] proved that the lower bound is essentially tight, proving that $f_{\max}(n) = 2^{(1/4+o(1))n}$. In this chapter we give the following exact solution to the problem.

Theorem 5.1.1. *For each $1 \leq i \leq 4$, there is a constant C_i such that, given any $n \equiv i \mod 4$, $[n]$ contains $(C_i + o(1))2^{n/4}$ maximal sum-free sets.*

The proof of Theorem 5.1.1 is given in Section 5.4, with the main work arising in Section 5.4.1. The proof draws on a number of ideas from [14]. In particular, as in [14] we make use of 'container' and 'removal' lemmas of Green [49, 48] as well as a result of Deshouillers, Freiman, Sós and Temkin [30] on the structure of sum-free sets. In order to avoid over-counting the number of maximal sum-free subsets of $[n]$, our present proof also develops a number of new ideas, thereby making the argument substantially more involved. We use a bound on the number of subsets of integers with small sumset by Green and Morris [50] as well as several new bounds on the number of maximal independent sets in various graphs. Further, the proof provides information about the typical structure of the maximal sum-free subsets of $[n]$. Indeed, we show that almost all of the maximal sum-free subsets of $[n]$ look like one of two particular extremal constructions (see Section 5.2.3 for more details).

In Section 5.2 we give an overview of the proof and highlight the new ideas that we develop. We state some useful results in Section 5.3 and prove Theorem 5.1.1 in Section 5.4. In Section 5.5 we give some results and open problems on the number of maximal sum-free subsets of abelian groups.

5.2 Background and an overview of the proof of Theorem 5.1.1

5.2.1 Independence and container theorems

An exciting recent development in combinatorics and related areas has been the emergence of 'independence' as a unifying concept. To be more precise, let V be a set and $\mathcal{E}$ a collection of subsets of V. We say that a subset I of V is an *independent set* if I does not contain any element of $\mathcal{E}$ as a subset. For example, if $V := [n]$ and $\mathcal{E}$ is the collection of all Schur triples in $[n]$ then an independent set I is simply a sum-free set. It is often helpful to think of $(V, \mathcal{E})$ as a hypergraph with vertex set V and edge set $\mathcal{E}$; thus an independent set I corresponds to an independent set in the hypergraph.

So-called 'container results' have emerged as a powerful tool for attacking many problems that concern counting independent sets. Roughly speaking, container results state that the independent sets of a given hypergraph H lie only in a 'small' number of subsets of the vertex set of H (referred to as *containers*), where each of these containers is an 'almost independent set'. Balogh, Morris and Samotij [17] and independently Saxton and Thomason [81], proved general container theorems for hypergraphs whose edge distribution satisfies certain boundedness conditions.

In the proof of Theorem 5.1.1 we will apply the following container theorem of Green [49].

Lemma 5.2.1 (Proposition 6 in [49])**.** *There exists a family $\mathcal{F}$ of subsets of $[n]$ with the following properties.*

(i) Every member of $\mathcal{F}$ has at most $o(n^2)$ Schur triples.

(ii) If $S \subseteq [n]$ is sum-free, then S is contained in some member of $\mathcal{F}$.

(iii) $|\mathcal{F}| = 2^{o(n)}$.

(iv) Every member of $\mathcal{F}$ has size at most $(1/2 + o(1))n$.

We refer to the sets in $\mathcal{F}$ as *containers*.

In [14] we used Lemma 5.2.1 to prove that $f_{\max}(n) = 2^{(1+o(1))n/4}$. Indeed, we showed that every $F \in \mathcal{F}$ contains at most $2^{(1+o(1))n/4}$ maximal sum-free subsets of $[n]$ which by (ii) and (iii) yields the desired result. To obtain an exact bound on $f_{\max}(n)$ it is not sufficient to give a tight general bound on the number of maximal sum-free subsets of $[n]$ that lie in a container $F \in \mathcal{F}$. Indeed, such an $F \in \mathcal{F}$ could contain $O(2^{n/4})$ maximal sum-free subsets of $[n]$, and thus together with (iii) this still gives an error term in the exponent. In general, since containers may overlap, applications of container results may lead to 'over-counting'.

We therefore need to count the number of maximal sum-free subsets of $[n]$ in a more refined way. To explain our method, we first need to describe the constructions which imply that $f_{\max}(n) \geq 2^{\lfloor n/4 \rfloor}$.

5.2.2 Lower bound constructions

The following construction of Cameron and Erdős [26] implies that $f_{\max}(n) \geq 2^{\lfloor n/4 \rfloor}$. Let $n \in \mathbb{N}$ and let $m = n$ or $m = n - 1$, whichever is even. Let S consist of m together with precisely one number from each pair $\{x, m - x\}$ for odd $x < m/2$. Then S is sum-free. Moreover, although S may not be maximal, no further odd numbers less than m can be added, so distinct S lie in distinct maximal sum-free subsets of $[n]$.

The following construction from [14] also yields the same lower bound on $f_{\max}(n)$. Suppose that $4|n$ and set $I_1 := \{n/2 + 1, \dots, 3n/4\}$ and $I_2 := \{3n/4 + 1, \dots, n\}$. First choose the element $n/4$ and a set $S' \subseteq I_2$. Then for every $x \in I_2 \setminus S'$, choose $x - n/4 \in I_1$. The resulting set S is sum-free but may not be maximal. However, no further element in I_2 can be added, thus distinct S lie in distinct maximal sum-free sets in $[n]$. There are $2^{|I_2|} = 2^{n/4}$ ways to choose S.

5.2.3 Counting maximal sum-free sets

The following result provides structural information about the containers $F \in \mathcal{F}$. Lemma 5.2.2 is implicitly stated in [14] and was essentially proven in [49]. It is an immediate consequence of a result of Deshouillers, Freiman, Sós and Temkin [30] on the structure of sum-free sets and a removal lemma of Green [48]. Here O denotes the set of odd numbers in $[n]$.

Lemma 5.2.2. *If $F \subseteq [n]$ has $o(n^2)$ Schur triples then either*
(a) $|F| \leq 0.47n$;
or one of the following holds for some $-o(1) \leq \gamma = \gamma(n) \leq 0.03$:
(b) $|F| = \left(\frac{1}{2} - \gamma\right) n$ and $F = A \cup B$ where $|A| = o(n)$ and $B \subseteq [(1/2-\gamma)n, n]$ is sum-free;
(c) $|F| = \left(\frac{1}{2} - \gamma\right) n$ and $F = A \cup B$ where $|A| = o(n)$ and $B \subseteq O$.

The crucial idea in the proof of Theorem 5.1.1 is that we show 'most' of the maximal sum-free subsets of $[n]$ 'look like' the examples given in Section 5.2.2: We first show that containers of type (a) house only a small (at most $2^{0.249n}$) number of maximal sum-free subsets of $[n]$ (see Lemma 5.4.3). For type (b) containers we split the argument into two parts. More precisely, we count the number of maximal sum-free subsets S of $[n]$ with the property that (i) the smallest element of S is $n/4 \pm o(n)$ and (ii) the second smallest element of S is at least $n/2 - o(n)$. (For this we use a direct argument rather than counting such sets within the containers.) We then show that the number of maximal sum-free subsets of $[n]$ that lie in type (b) containers but that fail to satisfy one of (i) and (ii) is small ($o(2^{n/4})$). We use a similar idea for type (c) containers. Indeed, we show directly that the number of maximal sum-free subsets of $[n]$ that contain *at most* one even number is $O(2^{n/4})$. We then show that the number of maximal sum-free subsets of $[n]$ that lie in type (c) containers and which contain two or more even numbers is small ($o(2^{n/4})$).

In each of our cases, we give an upper bound on the number of maximal sum-free sets in a container by counting the number of maximal independent sets in various auxiliary graphs. (Similar techniques were used in [90, 14], and in the graph setting in [19].) In Section 5.3.3 we collect together a number of results that are useful for this.

5.3 Notation and preliminaries

5.3.1 Notation

For a set $F \subseteq [n]$, denote by $\mathrm{MSF}(F)$ the set of all maximal sum-free subsets of $[n]$ that are contained in F and let $f_{\max}(F) := |\mathrm{MSF}(F)|$. Also, denote by $\min(F)$ and $\max(F)$ the minimum and the maximum element of F respectively. Let $\min_2(F)$ denote the second smallest element of F. Denote by E the set of all even and by O the set of all odd numbers in $[n]$. Given sets A, B, we let $A + B := \{a + b : a \in A, b \in B\}$. We say a real valued function $f(n)$ is exponentially smaller than another real valued function $g(n)$ if there exists a constant $\varepsilon > 0$ such that $f(n) \leq g(n)/2^{\varepsilon n}$ for n sufficiently large. We use log to denote the logarithm function of base 2.

Throughout, all graphs considered are simple unless stated otherwise. We say that G is a *graph possibly with loops* if G can be obtained from a simple graph by adding at most one loop at each vertex. We write $e(G)$ for the number of edges in G. Given a vertex x in G, we write $\deg_G(x)$ for the *degree* of x in G. Note that a loop at x contributes two to the degree of x. We write $\delta(G)$ for the *minimum degree* and $\Delta(G)$ for the *maximum degree* of G. Denote by $G[T]$ the induced subgraph of G on the vertex set T and $G \setminus T$ the induced subgraph of G on the vertex set $V(G) \setminus T$. Given $x \in V(G)$, we write $N_G(x)$ for the *neighourhood of x in G*. Given $S \subseteq V(G)$, we write $N_G(S)$ for the set of vertices $y \in V(G)$ such that $xy \in E(G)$ for some $x \in S$.

We write C_m for the cycle, and P_m for the path on m vertices. Given graphs G and H we write $G \Box H$ for the *cartesian product graph*. So $G \Box H$ has vertex set $V(G) \times V(H)$ and (x, y) and (x', y') are adjacent in $G \Box H$ if (i) $x = x'$ and y and y' are adjacent in H or (ii) $y = y'$ and x and x' are adjacent in G.

Throughout the chapter we omit floors and ceilings where the argument is unaffected. We write $0 < \alpha \ll \beta \ll \gamma$ to mean that we can choose the constants α, β, γ from right to left. More precisely, there are increasing functions f and g such that, given γ, whenever we choose some $\beta \leq f(\gamma)$ and $\alpha \leq g(\beta)$, all calculations needed in our proof are valid. Hierarchies of other lengths are defined in the obvious way.

5.3.2 The number of sets with small sumset

We need the following lemma of Green and Morris [50], which bounds the number of sets with small sumset.

Lemma 5.3.1. *Fix $\delta > 0$ and $R > 0$. Then the following hold for all integers $s \geq s_0(\delta, R)$.*

For any $D \in \mathbb{N}$ there are at most

$$2^{\delta s}\binom{\frac{1}{2}Rs}{s}D^{\lfloor R+\delta \rfloor}$$

sets $S \subseteq [D]$ with $|S| = s$ and $|S+S| \le R|S|$.

5.3.3 Maximal independent sets in graphs

In this section we collect together results on the number of maximal independent sets in a graph. Let $\mathrm{MIS}(G)$ denote the number of maximal independent sets in a graph G.

Moon and Moser [68] showed that for any simple graph G, $\mathrm{MIS}(G) \le 3^{|G|/3}$. When a graph is triangle-free, this bound can be improved significantly: A result of Hujter and Tuza [55] states that for any triangle-free graph G,

$$\mathrm{MIS}(G) \le 2^{|G|/2}. \tag{5.1}$$

The next result implies that the bound given in (5.1) can be further lowered if G is additionally not too sparse.

Lemma 5.3.2. *Let $n, D \in \mathbb{N}$ and $k \in \mathbb{R}$. Suppose that G is a triangle-free graph on n vertices with $\Delta(G) \le D$ and $e(G) \ge n/2 + k$. Then*

$$\mathrm{MIS}(G) \le 2^{n/2 - k/(100D^2)}.$$

The following result for 'almost triangle-free' graphs follows from Lemma 5.3.2.

Corollary 5.3.3. *Let $n, D \in \mathbb{N}$ and $k \in \mathbb{R}$. Suppose that G is a graph and T is a set such that $G' := G \setminus T$ is triangle-free. Suppose that $\Delta(G) \le D$, $|G'| = n$ and $e(G') \ge n/2 + k$. Then*

$$\mathrm{MIS}(G) \le 2^{n/2 - k/(100D^2) + 101|T|/100}.$$

We defer the proofs of Lemma 5.3.2 and Corollary 5.3.3 to the end of this subsection.

The following result gives an improvement on the Moon–Moser bound for graphs that are not too sparse, almost regular and of large minimum degree. (The result is proven as equation (3) in [14].)

Lemma 5.3.4 ([14])**.** *Let $k \ge 1$ and let G be a graph on n vertices possibly with loops.*

Suppose that $\Delta(G) \leq k\delta(G)$ and set $b := \sqrt{\delta(G)}$. Then

$$\mathrm{MIS}(G) \leq \sum_{0 \leq i \leq n/b} \binom{n}{i} 3^{\left(\frac{k}{k+1}\right)\frac{n}{3} + \frac{2n}{3b}}.$$

Fact 5.3.5. Suppose that G' is a (simple) graph. If G is a graph obtained from G' by adding loops at some vertices $x \in V(G')$ then

$$\mathrm{MIS}(G) \leq \mathrm{MIS}(G').$$

The following lemma from [12] gives an improvement on (5.1) when G additionally contains many vertex disjoint P_3s. Its proof is similar to that of Lemma 5.3.2.

Lemma 5.3.6 ([12]). *Let G be an n-vertex triangle-free graph, possibly with loops. If G contains k vertex-disjoint P_3s, then*

$$\mathrm{MIS}(G) \leq 2^{\frac{n}{2} - \frac{k}{25}}.$$

Here we give the proofs of Lemma 5.3.2 and Corollary 5.3.3. The following simple facts will be used in the proof of Lemma 5.3.2.

Fact 5.3.7. Suppose that G is a graph. For any maximal independent set I in G that contains x, $I \setminus \{x\}$ is a maximal independent set in $G \setminus (N_G(x) \cup \{x\})$.

Given $x \in V(G)$, let $\mathrm{MIS}_G(x)$ denote the number of maximal independent sets in G that contain x.

Fact 5.3.8. Suppose that G is a graph. Given any $x \in V(G)$,

$$\mathrm{MIS}(G) \leq \mathrm{MIS}_G(x) + \sum_{v \in N_G(x)} \mathrm{MIS}_G(v).$$

Notice that Fact 5.3.8 is not true in general if G is a graph with loops.

Lemma 5.3.9 (Füredi [45]). *For $m \geq 6$, $\mathrm{MIS}(C_m) = \mathrm{MIS}(C_{m-2}) + \mathrm{MIS}(C_{m-3})$.*

Lemma 5.3.9 implies the following simple result.

Lemma 5.3.10. *For all $m \geq 4$, $\mathrm{MIS}(C_m) < 2^{0.49m}$.*

Proof. It is easy to check that the lemma holds for $m = 4, 5, 6$. For $m \geq 7$, by induction, Lemma 5.3.9 implies that

$$\mathrm{MIS}(C_m) = \mathrm{MIS}(C_{m-2}) + \mathrm{MIS}(C_{m-3}) < 2^{0.49m}(2^{-0.98} + 2^{-1.47}) < 2^{0.49m}.$$

□

Corollary 5.3.11. *If G is the vertex-disjoint union of cycles of length at least 4 then* $\mathrm{MIS}(G) < 2^{0.49|G|}$.

We now combine the previous results to prove Lemma 5.3.2.

Proof of Lemma 5.3.2. We proceed by induction on n. The case when $n \leq 4$ is an easy calculation. We split the argument into several cases.

Case 1: There is a vertex $x \in V(G)$ of degree 0.

By induction $G' := G \setminus \{x\}$ is such that $\mathrm{MIS}(G') \leq 2^{(n-1)/2-k/(100D^2)}$ and clearly $\mathrm{MIS}(G) = \mathrm{MIS}(G')$.

Case 2: There is a vertex $x \in V(G)$ of degree 1.

First suppose that x is adjacent to a vertex y of degree 1. Then consider $G' := G \setminus \{x, y\}$. Note that $\mathrm{MIS}(G) = 2 \cdot \mathrm{MIS}(G')$. Further, $|G'| = n-2$, $e(G') \geq (n-2)/2+k$ and $\Delta(G') \leq D$. Thus, by induction we have that

$$\mathrm{MIS}(G) = 2 \cdot \mathrm{MIS}(G') \leq 2 \times 2^{(n-2)/2-k/(100D^2)} = 2^{n/2-k/(100D^2)},$$

as desired.

Otherwise x is adjacent to a vertex y of degree $d \geq 2$. Consider $G' := G \setminus \{x, y\}$. So $|G'| = n - 2$, $e(G') \geq (n - 2)/2 + k - d + 1$ and $\Delta(G') \leq D$. Therefore by induction and Fact 5.3.7,

$$\mathrm{MIS}_G(x) \leq \mathrm{MIS}(G') \leq 2^{(n-2)/2-(k-d+1)/(100D^2)} \leq 2^{n/2-k/(100D^2)}(2^{-1+d/(100D^2)}). \tag{5.2}$$

Consider $G'' := G \setminus (N_G(y) \cup \{y\})$. So $|G''| = n - d - 1$, $e(G'') \geq n/2 + k - (d - 1)D - 1 \geq (n - d - 1)/2 + (k - (d - 1)D)$ and $\Delta(G'') \leq D$. Thus, by induction and Fact 5.3.7,

$$\begin{aligned}\mathrm{MIS}_G(y) \leq \mathrm{MIS}(G'') &\leq 2^{(n-d-1)/2-(k-(d-1)D)/(100D^2)} \\ &= 2^{n/2-k/(100D^2)}(2^{-(d+1)/2+(d-1)/100D}).\end{aligned} \tag{5.3}$$

Now as $2 \le d \le D$ we have that

$$2^{-1+d/(100D^2)} + 2^{-(d+1)/2+(d-1)/100D} \le 2^{-1+1/100} + 2^{-3/2+1/100} < 1.$$

So (5.2) and (5.3) together with Fact 5.3.8 imply that

$$\mathrm{MIS}(G) \le \mathrm{MIS}_G(x) + \mathrm{MIS}_G(y) < 2^{n/2-k/(100D^2)},$$

as desired.

Case 3: $\delta(G) \ge 4$.
Let $v \in V(G)$ be the vertex of smallest degree in G and write $\deg_G(v) = i - 1 \ge 4$. Given any $w \in N_G(v) \cup \{v\}$ let $G' := G \setminus (N_G(w) \cup \{w\})$. So $|G'| = n - \deg_G(w) - 1$, $e(G') \ge n/2 + (k - \deg_G(w)D) \ge |G'|/2 + (k - \deg_G(w)D)$ and $\Delta(G') \le D$. Hence by induction and Fact 5.3.7

$$\mathrm{MIS}_G(w) \le \mathrm{MIS}(G') \le 2^{(n-\deg_G(w)-1)/2-(k-\deg_G(w)D)/100D^2)} \le 2^{(n-i)/2-(k-iD)/(100D^2)}.$$

Thus by Fact 5.3.8 we have that

$$\mathrm{MIS}(G) \le i \times 2^{(n-i)/2-(k-iD)/(100D^2)} \le (i2^{-i/2+i/100})2^{n/2-k/(100D^2)} < 2^{n/2-k/(100D^2)},$$

as desired. (Here we used that for $i \ge 5$, $i2^{-i/2+i/100} < 1$.)

Case 4: $\delta(G) = 2$ and there exist $v, w \in V(G)$ such that $\deg_G(v) = 2$, $\deg_G(w) \ge 3$ and $vw \in E(G)$.
By arguing as before (using induction and Facts 5.3.7 and 5.3.8) we have that

$$\begin{aligned}\mathrm{MIS}(G) &\le \mathrm{MIS}_G(v) + \sum_{u \in N_G(v)} \mathrm{MIS}_G(u) \le 2 \times 2^{(n-3)/2-(k-2D)/(100D^2)} + 2^{(n-4)/2-(k-3D)/(100D^2)} \\ &< 2^{n/2-k/(100D^2)},\end{aligned}$$

as desired. (Here we have used that $2 \cdot 2^{-3/2+1/50} + 2^{-2+3/100} < 1$.)

Cases 1–4 imply that we may now assume that G consists precisely of 2-regular components and components of minimum degree at least 3.

Case 5: There exist $v, w \in V(G)$ such that $\deg_G(v) = 3$, $\deg_G(w) \ge 4$ and $vw \in E(G)$.

By arguing similarly to before (using induction and Facts 5.3.7 and 5.3.8) we have that

$$\begin{aligned}\mathrm{MIS}(G) &\le \mathrm{MIS}_G(v) + \sum_{u \in N_G(v)} \mathrm{MIS}_G(u) \le 3 \times 2^{(n-4)/2-(k-3D)/(100D^2)} + 2^{(n-5)/2-(k-4D)/(100D^2)} \\ &< 2^{n/2-k/(100D^2)},\end{aligned}$$

as desired. (Here we have used that $3 \cdot 2^{-2+3/100} + 2^{-5/2+1/25} < 1$.)

We may now assume that G consists only of 2- and 3-regular components and components of minimum degree at least 4. However, if there is a component of minimum degree at least 4 then by arguing precisely as in Case 3, we obtain that $\mathrm{MIS}(G) \le 2^{n/2-k/(100D^2)}$. So we may now assume G consists of 2- and 3-regular components only.

Case 6: G contains a 3-regular component.
Here we use the fact that $\mathrm{MIS}(G) \le \mathrm{MIS}(G \setminus \{v\}) + \mathrm{MIS}(G \setminus (N_G(v) \cup \{v\}))$ for any $v \in V(G)$. Indeed, by induction we have

$$\mathrm{MIS}(G) \le 2^{(n-1)/2-(k-5/2)/(100D^2)} + 2^{(n-4)/2-(k-7)/(100D^2)} < 2^{n/2-k/(100D^2)},$$

as desired. (Here we have used that $2^{-1/2+1/40} + 2^{-2+7/100} < 1$.)

Case 7: G is 2-regular.
Since G is triangle-free, Corollary 5.3.11 implies that $\mathrm{MIS}(G) \le 2^{0.49n} \le 2^{n/2-k/(100D^2)}$, as desired. □

Finally, we show that Corollary 5.3.3 follows from Lemma 5.3.2.

Proof of Corollary 5.3.3. Every maximal independent set in G can be obtained in the following two steps:

(1) Choose an independent set $S \subseteq T$.

(2) Extend S in $V(G) \setminus T = V(G')$, i.e. choose a set $R \subseteq V(G')$ such that $R \cup S$ is a maximal independent set in G.

Note that although every maximal independent set in G can be obtained in this way, it is not necessarily the case that given an arbitrary independent set $S \subseteq T$, there exists a set $R \subseteq V(G')$ such that $R \cup S$ is a maximal independent set in G. Notice that if $R \cup S$ is maximal, R is also a maximal independent set in $G'' := G \setminus (T \cup N_G(S))$. The number of choices for S in (1) is at most $2^{|T|}$. Note that G'' is triangle-free, $\Delta(G'') \le D$ and $e(G'') \ge e(G') - |T|D^2 \ge |G''|/2 + (k - |T|D^2)$. Thus, Lemma 5.3.2 implies that the number of extensions in (2) is at most $2^{n/2-(k-|T|D^2)/(100D^2)}$. Therefore, we have $\mathrm{MIS}(G) \le 2^{|T|} \cdot 2^{n/2-(k-|T|D^2)/(100D^2)}$, as desired. □

5.4 Proof of Theorem 5.1.1

Let $1 \leq i \leq 4$ and $0 < \eta < 1$. To prove Theorem 5.1.1, we must show that there is a constant C_i (dependent only on i) such that if n is sufficiently large and $n \equiv i \mod 4$ then

$$(C_i - \eta)2^{n/4} \leq f_{\max}(n) \leq (C_i + \eta)2^{n/4}. \tag{5.4}$$

Given $\eta > 0$ and sufficiently large n with $n \equiv i \mod 4$, define constants $\alpha, \delta, \varepsilon > 0$ so that

$$0 < 1/n \ll \alpha \ll \delta \ll \varepsilon \ll \eta < 1. \tag{5.5}$$

Let $\mathcal{F}$ be the family of containers obtained from Lemma 5.2.1. Since n is sufficiently large, Lemma 5.2.2 implies that $|\mathcal{F}| \leq 2^{\alpha n}$ and for every $F \in \mathcal{F}$ either

(a) $|F| \leq 0.47n$;

or one of the following holds for some $-\alpha \leq \gamma = \gamma(n) \leq 0.03$:

(b) $|F| = \left(\frac{1}{2} - \gamma\right) n$ and $F = A \cup B$ where $|A| \leq \alpha n$ and $B \subseteq [(1/2 - \gamma)n, n]$ is sum-free;

(c) $|F| = \left(\frac{1}{2} - \gamma\right) n$ and $F = A \cup B$ where $|A| \leq \alpha n$ and $B \subseteq O$.

Throughout the rest of the chapter we refer to such containers as type (a), type (b) and type (c), respectively.

For any subsets $B, S \subseteq [n]$, let $L_S[B]$ be the *link graph of S on B* defined as follows. The vertex set of $L_S[B]$ is B. The edge set of $L_S[B]$ consists of the following two types of edges:
(i) Two vertices x and y are adjacent if there exists an element $z \in S$ such that $\{x, y, z\}$ forms a Schur triple;
(ii) There is a loop at a vertex x if $\{x, x, z\}$ forms a Schur triple for some $z \in S$ or if $\{x, z, z'\}$ forms a Schur triple for some $z, z' \in S$.

The following simple lemma from [14] will be applied in many cases throughout the proof.

Lemma 5.4.1 ([14])**.** *Suppose that B and S are both sum-free subsets of $[n]$. If $I \subseteq B$ is such that $S \cup I$ is a maximal sum-free subset of $[n]$, then I is a maximal independent set in $G := L_S[B]$.*

The next lemma will allow us to apply (5.1) to certain link graphs.

Lemma 5.4.2. *Suppose that $B, S \subseteq [n]$ such that S is sum-free and* $\max(S) < \min(B)$. *Then $G := L_S[B]$ is triangle-free.*

Proof. Suppose to the contrary that $z > y > x > \max(S)$ form a triangle in G. Then there exists $a, b, c \in S$ such that $z - y = a, y - x = b$ and $z - x = c$, which implies $a + b = c$ with

$a, b, c \in S$. This is a contradiction to S being sum-free. $\square$

In the proof we will use the simple fact that if $S \subseteq T \subseteq [n]$ then

$$f_{\max}(S) \leq f_{\max}(T). \tag{5.6}$$

The following lemma is a slightly stronger form of Lemma 3.2 from [14], which deals with containers of 'small' size. The proof is exactly the same as in [14].

Lemma 5.4.3. *If $F \in \mathcal{F}$ has size at most $0.47n$, then $f_{\max}(F) \leq 2^{0.249n}$.*

Thus, to show that (5.4) holds it suffices to show that there is a constant C_i such that in total, type (b) and (c) containers house $(C_i \pm \eta/2)2^{n/4}$ maximal sum-free subsets of $[n]$. In Section 5.4.1 we deal with containers of type (b) and in Section 5.4.2 we deal with containers of type (c).

5.4.1 Type (b) containers

The following lemma allows us to restrict our attention to type (b) containers that have at most εn elements from $[n/2]$.

Lemma 5.4.4. *Let $F \in \mathcal{F}$ be a container of type (b) so that $|F \cap [n/2]| \geq \varepsilon n$. Then $f_{\max}(F) \leq 2^{(1/4-\delta)n}$.*

Proof. Define $c \geq \varepsilon$ so that $|F \cap [n/2]| = cn$. Since F is of type (b), $F = A \cup B$ where $|A| \leq \alpha n$ and B is sum-free where $\min(B) \geq 0.47n$. Therefore $cn \leq (0.03 + \alpha)n$.

As $|F \cap [n/2]| = cn$, $|B \cap [0.47n, n/2]| \geq (c - \alpha)n$ and so trivially $|(B + B) \cap [0.94n, n]| \geq (2c - 4\alpha)n$. Therefore, since B is sum-free, F is missing at least $(2c - 4\alpha)n$ numbers from $[0.94n, n]$. Partition $F = F_1 \cup F_2$ where $F_1 := F \cap [n/2]$ and $F_2 := F \setminus F_1$. Note that $|F_2| \leq (1/2 - 2c + 4\alpha)n$.

The following observation is a key idea for the proof of this lemma. Every maximal sum-free subset of $[n]$ in F can be built in the following two steps. First, fix an arbitrary sum-free set $S \subseteq F_1$. Next, extend S in F_2 to a maximal one. Since $|F_1| = cn$, there are at most 2^{cn} ways to pick S. By Lemma 5.4.1, the number of choices for the second step is at most the number of maximal independent sets I in $L_S[F_2]$.

Claim 5.4.5. There are at most $2^{(1/4-\varepsilon/20)n}$ maximal sum-free subsets M of $[n]$ in F such that $|M \cap F_1| \leq cn/4$.

Proof. Choose an arbitrary sum-free set $S \subseteq F_1$ such that $|S| \leq cn/4$ (there are at most $cn\binom{cn}{cn/4}/4$ choices for S). By Lemma 5.4.2, $L := L_S[F_2]$ is triangle-free. So $\text{MIS}(L) \leq 2^{|F_2|/2} \leq 2^{(1/4-c+2\alpha)n}$ by (5.1). Thus, the number of maximal sum-free subsets of $[n]$ in F with at most $cn/4$ elements from F_1 is at most

$$\frac{cn}{4}\binom{cn}{\frac{cn}{4}} \cdot 2^{(1/4-c+2\alpha)n} \leq 2^{(1/4-c/10+2\alpha)n} \leq 2^{(1/4-\varepsilon/20)n},$$

where the last inequality follows since $\alpha \ll \varepsilon \leq c$. □

Let $S \subseteq F_1$ be sum-free such that $|S| > cn/4$. Claim 5.4.5 together with our earlier observation implies that to prove the lemma it suffices to show that $\text{MIS}(L_S[F_2]) \leq 2^{(1/4-c-2\delta)n}$.

By Lemma 5.4.2, $L_S[F_2]$ is triangle-free. We may assume that F is missing at most $(2c+4\delta)n$ numbers from $[0.94n, n]$. Indeed, otherwise by (5.1), $\text{MIS}(L_S[F_2]) \leq 2^{(1/4-c-2\delta)n}$, as required.

Claim 5.4.6. We may assume that $(2c-4\alpha)n \leq |[n/2+1, n] \setminus F| \leq (2c+9\delta)n$.

Proof. Since we already know that $(2c-4\alpha)n \leq |[0.94n, n] \setminus F| \leq (2c+4\delta)n$, to prove the claim we only need to prove that F is missing at most $5\delta n$ elements from $[0.5n, 0.94n]$. Suppose to the contrary that F is missing at least $5\delta n$ numbers from $[0.5n, 0.94n]$. Then $|F_2| \leq (1/2-2c+4\alpha-5\delta)n \leq (1/2-2c-4\delta)n$ and so by (5.1), $\text{MIS}(L_S[F_2]) \leq 2^{(1/4-c-2\delta)n}$. □

Claim 5.4.7. Set $m := \min(S)$. Suppose that $m < (1/4-2c)n$ or $m > (1/4+\varepsilon)n$. Then $\text{MIS}(L_S[F_2]) \leq 2^{(1/4-c-2\delta)n}$.

Proof. Suppose that $m > (1/4+\varepsilon)n$. Then in $L := L_S[F_2]$ a vertex $x \in [(3/4-\varepsilon)n, (3/4+\varepsilon)n] =: N$ is either isolated or adjacent only to itself. Thus $\text{MIS}(L) = \text{MIS}(L')$ where $L' := L \setminus N$. Recall that $(2c-4\alpha)n \leq |[0.94n, n] \setminus F|$. Hence, (5.1) implies that, $\text{MIS}(L) \leq 2^{(1/4-c+2\alpha-\varepsilon)n} \leq 2^{(1/4-c-2\delta)n}$.

Now suppose that $m < (1/4-2c)n$. Then $L := L_S[F_2]$ contains at least $100\delta n$ vertex-disjoint copies of P_3. Indeed, consider the set of all P_3s with vertex set $\{n/2+i, n/2+m+i, n/2+2m+i\}$ for all $1 \leq i \leq n/2-2m$. Since $m \leq (1/4-2c)n$, we have at least $n/2-2m \geq 4cn$ such P_3s. By Claim 5.4.6, at most $(2c+9\delta)n$ elements from $[n/2+1, n]$ are not in F. Hence, L contains at least $(2c-9\delta)n \geq 700\delta n$ of these copies of P_3. Note that these copies of P_3 may not be vertex-disjoint, but given one of these copies P of P_3, there are at most 6 copies of P_3 of this type that intersect P in L. So L contains a collection of $100\delta n$ vertex-disjoint copies of P_3. Using Lemma 5.3.6, we have $\text{MIS}(L) \leq 2^{(1/4-c+2\alpha)n-4\delta n} \leq 2^{(1/4-c-2\delta)n}$. □

By Claim 5.4.7 we may now assume that $(1/4-2c)n \leq m \leq (1/4+\varepsilon)n$.

Claim 5.4.8. Set $b := \min_2(S)$. If $b \leq (1/2-4c)n$ then $\mathrm{MIS}(L_S[F_2]) \leq 2^{(1/4-c-2\delta)n}$.

Proof. We claim that $L := L_S[F_2]$ contains at least $100\delta n$ vertex-disjoint copies of P_3. Consider the set of all P_3s with vertex set $\{n/2+i, n/2+b+i, n/2+b-m+i\}$ for all $1 \leq i \leq n/2-b$. Since $b \leq n/2-4cn$, we have at least $n/2-b \geq 4cn$ such P_3s. Note that F might be missing up to $(2c+9\delta)n$ elements from $[n/2+1, n]$. Hence, L contains at least $(2c-9\delta)n \geq 700\delta n$ of these copies of P_3. Note that these copies of P_3 may not be vertex-disjoint, but given one of these copies P of P_3, there are at most 6 copies of P_3 of this type that intersect P in L. So L contains a collection of $100\delta n$ vertex-disjoint copies of P_3. Hence, Lemma 5.3.6 implies that $\mathrm{MIS}(L_S[F_2]) \leq 2^{(1/4-c-2\delta)n}$. □

So now we may assume that $|S| > cn/4$, $(1/4-2c)n \leq m \leq (1/4+\varepsilon)n$ and $b \geq (1/2-4c)n$. Thus, at least $cn/4$ elements from $[(3/4-6c)n, (3/4+\varepsilon)n]$ lie in $S+m$. Every element of $S+m$ is either missing from F_2 or has a loop in $L_S[F_2]$. Recall that F_2 is missing $(2c-4\alpha)n$ elements from $[0.94n, n]$. Thus, altogether at least $2cn-4\alpha n+cn/4 \geq 2cn+4\delta n$ elements from $[n/2+1, n]$ are either missing from F_2 or have a loop in $L_S[F_2]$. Hence, we have,

$$\mathrm{MIS}(L_S[F_2]) \leq 2^{(1/4-c-2\delta)n}.$$

□

Lemma 5.4.9. *Let $F \in \mathcal{F}$ be a container of type (b) so that $|F \cap [n/2]| \leq \varepsilon n$. Let $f^*_{\max}(F)$ denote the number of maximal sum-free subsets M of $[n]$ in F that satisfy at least one of the following properties:*

(i) $\min(M) > (1/4+2\varepsilon)n$ *or* $\min(M) < (1/4-175\varepsilon)n$*;*

(ii) $\min_2(M) \leq (1/2-350\varepsilon)n$*.*

Then $f^*_{\max}(F) \leq 2^{(1/4-\varepsilon)n}$*.*

Proof. Since F is of type (b), $F = A \cup B$ for some A, B where $|A| \leq \alpha n$ and B is sum-free where $\min(B) \geq 0.47n$. Partition $F = F_1 \cup F_2$ where $F_1 := F \cap [n/2]$ and $F_2 := F \setminus F_1$. So $|F_1| \leq \varepsilon n$ by the hypothesis of the lemma. By (5.6) we may assume that $F_2 = [n/2+1, n]$.

Every maximal sum-free subset of $[n]$ in F that satisfies (i) or (ii) can be built in the following two steps. First, fix a sum-free set $S \subseteq F_1$. Next, extend S in F_2 to a maximal one. To give an upper bound on the sets M satisfying (i) we choose $S \subseteq F_1$ where $m := \min(S)$ is such that $m > (1/4+2\varepsilon)n$ or $m < (1/4-175\varepsilon)n$ (there are at most $2^{|F_1|} \leq 2^{\varepsilon n}$ choices for S). Then by arguing similarly to Claim 5.4.7 we have that $\mathrm{MIS}(L_S[F_2]) \leq 2^{(1/4-2\varepsilon)n}$.

To give an upper bound on the sets M satisfying (ii) we choose $S \subseteq F_1$ where $b := \min_2(S)$ is such that $b \leq n/2 - 350\varepsilon n$ (there are at most $2^{|F_1|} \leq 2^{\varepsilon n}$ choices for S). Then by arguing similarly to Claim 5.4.8 we have that $\mathrm{MIS}(L_S[F_2]) \leq 2^{(1/4-2\varepsilon)n}$.

Altogether, this implies that $f^*_{\max}(F) \leq 2^{(1/4-\varepsilon)n}$ as desired. □

Throughout this subsection, given a maximal sum-free set M we write $m := \min(M)$ and $b := \min_2(M)$ and define $S := (M \cap [n/2]) \setminus \{m\}$. Lemmas 5.4.4 and 5.4.9 imply that, to count the number of maximal sum-free subsets of $[n]$ lying in type (b) containers, it now suffices to count the number of maximal sum-free sets M with the following structure:

(α) $m \in [(1/4 - 175\varepsilon)n, (1/4 + 175\varepsilon)n]$.

(β) $b \geq (1/2 - 350\varepsilon)n$.

In particular, the next lemma shows that almost all of the maximal sum-free subsets of $[n]$ that satisfy (α) and (β) lie in type (b) containers only.

Lemma 5.4.10. *There are at most $\varepsilon 2^{n/4}$ maximal sum-free subsets of $[n]$ that satisfy (α) and (β) and that lie in type (a) or (c) containers.*

Proof. By Lemma 5.4.3, at most $2^{0.249n} \leq \varepsilon 2^{n/4}/2$ such maximal sum-free subsets of $[n]$ lie in type (a) containers.

Suppose that M is a maximal sum-free subset of $[n]$ that satisfies (α) and (β) and lies in a type (c) container F. Thus, $F = A \cup B$ where $|A| \leq \alpha n$ and $B \subseteq O$. Define $F' := B \cap [n/2 - 350\varepsilon n, n]$. So, $|F'| \leq (1/4 + 175\varepsilon)n$. By Lemma 5.4.1, $M = I \cup S$ where $\min(S) = m$ for some $m \in [(1/4 - 175\varepsilon)n, (1/4 + 175\varepsilon)n]$, $(S \setminus \{m\}) \subseteq A$ and I is a maximal independent set in $G := L_S[F']$. By the Moon–Moser bound,

$$\mathrm{MIS}(G) \leq 3^{(1/12+60\varepsilon)n} \leq 2^{(1/4-\varepsilon)n}.$$

In total, there are at most $2^{\alpha n}$ choices for F, at most $350\varepsilon n$ choices for m and at most $2^{\alpha n}$ choices for $S \setminus \{m\}$. Thus, there are at most

$$2^{\alpha n} \times 350\varepsilon n \times 2^{\alpha n} \times 2^{n/4-\varepsilon n} \leq \varepsilon 2^{n/4}/2$$

maximal sum-free subsets of $[n]$ that satisfy (α) and (β) and that lie in type (c) containers, as desired. □

For the rest of this subsection, we focus on counting the maximal sum-free sets that satisfy (α) and (β). Fix m, b such that $m \in [(1/4 - 175\varepsilon)n, (1/4 + 175\varepsilon)n]$ and $b \geq (1/2 - 350\varepsilon)n$. Define $t := |m - n/4|$ and $D := n/2 - b$, so $t, D \leq 350\varepsilon n$. (Notice that if $b > n/2$, then

D is negative.) Let $S \subseteq [b, n/2]$ such that $b \in S$, $S \cup \{m\}$ is sum-free and set $s := |S| \leq D$. In the case when $b > n/2$, we define $S := \emptyset$.

Denote by $L := L(n, m, S)$ the link graph of $S \cup \{m\}$ on vertex set $[n/2 + 1, n]$. So L is triangle-free by Lemma 5.4.2. We will need the following two bounds on the number of maximal independent sets in L.

Lemma 5.4.11. *We have the following two bounds on* $\mathrm{MIS}(L)$*.*

(i) $\mathrm{MIS}(L) \leq 2^{n/4 - D/25}$*;*

(ii) Let R be defined so that $|S + S| = Rs$. Then $\mathrm{MIS}(L) \leq 2^{n/4-(R+1)s/2}$.

Proof. If $D \leq 0$ then (i) follows from (5.1). So assume $D > 0$. Notice that there are D vertex-disjoint P_3s in L: $\{n/2 + i, n + i - D, n + i - D - m\}$ for each $1 \leq i \leq D$. (These paths are vertex-disjoint since $D \leq 350\varepsilon n$ and $m \in [(1/4 - 175\varepsilon)n, (1/4 + 175\varepsilon)n]$.) The bound follows immediately from Lemma 5.3.6.

For (ii), notice that in L we have loops at all vertices in $S + S$ and $S + m$ (in total $(R+1)s$ vertices). $\mathrm{MIS}(L) = \mathrm{MIS}(L')$ where L' is the graph obtained from L by deleting all the vertices with loops. The bound then follows from (5.1). □

The following lemma bounds the number of maximal sum-free sets M satisfying (α) and (β) and with b sufficiently bounded away from $n/2$ from above.

Lemma 5.4.12. *There exists a constant $K = K(\varepsilon)$ such that the number of maximal sum-free sets M in $[n]$ that satisfy (α), (β) and $b \leq n/2 - K$ is at most $\varepsilon 2^{n/4}$.*

Proof. Let K be such that $\delta \ll 1/K \ll \varepsilon$. Our first claim implies that there are not too many maximal sum-free subsets of $[n]$ with t or D 'large'.

Claim 5.4.13. There are at most $\varepsilon 2^{n/4}/5$ maximal sum-free sets M which satisfy (α) and (β) and with

(a) $b \leq n/2 - K$;

(b) $t \geq 3D$ or $D \geq 10^9 s$.

Proof. Fix any m, b such that $m \in [(1/4 - 175\varepsilon)n, (1/4 + 175\varepsilon)n]$ and $n/2 - 350\varepsilon n \leq b \leq n/2 - K$. Define t and D as before. Let $S \subseteq [b, n/2]$ such that $b \in S$, $S \cup \{m\}$ is sum-free and set $s := |S| \leq D$. Define the link graph L as before.

Suppose that $t \geq 3D$. If $m = n/4 - t$ then for each i with $D + 1 \leq i \leq 2t - D$ consider the subgraph H_i of L induced by $\{n/2 + i, 3n/4 + i - t, n + i - 2t\}$. Ignoring loops, H_i spans a P_3 component in L and so $\mathrm{MIS}(H_i) \leq 2$. Indeed, since $t, D \leq 350\varepsilon n$ and $\min(S) = b = n/2 - D$, the vertex $3n/4 + i - t$ has no neighbour in L generated by S. Also, since $n/2+i+b = n+i-D > n$ and $n+i-2t-b = n/2+i-2t+D \leq n/2$, neither $n/2+i$ nor $n+i-2t$ has a neighbour generated by S in L. Recall L and thus $L' := L \setminus \cup_{i=D+1}^{2t-D} H_i$ is triangle-free. Thus by (5.1) we have

$$\mathrm{MIS}(L) \leq \mathrm{MIS}(L') \cdot \prod_i \mathrm{MIS}(H_i) \leq 2^{[n/2-3(2t-2D)]/2} \cdot 2^{2t-2D} \leq 2^{n/4-(t-D)} \leq 2^{n/4-2t/3}.$$

Otherwise $m = n/4+t$ and then there are $2t$ isolated vertices $\{3n/4-t+1, \ldots, 3n/4+t\}$ in L. Then by (5.1), $\mathrm{MIS}(L) \leq 2^{n/4-t}$.

Given fixed t, there are 2 choices for m. There are at most $2^{t/3}$ choices for S so that $D \leq t/3$. Further, fixing S determines b and D. Altogether, this implies that the number of maximal sum-free subsets M of $[n]$ that satisfy (α), (β), (a) and $t \geq 3D$ is at most

$$2 \cdot \sum_{t \geq 3D \geq 3K} 2^{t/3} \cdot 2^{n/4-2t/3} \leq 2 \cdot \sum_{t \geq 3K} 2^{n/4-t/3} \leq \frac{\varepsilon}{10} \cdot 2^{n/4}, \tag{5.7}$$

where the last inequality follows since $1/K \ll \varepsilon$ and n is sufficiently large.

Suppose now that $t \leq 3D$ and $D/s \geq 10^9$. For fixed $D \geq K$ there are $3D$ choices for t and so at most $6D \leq 2^{2\log D}$ choices for m. Given fixed D, there are $D = 2^{\log D}$ choices for s. For fixed D, s there are $\binom{D}{s} \leq \left(\frac{eD}{s}\right)^s \leq 2^{s\log(eD/s)}$ choices for S. Note that when $D/s \geq 10^9$, $3\log D + s\log(eD/s) \leq D/50$. Together, with Lemma 5.4.11(i), this implies that the number of maximal sum-free subsets M of $[n]$ that satisfy (α), (β), (a) and with $t \leq 3D$ and $D/s \geq 10^9$ is at most

$$\sum_{D \geq K} 2^{2\log D} \cdot 2^{\log D} \cdot 2^{s\log(eD/s)} \cdot 2^{n/4-D/25} \leq \sum_{D \geq K} 2^{n/4-D/50} \leq \frac{\varepsilon}{10} \cdot 2^{n/4}. \tag{5.8}$$

□

By Claim 5.4.13, to complete the proof of the lemma it suffices to count the number of maximal sum-free subsets M of $[n]$ that satisfy (α), (β) and

(γ_1) $b \leq n/2 - K$;

(γ_2) $s \geq D/10^9 \geq K/10^9$;

(γ_3) $t < 3D$.

Fix any m, b such that $m \in [(1/4-175\varepsilon)n, (1/4+175\varepsilon)n]$ and $n/2-350\varepsilon n \leq b \leq n/2-K$. Let $S \subseteq [b, n/2]$ such that $b \in S$, $S \cup \{m\}$ is sum-free and set $s := |S| \leq D$. Define the link graph L as before.

Choose s and D such that $s \geq D/10^9$. For each fixed s there are at most $10^9 s$ choices for D. For a fixed $s \geq D/10^9$, there are at most $6D \leq 10^{10}s \leq 2^{2\log s}$ choices for m so that

$t < 3D$ and at most $\binom{10^9 s}{s}$ choices for S. So there are at most

$$10^9 s \cdot 2^{2\log s} \cdot \binom{10^9 s}{s} \leq 10^9 s \cdot 2^{2\log s} \cdot 2^{s\log(e\cdot 10^9)} \leq 2^{49s} \tag{5.9}$$

choices for the pair S, m given fixed s. Let R be defined so that $|S+S| = Rs$. We now distinguish two cases depending on the size of $S+S$.

The number of maximal sum-free subsets M in $[n]$ that satisfy (α), (β), (γ_1)–(γ_3) and $R \geq 100$ is at most

$$\sum_{s\geq K/10^9} 2^{49s} \cdot 2^{n/4-50s} \leq \sum_{s\geq K/10^9} 2^{n/4-s} \leq \frac{\varepsilon}{10} \cdot 2^{n/4}. \tag{5.10}$$

(Here we have applied (5.9) and Lemma 5.4.11 (ii).)

Let $s_0(1/9, 100)$ be the constant returned from Lemma 5.3.1. Since we chose K sufficiently large, we have that $s \geq K/10^9 \geq s_0(1/9, 100)$.

Now suppose $R \leq 100$. Then by Lemma 5.3.1 the number of choices for S is at most

$$2^{s/9}\binom{\frac{1}{2}Rs}{s} D^{\lfloor R+1/9 \rfloor} \leq 2^{s/9} \cdot 2^{Rs/2} \cdot 2^{4R\log s} \leq 2^{Rs/2+2s/9}. \tag{5.11}$$

Recall that for a fixed s, the number of choices for m is at most $2^{2\log s}$. Together with Lemma 5.4.11(ii) and (5.11), we have that the number of maximal sum-free subsets M in $[n]$ that satisfy (α), (β), (γ_1)–(γ_3) and $R \leq 100$ is at most

$$\begin{aligned} &\sum_{s\geq K/10^9} 2^{2\log s} \cdot 2^{Rs/2+2s/9} \cdot 2^{n/4-(R+1)s/2} \leq \sum_{s\geq K/10^9} 2^{n/4-s/2+s/3} \\ \leq &\sum_{s\geq K/10^9} 2^{n/4-s/6} \leq \frac{\varepsilon}{10} \cdot 2^{n/4}. \end{aligned} \tag{5.12}$$

Thus by Claim 5.4.13, (5.10) and (5.12), we have that the number of maximal sum-free sets that satisfy (α), (β) and $b \leq n/2 - K$ is at most $\varepsilon \cdot 2^{n/4}$. □

The following lemma bounds the number of maximal sum-free sets when t is large.

Lemma 5.4.14. *There are at most $\varepsilon 2^{n/4}$ maximal sum-free sets in $[n]$ that satisfy (α) and (β) and with $|m - n/4| = t$ and $b = n/2 - D$ such that $D \leq K$ and $t \geq 50K$.*

Proof. Let us first assume that $m = n/4 + t$. If $b \leq n/2$ then let $S \subseteq [b, n/2]$ where $b \in S$. Otherwise let $S = \emptyset$. Then in the link graph $L := L(n, m, S)$, every vertex in $\{3n/4 - t + 1, 3n/4 + t\} =: N$ is either isolated or adjacent only to itself. Since $D \leq K$, the number of choices for S is at most 2^K. Let $L' := L \setminus N$, then by (5.1) the number of maximal sum-free sets in this case is at most

$$\sum_{t \geq 50K} 2^K \cdot \mathrm{MIS}(L') \leq \sum_{t \geq 50K} 2^K \cdot 2^{n/4-t} \leq \varepsilon 2^{n/4}/2.$$

Otherwise, suppose $m = n/4 - t$. If $b \leq n/2$ then let $S \subseteq [b, n/2]$ where $b \in S$. Otherwise let $S = \emptyset$. The link graph $L := L(n, m, S)$ contains $2t$ vertex-disjoint P_3s on the vertex set $\{n/2 + i, 3n/4 - t + i, n - 2t + i\}$ where $1 \leq i \leq 2t$. Then by Lemma 5.3.6, the number of maximal sum-free sets in this case is at most

$$\sum_{t \geq 50K} 2^K \cdot \mathrm{MIS}(L) \leq \sum_{t \geq 50K} 2^K \cdot 2^{n/4-2t/25} \leq \varepsilon 2^{n/4}/2.$$

□

By Lemmas 5.4.12 and 5.4.14, we now need only focus on maximal sum-free sets with

$$t, D \leq 50K, \quad i.e. \quad S \subseteq [n/2 - 50K, n/2] \quad \text{and} \quad m \in [n/4 - 50K, n/4 + 50K], \quad (5.13)$$

where here D may be negative and $S = \emptyset$. Given any m, S satisfying (5.13) so that $2m \notin S$, define $C(n, m, S) := \frac{|\mathrm{MIS}(L(n,m,S))|}{2^{n/4}}$. Notice that not every maximal independent set in $L(n, m, S)$ necessarily gives a maximal sum-free set in $[n]$. This happens exactly when a set I is a maximal independent set in both $L(n, m, S)$ and $L(n, m, S^*)$ for some sum-free $S^* \supset S$ such that $S^* \subseteq [n/2] \setminus \{m, 2m\}$. Let $\mathcal{I}(n, m, S)$ be the set of all maximal independent sets in $L(n, m, S)$ that do not correspond to maximal sum-free sets in $[n]$. For each $I \in \mathcal{I}(n, m, S)$, define $S^*(I)$ to be the largest sum-free set such that $S \subseteq S^*(I) \subseteq [n/2] \setminus \{m, 2m\}$ and I is also a maximal independent set in $L(n, m, S^*(I))$. Further partition $\mathcal{I}(n, m, S) := \mathcal{I}_1(n, m, S) \cup \mathcal{I}_2(n, m, S)$, in which $\mathcal{I}_1(n, m, S)$ consists of all those $I \in \mathcal{I}(n, m, S)$ with $S^*(I) \subseteq [n/2 - 50K, n/2]$. Let $\mathrm{MSF}(n, m, S)$ be the number of maximal sum-free sets M in $[n]$ that satisfy (α) and (β) with $\min(M) = m$ and $(M \cap [n/2]) \setminus \{m\} = S$. For $i = 1, 2$, further define $C_i(n, m, S) := \frac{|\mathcal{I}_i(n,m,S)|}{2^{n/4}}$. Then clearly by the definition we have

$$\mathrm{MSF}(n, m, S) = [C(n, m, S) - C_1(n, m, S) - C_2(n, m, S)]2^{n/4}.$$

Notice that every set $I \in \mathcal{I}_2(n, m, S)$ is a maximal independent set in $L(n, m, S^*(I))$ with $\min(S^*(I)) \leq n/2 - 50K$, it then follows from Lemma 5.4.12 that $\sum_{m,S:\ t,D \leq 50K} C_2(n, m, S) \leq \varepsilon$.

Thus, the number of maximal sum-free sets M in $[n]$ that satisfy (α) and (β) is at least

$$\begin{aligned}\sum_{m,S:\ t,D\le 50K} \mathrm{MSF}(n,m,S) &= \sum_{m,S:\ t,D\le 50K} [C(n,m,S) - C_1(n,m,S) - C_2(n,m,S)]2^{n/4} \\ &\ge \sum_{m,S:\ t,D\le 50K} [C(n,m,S) - C_1(n,m,S)]2^{n/4} - \varepsilon 2^{n/4}.\end{aligned}$$

On the other hand, by Lemmas 5.4.12 and 5.4.14, the number of maximal sum-free sets M in $[n]$ that satisfy (α) and (β) is at most

$$\begin{aligned}\sum_{m,S} \mathrm{MSF}(n,m,S) &= \sum_{m,S:\ t,D\le 50K} \mathrm{MSF}(n,m,S) + \sum_{m,S:\ \max\{t,D\}>50K} \mathrm{MSF}(n,m,S) \\ &\le \sum_{m,S:\ t,D\le 50K} [C(n,m,S) - C_1(n,m,S)]2^{n/4} + 2\varepsilon 2^{n/4}.\end{aligned}$$

By defining $C(n) := \sum_{m,S:\ t,D\le 50K}[C(n,m,S) - C_1(n,m,S)]$, together with Lemmas 5.4.4, 5.4.9 and 5.4.10, we have that the number of maximal sum-free sets of $[n]$ contained in type (b) containers is $(C(n) \pm 4\varepsilon)2^{n/4}$.

We now proceed to prove that for any $n' \equiv n \mod 4$, $C(n') = C(n)$. We need the following lemma, which roughly states that for any "fixed" choice of m and S, the link graphs on $[n/2+1, n]$ and $[n'/2+1, n']$ differ by a component consisting of an induced matching of size $(n'-n)/4$. To be formal, fix $t \in [-50K, 50K]$, $S_0 \subseteq [50K]$ and $\ell \in \mathbb{N}$. Define

$$n' := n + 4\ell, \quad m := n/4 - t, \quad m' := n'/4 - t, \quad S := n/2 - S_0, \quad S' := n'/2 - S_0. \quad (5.14)$$

The proof of the following lemma for the case $m = n/4 + t$ and $m' = n'/4 + t$ is almost identical except only simpler, we omit it here.

Lemma 5.4.15. *Let n', m, m', S, S' be given as in (5.14). Then $L(n', m', S')$ is isomorphic to the disjoint union of $L(n, m, S)$ and a matching of size ℓ.*

Proof. Let $I_1 := [n'/2 + 200K + 1, 3n'/4 - 200K + t]$ and $I_2 := [3n'/4 + 200K + 1 - t, n' - 200K]$. Notice first that the induced subgraph of $L' := L(n', m', S')$ on $I_1 \cup I_2$ is a matching: $\{n'/2+200K+1, 3n'/4+200K+1-t\}, \dots, \{3n'/4-200K+t, n'-200K\}$. Let $\mathcal{M}$ be the first ℓ matching edges in $L'[I_1 \cup I_2]$, i.e. $\{n'/2+200K+1, 3n'/4+200K+1-t\}, \dots, \{n'/2+200K+\ell, 3n'/4+200K+\ell-t\}$. Define $L'' := L' \setminus \mathcal{M}$. It is a straightforward but tedious task to see that L'' is isomorphic to $L := L(n, m, S)$. We give here only the mapping $f : V(L) \to V(L'')$ that defines an isomorphism:

- $[n/2+1, n/2+200K] \quad \to \quad [n'/2+1, n'/2+200K]$;
- $[n/2+200K+1, 3n/4+200K-t] \quad \to \quad [n'/2+200K+\ell+1, 3n'/4+200K-t]$;
- $[3n/4+200K-t+1, n-200K] \quad \to \quad [3n'/4+200K+\ell-t+1, n'-200K]$;
- $[n-200K+1, n] \quad \to \quad [n'-200K+1, n']$.

□

Fix n', m, m', S, S' satisfying (5.13) and (5.14). By the definition of $C(n)$, to show that $C(n) = C(n')$, it suffices to show that $C(n,m,S) = C(n',m',S')$ and $C_1(n,m,S) = C_1(n,m,S)$. Let $\mathcal{M}$ and f be the matching of size ℓ and the mapping from Lemma 5.4.15. As an immediate consequence of Lemma 5.4.15, we have

$$C(n',m',S') = \frac{|\mathrm{MIS}(L(n',m',S'))|}{2^{n'/4}} = \frac{|\mathrm{MIS}(L(n,m,S))| \cdot \mathrm{MIS}(\mathcal{M})}{2^{n/4} \cdot 2^{\ell}} = C(n,m,S).$$

As for $C_1(n,m,S)$, it suffices to show that every $I \in \mathcal{I}_1(n,m,S)$ corresponds to precisely 2^{ℓ} sets in $\mathcal{I}_1(n',m',S')$. Fix an arbitrary $I \in \mathcal{I}_1(n,m,S)$ and recall that $S \subseteq S^*(I) \subseteq [n/2-50K, n/2]$. Let S^{**} be the "counterpart" (as in S' to S in (5.14)) of $S^*(I)$ in $[n']$, i.e. $S^{**} := n'/2-(n/2-S^*(I)) \subseteq [n'/2-50K, n'/2]$. By the definition of $\mathcal{M}$, edges generated by $S', S^{**} \subseteq [n'/2-50K, n'/2]$ on $[n'/2, n']$ are not incident to any vertex in $\mathcal{M}$. Hence by adding any maximal independent set of $\mathcal{M}$ to $f(I)$, we obtain $|\mathrm{MIS}(\mathcal{M})| = 2^{\ell}$ many maximal independent sets I' in $\mathcal{I}_1(n',m',S')$ with $S^*(I') = S^{**}$ as required. We have concluded the following main result of this subsection.

Lemma 5.4.16. *For each $1 \le i \le 4$, there is a constant D_i such that, if $n \equiv i \mod 4$ then the number of maximal sum-free subsets of $[n]$ in type (b) containers is $(D_i \pm 4\varepsilon)2^{n/4}$.*

5.4.2 Type (c) containers

The next result implies that the number of maximal sum-free subsets of $[n]$ that contain at least two even numbers and that lie in type (c) containers is 'small'.

Lemma 5.4.17. *Let $F \in \mathcal{F}$ be a container of type (c). Then F contains at most $2^{(1/4-\varepsilon/2)n}$ maximal sum-free subsets of $[n]$ that contain at least two even numbers.*

Proof. Let $F \in \mathcal{F}$ be as in the statement of the lemma. Let K be a sufficiently large constant so that

$$\sum_{0 \le i \le n/K} \binom{n}{i} 3^{\frac{5n}{36}+\frac{n}{3K}} \le 2^{0.249n}. \tag{5.15}$$

Since $1/n \ll \varepsilon \ll 1$, we have that $\varepsilon \ll 1/K^2$. By (5.6), we may assume that $F = O \cup C$ with $C \subseteq E$ and $|C| \le \alpha n$. Similarly as before, every maximal sum-free subset of $[n]$ in F can be built from choosing a sum-free set $S \subseteq C$ (at most $2^{|C|} \le 2^{\alpha n}$ choices) and extending S in O to a maximal one. Fix an arbitrary sum-free set S in C where $|S| \ge 2$ and let $G := L_S[O]$ be the link graph of S on vertex set O. Since O is sum-free and $\alpha \ll \varepsilon$, Lemma 5.4.1 implies that, to prove the lemma, it suffices to show that $\mathrm{MIS}(G) \le 2^{(1/4-\varepsilon)n}$. We will achieve this in two cases depending on the size of S.

Case 1: $|S| \ge 2K^2$.

In this case, we will show that G is 'not too sparse and almost regular'. Then we apply Lemma 5.3.4.

We first show that $\delta(G) \ge |S|/2$ and $\Delta(G) \le 2|S| + 2$, thus $\Delta(G) \le 5\delta(G)$. Let x be any vertex in O. If $s \in S$ such that $s < \max\{x, n-x\}$ then at least one of $x - s$ and $x + s$ is adjacent to x in G. If $s \in S$ such that $s \ge \max\{x, n-x\}$ then $s - x$ is adjacent to x in G. By considering all $s \in S$ this implies that $\deg_G(x) \ge |S|/2$ (we divide by 2 here as an edge xy may arise from two different elements of S). For the upper bound consider $x \in O$. If $xy \in E(G)$ then $y = x + s$, $x - s$ or $s - x$ for some $s \in S$ and only two of these terms are positive. Further, there may be a loop at x in G (contributing 2 to the degree of x in G). Thus, $\deg_G(x) \le 2|S| + 2$, as desired.

Note that $\delta(G)^{1/2} \ge K$. Thus, applying Lemma 5.3.4 to G with $k = 5$ we obtain that

$$\mathrm{MIS}(G) \le \sum_{0 \le i \le n/K} \binom{n}{i} 3^{\frac{5n}{36}+\frac{n}{3K}} \overset{(5.15)}{\le} 2^{0.249n}.$$

Case 2: $2 \le |S| \le 2K^2$.

As in Case 1 we have that $\Delta(G) \le 2|S| + 2 \le 5K^2$. Additionally, we need to count triangles in G.

Claim 5.4.18. G contains at most $24|S|^3$ triangles.

The claim is shown in the proof of Lemma 3.4 in [14], so we omit the proof here. Let $T \subseteq V(G)$ such that $|T| \le 24|S|^3$ and $G \setminus T$ is triangle-free.

Let G_1 denote the graph obtained from G by removing all loops. Given any $x \in O$

and $s \in S$, one of $x - s, s - x$ is adjacent to x in G. In particular, if $2x \neq s$, then one of $x - s, s - x$ is adjacent to x in G_1. Therefore each $s \in S$ gives arise to at least $(|O| - 1)/2$ edges in G_1. Given distinct $s, s' \in S$, there is at most one pair $x, y \in O$ such that s, x, y and s', x, y are both Schur triples. Thus, since $|S| \geq 2$, this implies that $e(G_1) \geq |O| - 2$. Set $G' := G_1 \setminus T$. Note that $\Delta(G_1) \leq 5K^2$, $|G'| \leq |O|$ and $e(G') \geq |O| - 2 - |T|5K^2 \geq 3|O|/4$. Thus Corollary 5.3.3 implies that $\mathrm{MIS}(G_1) \leq 2^{(1/4-\varepsilon)n}$. Fact 5.3.5 therefore implies that $\mathrm{MIS}(G) \leq 2^{(1/4-\varepsilon)n}$, as desired. □

Note that the argument in Case 2 of Lemma 5.4.17 immediately implies the following result.

Lemma 5.4.19. *Given any distinct $x, x' \in E$,*

$$\mathrm{MIS}(L_{\{x,x'\}}[O]) \leq 2^{(1/4-\varepsilon)n}.$$

Given $n \in \mathbb{N}$, let $f'_{\max}(n)$ denote the number of maximal sum-free subsets of $[n]$ that contain precisely one even number. The next result implies that $f'_{\max}(n)$ is approximately equal to the number of maximal independent sets in the link graphs $L_x[O]$ where $x \in E$.

Lemma 5.4.20.

$$\sum_{x \in E} \mathrm{MIS}(L_x[O]) - 2 \cdot \sum_{x \neq x' \in E} \mathrm{MIS}(L_{\{x,x'\}}[O]) \leq f'_{\max}(n) \leq \sum_{x \in E} \mathrm{MIS}(L_x[O]). \tag{5.16}$$

In particular,

$$\sum_{x \in E} \mathrm{MIS}(L_x[O]) - 2^{(1/4-\varepsilon/2)n} \leq f'_{\max}(n) \leq \sum_{x \in E} \mathrm{MIS}(L_x[O]). \tag{5.17}$$

Proof. Given any maximal sum-free subset M of $[n]$ that contains precisely one even number x, $M \setminus \{x\}$ is a maximal independent set in $L_x[O]$. So the upper bound in (5.16) follows.

Claim 5.4.21. Suppose $x \in E$ and S is a maximal independent set in $L_x[O]$. Let M denote the maximal sum-free subset of $[n]$ that contains $S \cup \{x\}$. Then $M \setminus S \subseteq E$.

Proof. Suppose not. Then there exists $S' \subseteq M$ such that $S \subset S' \subseteq O$. But as M is sum-free, S' is an independent set in $L_x[O]$, a contradiction to the maximality of S. □

Suppose $y \in E$ and S is a maximal independent set in $L_y[O]$. If $S \cup \{y\}$ is not a maximal sum-free subset of $[n]$ then Claim 5.4.21 implies that there exists $y' \in E \setminus \{y\}$ such that

$S \cup \{y, y'\}$ is sum-free. In particular, S is a maximal independent set in $L_{\{y,y'\}}[O]$. In total there are at most

$$2 \cdot \sum_{x \neq x' \in E} \mathrm{MIS}(L_{\{x,x'\}}[O])$$

such pairs S, y. Thus, the lower bound in (5.16) follows.

The lower bound in (5.17) follows since, by Lemma 5.4.19,

$$2 \cdot \sum_{x \neq x' \in E} \mathrm{MIS}(L_{\{x,x'\}}[O]) \leq 2n^2 \cdot 2^{(1/4-\varepsilon)n} \leq 2^{(1/4-\varepsilon/2)n},$$

where the last inequality follows since n is sufficiently large. □

The next result determines $\sum_{x \in E} \mathrm{MIS}(L_x[O])$ asymptotically and thus, together with Lemma 5.4.20 determines, asymptotically, $f'_{\max}(n)$.

Lemma 5.4.22. *Given $1 \leq i \leq 4$, there exists a constant D'_i such that, if $n \equiv i \mod 4$,*

$$(D'_i - \varepsilon)2^{n/4} \leq \sum_{x \in E} \mathrm{MIS}(L_x[O]) \leq (D'_i + \varepsilon)2^{n/4}.$$

Proof. Suppose that $n \equiv 0 \mod 4$. The proofs for the other cases are essentially identical, so we omit them. Let $2n/3 < m \leq n$ be even. Consider $G := L_m[O]$. The edge set of G consists of precisely the following edges:

- An edge between i and $m - i$ for every odd $i < m/2$;
- A loop at $m/2$ if $m/2$ is odd;
- An edge between i and $m + i$ for all odd $i \leq n - m < n/3$.

In particular, since $m > 2n/3$, if $i < m/2$ is odd then in G, $m - i$ is only adjacent to i. Altogether this implies that if $m/2$ is even then G is the disjoint union of:

- $(n - m)/2$ copies of P_3;
- A matching containing $(3m - 2n)/4$ edges.

In this case $\mathrm{MIS}(G) = 2^{(n-m)/2} \times 2^{(3m-2n)/4} = 2^{m/4}$. If $m/2$ is odd then G is the disjoint union of:

- $(n - m)/2$ copies of P_3;
- A single loop;

- A matching containing $(3m-2n-2)/4$ edges.

In this case $\text{MIS}(G) = 2^{(m-2)/4}$.

Thus,

$$\sum_{m\in E\,:\, m>2n/3} \text{MIS}(L_m[O]) \le \sum_{m=4\,:\, m\equiv 0 \bmod 4}^{n} 2^{m/4} + \sum_{m=2\,:\, m\equiv 2 \bmod 4}^{n} 2^{(m-2)/4}$$
$$= \sum_{m=1}^{n/4} 2^m + \sum_{m=0}^{n/4-1} 2^m \le (3+\varepsilon/2)2^{n/4}. \tag{5.18}$$

Further,

$$\sum_{m\in E\,:\, m>2n/3} \text{MIS}(L_m[O]) \ge (3-\varepsilon/2)2^{n/4} - \sum_{m=1}^{2n/3} 2^{m/4} \ge (3-\varepsilon)2^{n/4}. \tag{5.19}$$

Consider $m \in E$ where $m \le 2n/3$ and set $G := L_m[O]$. It is easy to see that G is the disjoint union of paths that contain at least 3 vertices and in the case when $m/2$ is odd, an additional path of length at least 2 which contains a vertex (namely $m/2$) with a loop. Every such graph on $n/2$ vertices contains at least $n/10-1$ vertex-disjoint copies of P_3. Therefore, by Lemma 5.3.6 we have that

$$\sum_{m\in E\,:\, m\le 2n/3} \text{MIS}(L_m[O]) \le n2^{n/4-n/250+1}. \tag{5.20}$$

Overall, we have that

$$(3-\varepsilon)2^{n/4} \overset{(5.19)}{\le} \sum_{x\in E} \text{MIS}(L_x[O]) \overset{(5.18),(5.20)}{\le} (3+\varepsilon/2)2^{n/4} + n2^{n/4-n/250+1} \le (3+\varepsilon)2^{n/4},$$

as desired. □

We showed that the constant D'_4 in Lemma 5.4.22 is equal to 3. By following the argument given in the proof, it is easy to see that $D'_1 = 3\cdot 2^{-1/4}$, $D'_2 = 2^{3/2}$ and $D'_3 = 2^{5/4}$.

The next lemma shows that almost all of the maximal sum-free subsets of $[n]$ that contain precisely one even number lie in type (c) containers only.

Lemma 5.4.23. *There are at most $\varepsilon 2^{n/4}$ maximal sum-free subsets of $[n]$ that contain precisely one even number and that lie in type (a) or (b) containers.*

Proof. By Lemma 5.4.3, at most $2^{0.249n} \le \varepsilon 2^{n/4}/2$ such maximal sum-free subsets of $[n]$ lie in type (a) containers.

Suppose that M is a maximal sum-free subset of $[n]$ that lies in a type (b) container F and only contains one even number. Define $F' := F \cap O$. Since F is of type (b), $|F'| \le (0.53n)/2 + \alpha n \le 0.27n$. By Lemma 5.4.1, $M = I \cup \{m\}$ where m is even and I is a maximal independent set in $G := L_m[F']$. By the Moon–Moser bound,

$$\mathrm{MIS}(G) \le 3^{0.09n} \le 2^{(1/4-\varepsilon)n}.$$

In total, there are at most $2^{\alpha n}$ choices for F and at most $n/2$ choices for m. Thus, there are at most

$$2^{\alpha n} \times \frac{n}{2} \times 2^{n/4-\varepsilon n} \le \varepsilon 2^{n/4}/2$$

maximal sum-free subsets of $[n]$ that that lie in type (b) containers and only contain one even number, as desired. □

Notice that this completes the proof of Theorem 5.1.1. Indeed, for each $1 \le i \le 4$, set $C_i := D_i + D_i'$. Lemmas 5.4.3, 5.4.16, 5.4.17, 5.4.20, 5.4.22 and 5.4.23 together imply that if $n \equiv i \mod 4$, then

$$(C_i - \eta)2^{n/4} \le f_{\max}(n) \le (C_i + \eta)2^{n/4},$$

as desired.

5.5 Maximal sum-free sets in abelian groups

Throughout this section, unless otherwise specified, G will be an abelian group of order n and we denote by $\mu(G)$ the size of the largest sum-free subset of G. Denote by $f(G)$ the number of sum-free subsets of G and by $f_{\max}(G)$ the number of maximal sum-free subsets of G. Given a set $F \subseteq G$, we write $f_{\max}(F)$ for the number of maximal sum-free subsets of G that lie in F.

The study of sum-free sets in abelian groups dates back to the 1960s. Although Diananda and Yap [31] determined $\mu(G)$ for a large class of abelian groups G, it was not until 2005 that Green and Ruzsa [52] determined $\mu(G)$ for all such G. In particular, for every finite abelian group G, $2n/7 \le \mu(G) \le n/2$. Further, Green and Ruzsa [52] determined $f(G)$ up to an error term in the exponent for all G, showing that $f(G) = 2^{(1+o(1))\mu(G)}$.

Given G, what can we say about $f_{\max}(G)$? Is it also the case that $f_{\max}(G)$ is exponentially smaller than $f(G)$? Wolfovitz [90] proved that $f_{\max}(G) \le 2^{0.406n+o(n)}$ for every finite group

G. For even order abelian groups G this answers the second question in the affirmative since $\mu(G) = n/2$ for such groups.

Our next result strengthens the result of Wolfovitz for abelian groups, and implies that indeed $f_{\max}(G)$ is exponentially smaller than $f(G)$ for all finite abelian groups G. Let G be fixed. By a container lemma [52, Proposition 2.1] and a removal lemma [48, Theorem 1.4] for abelian groups, there exists a collection of containers $\mathcal{F}$ such that:

(i) $|\mathcal{F}| = 2^{o(n)}$ and $F \subseteq G$ for all $F \in \mathcal{F}$;

(ii) Given any $F \in \mathcal{F}$, $F = B \cup C$ where B is sum-free with size $|B| \leq \mu(G)$ and $|C| = o(n)$;

(iii) Given any sum-free subset S of G, there is an $F \in \mathcal{F}$ such that $S \subseteq F$.

Given sets $S, T \subseteq G$, we can define the link graph $L_S[T]$ analogously to the integer case. In particular, it is easy to check that an analogue of Lemma 5.4.1 holds for such link graphs.

Let $F \in \mathcal{F}$ be fixed. Every maximal sum-free subset of G contained in F can be chosen by picking a sum-free set S in C (at most $2^{o(n)}$ choices by (ii)), and extending it in B (at most $\text{MIS}(L_S[B]) \leq 3^{|B|/3} \leq 3^{\mu(G)/3}$ choices by Lemma 5.4.1 for abelian groups and the Moon-Moser theorem). Therefore, together this implies the following result.

Proposition 5.5.1. *Let G be an abelian group of order n. Then*

$$f_{\max}(G) \leq 3^{\mu(G)/3+o(n)}. \tag{5.21}$$

We do not know how far from tight the bound in Proposition 5.5.1 is. In particular, it would be interesting to establish whether the following bound holds.

Question 5.5.2. Given an abelian group G of order n, is it true that $f_{\max}(G) \leq 2^{\mu(G)/2+o(n)}$?

For the group $Z_2^k := Z_2 \otimes Z_2 \otimes \cdots \otimes Z_2$, the answer to the above question is affirmative and the upper bound is essentially tight.

Proposition 5.5.3. *The number of maximal sum-free subsets of Z_2^k is $2^{(1+o(1))\mu(Z_2^k)/2}$.*

Proof. Let $n := |Z_2^k|$. It is known that $\mu(Z_2^k) = n/2$. We first give a lower bound $f_{\max}(Z_2^k) \geq 2^{n/4}$. Write $Z_2^k = Z_2 \otimes Z_2 \otimes H$, where $H := Z_2^{k-2}$. Let $x := (0, 1, 0_H)$ and $U := \{1\} \otimes Z_2 \otimes H$. Notice that the link graph $L_x[U]$ is a perfect matching. Indeed, for any vertex $y = (1, a, h) \in U$, all of its possible neighbours in U are $x + y = (1, 1 + a, h)$, $x - y = (1, 1 - a, -h)$ and $y - x = (1, a - 1, h)$ and these elements of Z_2^k are identical. To build a collection of sum-free subsets, we first pick x and then pick exactly one of the endpoints of each edge in $L_x[U]$. Since $|U| = n/2$, we obtain $2^{n/4}$ sum-free subsets S in this way. These sets might not be maximal, but no further elements from U can be added into any of these sets. Hence distinct S lie in distinct maximal sum-free subsets. Therefore we have

$$f_{\max}(Z_2^k) \geq 2^{n/4}.$$

We now proceed with the proof of the upper bound. Let $\mathcal{F}$ be the family of $2^{o(n)}$ containers defined before Proposition 5.5.1. It suffices to show that $f_{\max}(F) \leq 2^{(1/4+o(1))n}$ for every container $F \in \mathcal{F}$. Fix a container $F \in \mathcal{F}$. We have $F = B \cup C$ with B sum-free, $|B| \leq \mu(Z_2^k) = n/2$ and $|C| = o(n)$. Every maximal sum-free subset of Z_2^k in F can be built by choosing a sum-free set S in C and extending S in B to a maximal one. The number of choices for S is at most $2^{|C|} = 2^{o(n)}$. For a fixed S, let $\Gamma := L_S[B]$ be the link graph of S on B. Then Lemma 5.4.1 (for abelian groups) implies that the number of extensions is at most $\mathrm{MIS}(\Gamma)$. Observe that Γ is triangle-free. Indeed, suppose to the contrary that there exists a triangle on vertices $a, b, c \in B \subseteq Z_2^k$. Since for any $x \in Z_2^k$, $x = -x$, we may assume that $a+b = s_1$, $b+c = s_2$ and $a+c = s_3$ for some $s_1, s_2, s_3 \in S$. Furthermore, s_1, s_2, s_3 are distinct elements in S since a, b, c are distinct in B. Then we have $s_1 + s_2 = a + 2b + c = a + c = s_3$, contradicting S being sum-free. Thus by (5.1), we have

$$\mathrm{MIS}(\Gamma) \leq 2^{|B|/2} \leq 2^{n/4}$$

and so

$$f_{\max}(F) \leq 2^{|C|} \cdot 2^{n/4} = 2^{(1/4+o(1))n},$$

as desired. □

The following construction gives a lower bound $f_{\max}(Z_n) \geq 6^{(1/18-o(1))n}$. Let $n = 9k+i$ for some $0 \leq i \leq 8$ and $M := [3k+1, 6k]$. Set $\Gamma := L_{\{k,-2k\}}[M]$. Then $|M|/6 - o(n)$ components of Γ are copies of $K_3 \square K_2$ as there are at most a constant number of components of Γ that are not copies of $K_3 \square K_2$. Observe that $K_3 \square K_2$ contains 6 maximal independent sets. Thus, $\mathrm{MIS}(\Gamma) \geq 6^{(1/18-o(1))n}$, yielding the desired lower bound on $f_{\max}(Z_n)$. It is known that $\mu(Z_p) = (1/3 + o(1))p$, if p is *prime*, so together with (5.21), we obtain the following result.

Proposition 5.5.4. *If p is prime then*

$$1.1^{p-o(p)} \leq 6^{(1/18-o(1))p} \leq f_{\max}(Z_p) \leq 3^{(1/9+o(1))p} \leq 1.13^{p+o(p)}.$$

It would be interesting to close the gap in Proposition 5.5.4.

We end this section with two more constructions that would match the upper bound in

Question 5.5.2 if it is true. For this, we need the following simple fact.

Fact 5.5.5. Suppose G is an abelian group of odd order. Then given a fixed $x \in G$, there is a unique solution in G to the equation $2y = x$.

Notice that Fact 5.5.5 is false for abelian groups of even order.

Proposition 5.5.6. *Suppose that* $3|n$ *where* n *is not divisible by a prime* p *with* $p \equiv 2 \mod 3$*. Then* $f_{\max}(G) \geq 2^{(n-9)/6} = 2^{(\mu(G)-3)/2}$.

Proof. First note that $\mu(G) = n/3$ for such groups (see [52]). Let $H \leq G$ be a subgroup of index 3. Then there are three cosets $0 + H, 1 + H, 2 + H$. Pick some $x \in 2 + H$. Then consider the link graph $\Gamma := L_x[1 + H]$ on $n/3$ vertices. There is a loop at $2x \in V(\Gamma)$. For every $y \in 1 + H$, $x + y \in 0 + H$, $y - x \in 2 + H$ and $x - y \in 1 + H$. So y has only one neighbour $x - y$ in $1 + H$ (unless $y = 2x$, which has a loop). By Fact 5.5.5, there is a unique $y \in 1 + H$ such that $x - y = y$. Overall this implies that Γ consists of the disjoint union of a matching M of size $(n-3)/6$, with a loop at at most one of the vertices in M, together with an additional vertex with a loop. Clearly $\mathrm{MIS}(\Gamma) \geq 2^{(n-9)/6}$ and so $f_{\max}(G) \geq 2^{(n-9)/6}$. □

Proposition 5.5.7. *Suppose that* n *is only divisible by primes* p *such that* $p \equiv 1 \mod 3$*. Suppose further that the exponent of* G *(the largest order of any element of* G*) is* 7*. Then* $f(G) \geq 2^{n/7-1} = 2^{\mu(G)/2-1}$.

Proof. First note that $\mu(G) = 2n/7$ for such groups (see [52]). Let $H \leq G$ be a subgroup of index 7. Then pick some $x \in 1 + H$. Consider the link graph $\Gamma := L_x[(2 + H) \cup (3 + H)]$ on $2n/7$ vertices. There is a loop at $2x \in 2 + H$ in Γ. The remaining edges of Γ form a perfect matching between $2+H$ and $3+H$. Therefore $\mathrm{MIS}(\Gamma) = 2^{n/7-1}$ and so $f_{\max}(G) \geq 2^{n/7-1}$. □

We conclude the section with two conjectures.

Conjecture 5.5.8. *For every abelian group* G *of order* n*,*

$$2^{n/7} \leq f_{\max}(G) \leq 2^{n/4+o(n)},$$

where the bounds, if true, are best possible.

We also suspect that there is an infinite class of finite abelian groups for which the upper bounds in Conjecture 5.5.8 and Question 5.5.2 are far from tight.

Conjecture 5.5.9. *There is a sequence of finite abelian groups* $\{G_i\}$ *of increasing order such that for all* i*,*

$$f_{\max}(G_i) \leq 2^{\mu(G_i)/2.01}.$$

Chapter 6

k-AP-free sets

Addressing a question of Cameron and Erdős [25], we show that, for infinitely many values of n, the number of subsets of $\{1, 2, \ldots, n\}$ that do not contain a k-term arithmetic progression is at most $2^{O(r_k(n))}$, where $r_k(n)$ is the maximum cardinality of a subset of $\{1, 2, \ldots, n\}$ without a k-term arithmetic progression. This bound is optimal up to a constant factor in the exponent. For all values of n, we prove a weaker bound, which is nevertheless sufficient to transfer the current best upper bound on $r_k(n)$ to the sparse random setting. To achieve these bounds, we establish a new supersaturation result, which roughly states that sets of size $\Theta(r_k(n))$ contain superlinearly many k-term arithmetic progressions.

Our proof uses the hypergraph container method, which has proven to be a very powerful tool in extremal combinatorics, and a new supersaturation theorem for arithmetic progressions.

6.1 Introduction

A subset of $[n] := \{1, 2, \ldots, n\}$ is *k-AP-free* if it does not contain a k-term arithmetic progression. Denote by $r_k(n)$ the maximum size of a k-AP-free subset of $[n]$. Cameron and Erdős [25] raised the following question: How many subsets of $[n]$ that does not contain a k-term arithmetic progression are there? In particular, they asked the following question.

Question 6.1.1 (Cameron-Erdős)**.** Is it true that the number of k-AP-free subsets of $[n]$ is

$$2^{(1+o(1))r_k(n)}?$$

Since every subset of a k-AP-free set is also k-AP-free, one can easily obtain $2^{r_k(n)}$ many k-AP-free subsets of $[n]$. In fact, Cameron and Erdős [25] slightly improved this obvious lower bound: writing $R_k(n)$ for the number of k-AP-free subsets of $[n]$, they proved

$$\limsup_{n\to\infty} \frac{R_k(n)}{2^{r_k(n)}} = \infty. \qquad (6.1)$$

The only progress on the upper bound in the last 30 years was improving the bounds on $r_k(n)$, until recently Balogh, Morris and Samotij [17], and independently Saxton and Thomason [81], proved that: for any $\beta > 0$ and integer $k \geq 3$, there exists $C > 0$ such that for $m \geq Cn^{1-1/(k-1)}$, the number of k-AP-free m-sets in $[n]$ is at most $\binom{\beta n}{m}$. This deep counting result implies the sparse random analogue of Szemerédi's theorem [87], however, this bound is far from settling Question 6.1.1.

One of the reasons for the difficulty in finding good upper bounds on $R_k(n)$ is our limited understanding of $r_k(n)$. Indeed, despite much effort, the gap between the current known lower and upper bounds on $r_3(n)$ is still rather large; closing this gap remains one of the most difficult problems in additive number theory. For the lower bound on $r_3(n)$, the celebrated construction of Behrend [20] shows that

$$r_3(n) = \Omega\left(\frac{n}{2^{2\sqrt{2}\sqrt{\log_2 n}} \cdot \log^{1/4} n}\right).$$

This was recently improved by Elkin [32] by a factor of $\sqrt{\log n}$, see also Green and Wolf [53]. Roth [77] gave the first non-trivial upper bound on $r_3(n)$, followed by the improvements of Heath-Brown [54], Szemerédi [86] and Bourgain [24]. The current best bound was provided by a recent breakthrough result of Sanders [78]:

$$r_3(n) = O\left(\frac{n(\log\log n)^5}{\log n}\right). \tag{6.2}$$

For $k \geq 4$, the best known estimates are as follows: there exists $c_k, c'_k > 0$ such that

$$\frac{n}{2^{c_k(\log n)^{1/(k-1)}}} \leq r_k(n) \leq \frac{n}{(\log\log n)^{c'_k}}, \tag{6.3}$$

where the lower bound is due to Rankin [76] and upper bound is by Gowers [46, 47].

Notice that using the lower bound in (6.3), we can get the following trivial upper bound for $R_k(n)$:

$$R_k(n) \leq \sum_{i=0}^{r_k(n)} \binom{n}{i} < 2\binom{n}{r_k(n)} < 2\left(\frac{en}{r_k(n)}\right)^{r_k(n)} = 2^{O(r_k(n)\cdot(\log n)^{\frac{1}{k-1}})}.$$

We show that for infinitely many n, the $(\log n)^{\frac{1}{k-1}}$ term in the exponent is not needed, i.e. our result is optimal up to a constant factor in the exponent.

Theorem 6.1.2. *The number of k-AP-free subsets of $[n]$ is at most $2^{O(r_k(n))}$ for infinitely* *many values of n.*

An immediate corollary of Theorem 6.1.2 is the following.

Corollary 6.1.3. *For every $\varepsilon > 0$, there exists a constant $b > 0$ such that the following holds. Let $A(b) \subseteq \mathbb{Z}$ consist of all integers n such that the number of k-AP-free subsets of $[n]$ is at most $2^{b \cdot r(n)}$. Then*

$$\limsup_{n\to\infty} \frac{|A(b)\cap[n]|}{n} \geq 1-\varepsilon.$$

Enumerating discrete structures with certain local constraints is a central topic in combinatorics. Theorem 6.1.2 is the first such result in which the order of magnitude of the corresponding extremal function is not known. It is also worth mentioning that two other natural conjectures of Erdős are false: it was conjectured that the number of Sidon sets[1] in $[n]$, denoted by $S(n)$, is $2^{(1+o(1))s(n)}$, where $s(n)$ denotes the size of a maximum Sidon set. However, it is known that $2^{1.16s(n)} \leq S(n) = 2^{O(s(n))}$, where the lower bound is by Saxton and Thomason [81] and the upper bound is by Kohayakawa, Lee, Rödl and Samotij [62] (see also [81]). Another conjecture of Erdős states that the number of C_6-free[2] graphs on vertex set $[n]$, denoted by $H(n)$, is $2^{(1+o(1))\mathrm{ex}(n,C_6)}$, where $\mathrm{ex}(n, C_6)$ is the maximum number of edges in a C_6-free graph. However, $2^{1.16\mathrm{ex}(n,C_6)} \leq H(n) = 2^{O(\mathrm{ex}(n,C_6))}$, where the lower bound is by Morris and Saxton [69] and the upper bound is by Kleitman and Wilson [61]. In view of these examples and (6.1), it is not inconceivable that the answer to Question 6.1.1 is "no".

The proof of Theorem 6.1.2 uses the hypergraph container method, developed by Balogh, Morris and Samotij [17], and independently by Saxton and Thomason [81]. In order to apply the hypergraph container method, we need a supersaturation result. Supersaturation problems are reasonably well-understood if the extremal family is of positive density. For example, the maximum size sum-free subset of $[n]$ has size $\lceil n/2 \rceil$, while any set of size $(\frac{1}{2}+\varepsilon)n$ has $\Omega(n^2)$ many triples satisfying $x + y = z$, see [48]. In the context of graphs, the Erdős-Stone theorem gives[3] $\mathrm{ex}(n, G) = (1 - \frac{1}{\chi(G)-1} + o(1))\frac{n^2}{2}$, while any n-vertex graph with $(1 - \frac{1}{\chi(G)-1} + \varepsilon)\frac{n^2}{2}$ many edges contains $\Omega(n^{|V(G)|})$ many copies of G. However, the degenerate case is significantly harder. Indeed, a famous open conjecture in extremal graph theory asks whether an n-vertex graph with $\mathrm{ex}(n, C_4)+1$ edges has at least two copies of C_4. In terms of arithmetic progressions, a weak supersaturation result by Varnavides [89] states that any subset of $[n]$ of size $\Omega(n)$ has $\Omega(n^2)$ many k-APs. However, nothing about sets of

[1] A set $A \subseteq [n]$ is a *Sidon* set if there do not exist distinct $a, b, c, d \in A$ such that $a + b = c + d$.

[2] Denote by C_k the cycle of length k. Given a graph H, a graph G is *H-free* if G does not contain H as a subgraph.

[3] We use standard graph theoretical terminology. Denote by $\mathrm{ex}(n, G)$ the maximum number of edges a G-free graph can have. The *chromatic number* of G, denoted by $\chi(G)$, is the minimum number of colors needed to color the vertices of G such that no two adjacent vertices receive the same color.

sublinear size is known. We need a much stronger supersaturation for sets of size $\Theta(r_k(n))$ for the application of the container method. Our second main result shows that the number of k-APs in any set A of size somewhat larger than $r_k(n)$ is superlinear in $|A|$.

Theorem 6.1.4. *Given $k \geq 3$, there exists a constant $C' = C'(k) > 0$ and an infinite sequence $\{n_i\}_{i=1}^{\infty}$, such that the following holds. For any $n \in \{n_i\}_{i=1}^{\infty}$ and any $A \subseteq [n]$ of size $C'r_k(n)$, the number of k-APs in A is at least*

$$\log^{3k-2} n \cdot \left(\frac{n}{r(n)}\right)^{k-1} \cdot n.$$

For all values of n, we obtain the following weaker estimate.

Theorem 6.1.5. *If $r_k(n) \leq \frac{n}{h(n)}$, where $h(n) \leq \log^c n$ for some $c > 0$, then the number of k-AP-free subsets of $[n]$ is at most $2^{O(n/h(n))}$. Furthermore, for any $\gamma > 0$, there exists $C = C(k, c, \gamma) > 0$ such that for any $m \geq n^{1-\frac{1}{k-1}+\gamma}$, the number of k-AP-free m-subsets of $[n]$ is at most*

$$\binom{Cn/h(n)}{m}.$$

Theorem 6.1.5 improves the counting result of Balogh-Morris-Samotij [17] and Saxton-Thomason [81] with a slightly weaker bound on m. We say that a set $A \subseteq \mathbb{N}$ is (δ, k)-*Szemerédi* if every subset of A of size at least $\delta|A|$ contains a k-AP. Denote by $[n]_p$ the p-random subset of $[n]$, where each element of $[n]$ is chosen with probability p independently of others. As mentioned earlier, the counting result of [17] and [81] implies the following sparse analogue of Szemerédi's theorem, which was only recently proved by a breakthrough transference theorem of Conlon and Gowers [27] and Schacht [82]: For any constant $\delta > 0$ and integer $k \geq 3$, there exists $C > 0$, such that $[n]_p$ is (δ, k)-Szemerédi almost surely for $p \geq Cn^{-\frac{1}{k-1}}$. As an easy corollary of Theorem 6.1.5, we obtain the following sharper version, in which δ could be taken as a function of n. In fact, it transfers current best bounds on $r_k(n)$ of Sanders [78] and Gowers [46, 47] to the random setting. Proving Corollary 6.1.6 from Theorem 6.1.5 is similar as in [17], thus we omit the proof here. We remark that the bound on p is optimal up to the additive error term γ in the exponent.

Corollary 6.1.6. *If $r_k(n) \leq \frac{n}{h(n)}$, where $h(n) \leq (\log n)^c$ for some $c > 0$, then for any $\gamma > 0$, there exists $C = C(k, c, \gamma) > 0$ such that the following holds. If $p_n \geq n^{-\frac{1}{k-1}+\gamma}$ for all sufficiently large n, then*

$$\lim_{n\to\infty} \mathbb{P}\left([n]_{p_n} \text{ is } \left(\frac{C}{h(n)}, k\right)\text{-Szemerédi}\right) = 1.$$

Combining the upper bounds in (6.2) and (6.3) with Corollary 6.1.6, for some $C > 0$, we have that almost surely $[n]_p$ is $\left(\frac{C(\log\log n)^5}{\log n}, 3\right)$-Szemerédi for $p \geq n^{-\frac{1}{2}+o(1)}$; and for $k \geq 4$ that almost surely $[n]_p$ is $\left(\frac{C}{(\log\log n)^{c'_k}}, k\right)$-Szemerédi for $p \geq n^{-\frac{1}{k-1}+o(1)}$.

Another immediate corollary of Theorem 6.1.5, together with bounds in (6.2) and (6.3), is the following.

Corollary 6.1.7. *The number of 3-AP-free subsets of* $[n]$ *is at most* $2^{O\left(n(\log\log n)^5/\log n\right)}$. *For* $k \geq 4$, *the number of* k*-AP-free subsets of* $[n]$ *is at most* $2^{O\left(n/(\log\log n)^{c'_k}\right)}$.

Organization. The rest of the chapter will be organized as follows. In Section 6.2, we introduce the hypergraph container method and some lemmas needed for proving supersaturation. In Section 6.3, we prove our main result, Theorem 6.1.2, and also Corollary 6.1.3, Theorem 6.1.4 and Theorem 6.1.5.

Notation. We write $[a, b]$ for the interval $\{a, a+1, \ldots, b\}$ and $[n] := [1, n]$. Given a set $A \subseteq [n]$, denote by $\Gamma_k(A)$ the number of k-APs in A. We write log for logarithm with base 2. Throughout the chapter we omit floors and ceilings where they are not crucial.

6.2 Preliminaries

In the next subsection, we present the hypergraph container theorem and derive a version tailored for arithmetic progressions. We then prove some supersaturation results needed for the proof of Theorem 6.1.4 in Section 6.2.2.

To see how they work, we give a quick overview of the proof of Theorem 6.1.2. We first apply the hypergraph container theorem (Corollary 6.2.2) to obtain a small collection of containers covering all k-AP-free sets in $[n]$, each of these containers having only few copies of k-APs. Then we apply the supersaturation result (Theorem 6.1.4) to show that every container necessarily has to be small in size ($O(r_k(n))$), from which our main result follows.

6.2.1 The hypergraph container theorem

An r-uniform hypergraph $\mathcal{H} = (V, E)$ consists of a vertex set V and an edge set E, in which every edge is a set of r vertices in V. An *independent* set in $\mathcal{H}$ is a set of vertices inducing no edge in E. The *independence number* $\alpha(\mathcal{H})$ is the maximum cardinality of an independent set in $\mathcal{H}$. Denote by $\chi(\mathcal{H})$ the *chromatic number* of $\mathcal{H}$, i.e., the minimum integer ℓ, such that $V(\mathcal{H})$ can be colored by ℓ colors with no monochromatic edge.

Many classical theorems in combinatorics can be phrased as statements about independent sets in a certain auxiliary hypergraph. For example, the celebrated theorem of Szemerédi [87] states that for $V(\mathcal{H}) = [n]$ and $E(\mathcal{H})$ consisting of all k-term arithmetic progressions in $[n]$, $\alpha(\mathcal{H}) = o(n)$. The cornerstone result of Erdős and Stone [41] in extremal graph theory characterizes the structure of all maximum independent sets in $\mathcal{H}$, where $V(\mathcal{H})$ is the edge set of K_n and $E(\mathcal{H})$ is the edge set of copies of some fixed graph G.

We will use the method of hypergraph containers for the proof of Theorem 6.1.2. This powerful method was recently introduced independently by Balogh, Morris and Samotij [17], and by Saxton and Thomason [81]. Roughly speaking, it says that if a hypergraph $\mathcal{H}$ has a somewhat uniform edge-distribution, then one can find a relatively small collection of sets covering all independent sets in $\mathcal{H}$. Among others, this method provides an alternative proof of a recent breakthrough transference theorem of Conlon and Gowers [27] and Schacht [82] for extremal results in sparse random setting. We refer the readers to [17, 81] for more details and applications, see also [14] for more recent applications of container-type results in the arithmetic setting.

Let $\mathcal{H}$ be an r-uniform hypergraph with average degree d. For every $S \subseteq V(\mathcal{H})$, its co-degree, denoted by $d(S)$, is the number of edges in $\mathcal{H}$ containing S, i.e.,

$$d(S) = |\{e \in E(\mathcal{H}) : S \subseteq e\}|.$$

For every $j \in [r]$, denote by Δ_j the j-th maximum co-degree of $\mathcal{H}$, i.e.,

$$\Delta_j = \max\{d(S) : S \subseteq V(\mathcal{H}), |S| = j\}.$$

For any $\tau \in (0, 1)$, define

$$\Delta(\mathcal{H}, \tau) = 2^{\binom{r}{2}-1} \sum_{j=2}^{r} 2^{-\binom{j-1}{2}} \frac{\Delta_j}{d\tau^{j-1}}.$$

We need the following version of the hypergraph container theorem (Corollary 3.6 in [81]).

Theorem 6.2.1. *Let $\mathcal{H}$ be an r-uniform hypergraph on vertex set $[n]$. Let $0 < \varepsilon, \tau < 1/2$. Suppose that $\tau < 1/(200r!^2 r)$ and $\Delta(\mathcal{H}, \tau) \leq \varepsilon/(12r!)$. Then there exists $c = c(r) \leq 1000r!^3 r$ and a collection of vertex subsets $\mathcal{C}$ such that*

(i) every independent set in $\mathcal{H}$ is a subset of some $A \in \mathcal{C}$;

(ii) for every $A \in \mathcal{C}$, $e(\mathcal{H}[A]) \leq \varepsilon e(\mathcal{H})$;

(iii) $\log |\mathcal{C}| \leq cn\tau \log(1/\varepsilon) \log(1/\tau)$.

Given an integer $k \geq 3$, consider the k-uniform hypergraph $\mathcal{H}_k$ encoding the set of all k-APs in $[n]$: $V(\mathcal{H}_k) = [n]$ and the edge set of $\mathcal{H}_k$ consists of all k-tuples that form a k-AP. It is easy to check that the number of k-APs in $[n]$ is $n^2/(2k) < e(\mathcal{H}_k) < n^2/k$. Note that $\Delta_1 \leq k \cdot \frac{n}{k-1} < 2n$ and

$$d = d(\mathcal{H}_k) \geq \frac{n}{2}, \quad \Delta_k = 1, \quad \Delta_i \leq \Delta_2 \leq \binom{k}{2} < k^2 \text{ for } 2 \leq i \leq k-1. \tag{6.4}$$

Using the k-AP-hypergraph $\mathcal{H}_k$, we obtain the following adaptation of Theorem 6.2.1 to the arithmetic setting.

Corollary 6.2.2. *Fix an arbitrary integer $k \geq 3$ and let $0 < \varepsilon, \tau < 1/2$ be such that*

$$\tau < 1/(200k^{2k}) \quad \text{and} \quad \varepsilon n \tau^{k-1} > k^{3k}. \tag{6.5}$$

Then for sufficiently large n, there exists a collection $\mathcal{C}$ of subsets of $[n]$ such that

(i) every k-AP-free subset of $[n]$ is contained in some $F \in \mathcal{C}$;

(ii) for every $F \in \mathcal{C}$, the number of k-APs in F is at most εn^2;

(iii) $\log |\mathcal{C}| \leq 1000k^{3k} n\tau \log(1/\varepsilon) \log(1/\tau)$.

Proof. Consider the k-AP hypergraph $\mathcal{H}_k$. Fix any $0 < \varepsilon, \tau < \frac{1}{2}$ such that $\tau < \frac{1}{200k^{2k}} < 2^{-3k}$ and $\varepsilon n \tau^{k-1} > k^{3k}$. Define $\alpha_j := 2^{-\binom{j-1}{2}} \cdot \tau^{-(j-1)}$ for $2 \leq j \leq k$. Since $\tau < 2^{-3k}$, we have that for $2 \leq j \leq k-2$,

$$\frac{\alpha_j}{\alpha_{j+1}} = \frac{2^{\binom{j}{2}} \cdot \tau^j}{2^{\binom{j-1}{2}} \cdot \tau^{j-1}} = 2^{j-1}\tau < 2^k \tau < 1 \quad \text{and} \quad \frac{k^3 \alpha_{k-1}}{\alpha_k} = k^3 2^{k-2} \tau < 1. \tag{6.6}$$

Note that for any $k \geq 3$, we have that $\tau < 1/(200k^{2k}) < 1/(200k!^2 k)$. We now bound the function $\Delta(\mathcal{H}_k, \tau)$ from above as follows:

$$\begin{aligned}\Delta(\mathcal{H}_k, \tau) &= 2^{\binom{k}{2}-1} \sum_{j=2}^{k} \alpha_j \frac{\Delta_j}{d} \overset{(6.4)}{\leq} 2^{\binom{k}{2}-1} \left(\sum_{j=2}^{k-1} \alpha_j \frac{k^2}{d} + \frac{\alpha_k}{d} \right) \overset{(6.6)}{\leq} 2^{\binom{k}{2}-1} \left((k-2)\alpha_{k-1} \frac{k^2}{d} + \frac{\alpha_k}{d} \right) \\ &\overset{(6.6)}{\leq} 2^{\binom{k}{2}-1} \cdot \frac{2\alpha_k}{d} = \frac{2^{k-1}}{d\tau^{k-1}} \overset{(6.4)}{\leq} \frac{2^k}{n\tau^{k-1}} \overset{(6.5)}{\leq} \frac{\varepsilon}{12k!}.\end{aligned}$$

We can now apply Theorem 6.2.1 on $\mathcal{H}_k$ to obtain $\mathcal{C}$. Then the conclusions follow from the observation that every independent set in $\mathcal{H}_k$ is a k-AP-free subset of $[n]$. □

6.2.2 Supersaturation

In this subsection, we present the second main ingredient for the proof of Theorem 6.1.2: a supersaturation result, Lemma 6.2.5, which states that many k-APs start to appear in a set once its size is larger than $r_k(n)$.

First notice that for any $A \subseteq [n]$ of size $K \cdot r_k(n)$, the following greedy algorithm gives

$$\Gamma_k(A) \geq (K-1) \cdot r_k(n). \tag{6.7}$$

Set $B := A$. Repeat the following process $(K-1) \cdot r_k(n)$ times: since $|B| > r_k(n)$, there is a k-AP in B; update B by removing an arbitrary element in this k-AP. We can use a random sparsening trick to improve this simple argument.

Lemma 6.2.3. *For every $A \subseteq [n]$ of size $K \cdot r_k(n)$ with $K \geq 2$, we have*

$$\Gamma_k(A) \geq \left(\frac{K}{2}\right)^k \cdot r_k(n).$$

Proof. Let T be a set chosen uniformly at random among all subsets of A of size $2r_k(n)$. Then the expected number of k-APs in T is

$$\mathbb{E}[\Gamma_k(T)] = \frac{\binom{|A|-k}{|T|-k}}{\binom{|A|}{|T|}} \cdot \Gamma_k(A) \leq \left(\frac{|T|}{|A|}\right)^k \cdot \Gamma_k(A) = \frac{\Gamma_k(A)}{(K/2)^k}.$$

Thus, there exists a choice of T such that $\Gamma_k(T) \leq \frac{\Gamma_k(A)}{(K/2)^k}$. On the other hand, from (6.7), $\Gamma_k(T) \geq r_k(n)$, hence $\Gamma_k(A) \geq \left(\frac{K}{2}\right)^k \cdot r_k(n)$ as desired. □

However, the bound given above is still linear in $|A|$, which is not sufficient for our purposes. A superlinear bound is provided in the following lemma, which implies that $\Gamma_k(A) \geq |A| \cdot \mathrm{polylog}(n)$[4] for infinitely many values of n (as in Theorem 6.1.4). We remark that all previously known supersaturation results only apply to sets of size linear in n, see Varnavides [89] and also Croot-Sisask [28]. A key new idea in our proof is that an averaging argument is carried out only over a set of carefully chosen arithmetic progressions with prime common differences. We first state the Prime Number Theorem (see e.g. [84]) that will be used later.

Theorem 6.2.4 (Prime Number Theorem)**.** *Let $\pi(x)$ be the number of primes less than*

[4] We write $\mathrm{polylog}(n)$ for a function that is a polynomial in $\log n$.

equal to x. Then

$$\lim_{x\to\infty}\frac{\pi(x)}{x/\log x}=1.$$

Lemma 6.2.5. *For any $1\le M\le n$ and $A\subseteq[n]$, if $\frac{|A|}{M}$ is sufficiently large and $\frac{|A|}{n}\ge 8K\cdot\frac{r_k(M)}{M}$ with $K\ge 2$, then*

$$\Gamma_k(A)\ge\frac{|A|^2}{M^2}\cdot\frac{K^k\cdot r_k(M)}{2^{k+4}\log^2 n}.$$

Proof. Define $x=|A|/(4M)$, and assume that it is sufficiently large that the Prime Number Theorem [84] holds, i.e., the number of prime numbers less than x is at least $x/\log x$ and at most $2x/\log x$. Let $\mathcal{B}_d$ denote the set of M-term arithmetic progressions with common difference d in $[n]$ and set

$$\mathcal{B}:=\bigcup_{\substack{d\text{ is prime}\\ d\le x}}\mathcal{B}_d,$$

that is, $\mathcal{B}$ consists of all M-APs whose common difference is a prime number not larger than x. We notice first that any k-AP can occur in at most $M\log n$ many members of $\mathcal{B}$. Indeed, fix an arbitrary k-AP, say Q', with common difference d'. Note that every M-AP Q containing Q' can be constructed in two steps:

(i) choose $1\le i\le M$ and set the i-th term of Q to $\min(Q')$;

(ii) choose the common difference d for Q.

There are clearly at most M choices for (i). As for (ii), in order to have $Q'\subseteq Q$, we need $d|d'$. Since $Q\in\mathcal{B}$, d must be a prime divisor of d'. Since the number of prime divisors of d' is at most $\log d'\le\log n$, the number of such choices is at most $\log n$. As a consequence, we have that

$$\Gamma_k(A)\ge\frac{1}{M\log n}\sum_{B\in\mathcal{B}}\Gamma_k(A\cap B). \tag{6.8}$$

Let $\mathcal{G}\subseteq\mathcal{B}$ consists of all $B\in\mathcal{B}$ such that $|A\cap B|\ge K\cdot r_k(M)$. Then by Lemma 6.2.3, we have $\Gamma_k(A\cap B)\ge(K/2)^k\cdot r_k(M)$ for every $B\in\mathcal{G}$. Together with (6.8), this gives that

$$\Gamma_k(A)\ge\frac{1}{M\log n}\sum_{B\in\mathcal{G}}\Gamma_k(A\cap B)\ge|\mathcal{G}|\cdot\frac{K^k\cdot r_k(M)}{2^kM\log n}. \tag{6.9}$$

Our next goal is to give a lower bound on $|\mathcal{G}|$, to achieve this, we will do a double-counting on $\sum_{B\in\mathcal{B}}|A\cap B|$.

For each $d\le x$, define $I_d:=[(M-1)d+1,n-(M-1)d]$. Then every $z\in I_d$ appears in exactly M members of $\mathcal{B}_d$. Since $x = |A|/(4M)$,

$$|A \cap I_d| = |A| - 2(M-1)d \geq |A| - 2Mx \geq \frac{|A|}{2}.$$

As an immediate consequence of the Prime Number Theorem [84], the number of primes less than x, which is the number of choices for d, is at least $x/\log x$ and at most $2x/\log x$ for sufficiently large x. Therefore,

$$\sum_{B \in \mathcal{B}} |A \cap B| = \sum_{\substack{d \text{ is prime} \\ d \leq x}} \sum_{B \in \mathcal{B}_d} |A \cap B| \geq M \sum_{\substack{d \text{ is prime} \\ d \leq x}} |A \cap I_d| \geq M \cdot \frac{x}{\log x} \cdot \frac{|A|}{2}. \quad (6.10)$$

On the other hand, since $|\mathcal{B}_d| < n$, for each d we have $|\mathcal{B}| \leq \frac{2x}{\log x} \cdot n$, hence

$$\sum_{B \in \mathcal{B}} |A \cap B| \leq M|\mathcal{G}| + K \cdot r_k(M) \cdot |\mathcal{B} \setminus \mathcal{G}| \leq M|\mathcal{G}| + K \cdot r_k(M) \cdot \frac{2xn}{\log x}. \quad (6.11)$$

Combining (6.10) and (6.11), we get

$$\begin{aligned} |\mathcal{G}| &\geq \frac{x}{\log x} \cdot \frac{|A|}{2} - K \cdot \frac{r_k(M)}{M} \cdot \frac{2xn}{\log x} = \frac{x}{\log x} \left(\frac{|A|}{2} - 2K \cdot \frac{r_k(M)}{M} \cdot n \right) \\ &\geq \frac{x}{\log n} \cdot \frac{|A|}{4} = \frac{|A|^2}{16M \log n}, \end{aligned}$$

where the last inequality follows from $\frac{|A|}{n} \geq 8K \cdot \frac{r_k(M)}{M}$. Thus, by (6.9), we have

$$\Gamma_k(A) \geq \frac{|A|^2}{16M \log n} \cdot \frac{K^k \cdot r_k(M)}{2^k M \log n} = \frac{|A|^2}{M^2} \cdot \frac{K^k \cdot r_k(M)}{2^{k+4} \log^2 n}.$$

□

6.3 Proof of Theorem 6.1.2

Throughout this section, we fix k a positive integer and write $r(n)$ instead of $r_k(n)$ and define $f(n) = r(n)/n$. We will use the following functions:

$$M(n) = \frac{n}{\log^{3k} n} \left(\frac{r(n)}{n} \right)^{k+2}, \ \varepsilon(n) = \frac{\log^{3k-2} n}{n} \left(\frac{n}{r(n)} \right)^{k-1}, \ \tau(n) = \frac{r(n)}{n} \frac{1}{\log^3 n}. \quad (6.12)$$

We first observe a simple fact about the function $r(n)$. Since the property of having no k-AP is invariant under translation, for any given $m < n$, if we divide $[n]$ into consecutive intervals of length m, then any given k-AP-free subset of $[n]$ contains at most $r(m)$ elements from each interval. Thus,

$$r(n) \leq \left\lceil \frac{n}{m} \right\rceil \cdot r(m). \tag{6.13}$$

Since $\frac{1}{n} \cdot \lceil \frac{n}{m} \rceil < \frac{2}{m}$ for any $m < n$, dividing by n on both sides of (6.13) yields:

Fact 6.3.1. For every $m < n$, $f(n) < 2f(m)$.

We also need that the function $r(n)$ is "smooth".

Lemma 6.3.2. *Given $k \geq 3$, there exists $C := C(k) > 4$ and an infinite sequence $\{n_i\}_{i=1}^{\infty}$, such that*

$$C\frac{r(n_i)}{n_i} \geq \frac{r(M(n_i))}{M(n_i)}$$

for all $i \geq 1$, where $M(n)$ is defined as (6.12).

Proof. Fix $C = C(k) > 4$ a sufficiently large constant. From Behrend's construction, we know that $f(n) > 2^{-5\sqrt{\log n}}$. We need to show that, for infinitely many n, $Cf(n) \geq f(M(n)) = f\left(\frac{n}{\log^{3k} n} f(n)^{k+2}\right)$. Suppose to the contrary, that for all but finitely many n, $f(n) \leq C^{-1} f(M(n))$. Let n_0 be the largest integer such that $f(n) > C^{-1} f(M(n))$.

Define a decreasing function $g(x) = 2^{-(5k+11)\sqrt{\log x}}$ for $x \geq 1$. Note that for sufficiently large n, since $f(n) > 2^{-5\sqrt{\log n}}$,

$$M(n) = \frac{n}{\log^{3k} n} f(n)^{k+2} > \frac{n}{\log^{3k} n} \cdot 2^{-5(k+2)\sqrt{\log n}} > n \cdot 2^{-(5k+11)\sqrt{\log n}} = n \cdot g(n).$$

Then by Fact 6.3.1, we have $f(M(n)) < 2f(n \cdot g(n))$. Therefore, by the definition of n_0, we have that for any $n > n_0$,

$$f(n) \leq C^{-1} f(M(n)) < \left(\frac{C}{2}\right)^{-1} f(n \cdot g(n)). \tag{6.14}$$

Fix an integer $n > n_0^2$ and set integer $t = \lfloor \frac{1}{2} \frac{\sqrt{\log n}}{5k+11} \rfloor$. We will show by induction that for every $1 \leq j \leq t$,

$$f(n) < \left(\frac{C}{4}\right)^{-j} f(n \cdot g(n)^j). \tag{6.15}$$

The base case is given by (6.14). Suppose (6.15) holds for some $1 \le j < t$. Define $n' := n \cdot g(n)^j$. Then

$$n' > n \cdot g(n)^t = n \cdot 2^{-(5k+11)\sqrt{\log n} \cdot \lfloor \frac{1}{2} \frac{\sqrt{\log n}}{5k+11} \rfloor} \ge n \cdot 2^{-\frac{1}{2}\log n} = \sqrt{n} > n_0.$$

So by (6.14), $f(n') < \left(\frac{C}{2}\right)^{-1} f(n' \cdot g(n'))$. Since $n' < n$ and $g(x)$ is decreasing, $n' \cdot g(n') > n' \cdot g(n)$. Then by Fact 6.3.1, $f(n' \cdot g(n')) < 2f(n' \cdot g(n))$. Hence, $f(n') < \left(\frac{C}{4}\right)^{-1} f(n' \cdot g(n))$. Thus by the induction hypothesis, we have

$$\begin{aligned} f(n) &< \left(\frac{C}{4}\right)^{-j} f(n \cdot g(n)^j) = \left(\frac{C}{4}\right)^{-j} f(n') \\ &< \left(\frac{C}{4}\right)^{-j} \left(\frac{C}{4}\right)^{-1} f(n' \cdot g(n)) = \left(\frac{C}{4}\right)^{-(j+1)} f(n \cdot g(n)^{j+1}). \end{aligned}$$

Write $j = t$ in (6.15) and note that $f(n) \le 1$ and that $f(n \cdot g(n)^t) < 2f(\sqrt{n})$ by Fact 6.3.1:

$$f(n) < \left(\frac{C}{4}\right)^{-t} f(n \cdot g(n)^t) < \left(\frac{C}{4}\right)^{-t} \cdot 2f(\sqrt{n}) \le 2\left(\frac{C}{4}\right)^{-t} = 2\left(\frac{C}{4}\right)^{-\lfloor \frac{1}{2} \frac{\sqrt{\log n}}{5k+11} \rfloor} < 2^{-5\sqrt{\log n}}$$

for C sufficiently large, a contradiction. □

Theorem 6.1.4 follows immediately from Lemmas 6.2.5 and 6.3.2.

Proof of Theorem 6.1.4. Let K be the constant from Lemma 6.2.5. Let C be the constant and $\{n_i\}_{i=1}^{\infty}$ be the sequence from Lemma 6.3.2. Define $C' = 8CK$. Fix an arbitrary $n \in \{n_i\}_{i=1}^{\infty}$ and write $M = M(n)$ as defined in (6.12). Let $A \subseteq [n]$ be an arbitrary set of size $C'r(n)$. Then by Lemma 6.3.2,

$$\frac{|A|}{n} = \frac{8CK \cdot r(n)}{n} \ge 8K \frac{r(M(n))}{M(n)}.$$

By Fact 6.3.1, $\frac{2r(M)}{M} > \frac{r(n)}{n}$. Thus by Lemma 6.2.5 and that $K \ge 2$, $C > 4$, we have

$$\begin{aligned} \Gamma_k(A) &> \frac{|A|^2}{M^2} \cdot \frac{K^k \cdot r(M)}{2^{k+4}\log^2 n} = \frac{(8CK)^2 r(n)^2}{M^2} \cdot \frac{K^k \cdot r(M)}{2^{k+4}\log^2 n} = \frac{r(n)^2}{M\log^2 n} \cdot \frac{2r(M)}{M} \cdot \frac{(8CK)^2 K^k}{2^{k+5}} \\ &> \frac{r(n)^2}{M\log^2 n} \cdot \frac{2r(M)}{M} > \frac{r(n)^2}{M\log^2 n} \cdot \frac{r(n)}{n} = \log^{3k-2} n \left(\frac{n}{r(n)}\right)^{k-1} n. \end{aligned}$$

□

Proof of Theorem 6.1.2. Let $\{n_i\}_{i=1}^{\infty}$ be the infinite sequence guaranteed by Lemma 6.3.2.

We will show that the conclusion holds for this sequence of values of n. Let $M = M(n), \varepsilon = \varepsilon(n)$ and $\tau = \tau(n)$ be as defined in (6.12). For sufficiently large n, we have that $\tau < \frac{1}{200k^{2k}}$ and

$$\varepsilon n \tau^{k-1} = \frac{\log^{3k-2} n}{n} \left(\frac{n}{r(n)}\right)^{k-1} \cdot n \cdot \left(\frac{r(n)}{n}\frac{1}{\log^3 n}\right)^{k-1} = \log n > k^{3k}.$$

Thus by Corollary 6.2.2, there is a family $\mathcal{C}$ of containers such that every k-AP-free subset of $[n]$ is a subset of some container in $\mathcal{C}$. By (6.12), $\log \frac{1}{\varepsilon} \log \frac{1}{\tau} < \log^2 n$, thus

$$\log |\mathcal{C}| \leq 1000k^{3k} n\tau \log \frac{1}{\varepsilon} \log \frac{1}{\tau} < 1000k^{3k} n \cdot \frac{r(n)}{n}\frac{1}{\log^3 n} \cdot \log^2 n = o(r(n)).$$

Since for every container $A \in \mathcal{C}$, the number of k-APs in A is at most εn^2, then by Theorem 6.1.4, $|A| < C'r(n)$ for every $A \in \mathcal{C}$. Recall that every k-AP-free subset is contained in some member of $\mathcal{C}$. Hence, the number of k-AP-free subsets of $[n]$ is at most

$$\sum_{A\in\mathcal{C}} 2^{|A|} \leq |\mathcal{C}| \cdot \max_{A\in\mathcal{C}} 2^{|A|} < 2^{o(r(n))} \cdot 2^{C'r(n)} = 2^{O(r(n))}.$$

□

Proof of Corollary 6.1.3. Let $\{n_i\}_{i=1}^{\infty}$ be a sequence of integers for which the conclusion of Theorem 6.1.2 holds. Fix an arbitrary $\varepsilon > 0$ and n_i. From Theorem 6.1.2, we know that the number of k-AP-free subsets of $[n_i]$ is at most $2^{c\cdot r(n_i)}$ for some absolute constant $c > 0$. For any $\varepsilon n_i \leq m < n_i$, by (6.13), we have that $r(n_i) \leq \lceil \frac{1}{\varepsilon} \rceil \cdot r(m) < \frac{2}{\varepsilon} \cdot r(m)$. Therefore, by setting $b = 2c/\varepsilon$, we have that the number of k-AP-free subsets of $[m]$ is at most $2^{c\cdot r(n_i)} \leq 2^{b\cdot r(m)}$. It then follows that $m \in A(b)$ for any $\varepsilon n_i \leq m < n_i$ and that $|A(b) \cap [n_i]|/n_i \geq 1 - \varepsilon$ as desired. □

The proof of Theorem 6.1.5 is along the same lines as of the proof of Theorem 6.1.2, hence we provide here only a sketch of it.

Proof of Theorem 6.1.5. Fix an arbitrary $0 < \gamma < 1$. We apply Corollary 6.2.2 with $\varepsilon = n^{-\gamma/2}$, $\tau = n^{-\frac{1}{k-1}+\gamma/2}$ and let $\mathcal{C}$ be the family of containers of size $\log |\mathcal{C}| = o(n^{1-\frac{1}{k-1}+\gamma})$. Each container contains at most $\varepsilon n^2 = n^{2-\gamma/2}$ many k-APs. It follows that for every $A \in \mathcal{C}$, $|A| \leq \frac{C'n}{h(n)}$ for some $C' = C'(k, c, \gamma)$, since otherwise applying Lemma 6.2.5 on A with $M = n^{\gamma/4}$ would imply $\Gamma_k(A) > n^{2-\gamma/3} > \varepsilon n^2$, a contradiction. Thus, the number of k-AP-free subsets of $[n]$ is at most $|\mathcal{C}| \cdot 2^{C'n/h(n)} = 2^{2C'n/h(n)}$, as desired. Similarly, the number of k-AP-free m-subsets of $[n]$ is at most $|\mathcal{C}| \cdot \binom{C'n/h(n)}{m} \leq 2^m \cdot \binom{C'n/h(n)}{m} \leq \binom{2C'n/h(n)}{m}$, where the first inequality follows from $m \geq n^{1-\frac{1}{k-1}+\gamma} \geq \log |\mathcal{C}|$. □

References

[1] N. Alon, An extremal problem for sets with applications to graph theory. *J. Combin. Theory, Ser. A*, (40), 1985, 82–89.

[2] N. Alon, J. Balogh, B. Bollobás and R. Morris, The structure of almost all graphs in a hereditary property, *J. Combin. Theory, Ser. B*, (101), 2011, 85–110.

[3] N. Alon, J. Balogh, P. Keevash and B. Sudakov, The number of edge colorings with no monochromatic cliques, *J. London Math. Soc.*, (70), 2004, 273–288.

[4] N. Alon, J. Balogh, R. Morris and W. Samotij, A refinement of the Cameron-Erdős conjecture, *Proc. London Math. Soc.*, (108), 2014, 44–72.

[5] N. Alon and J.H. Spencer, The probabilistic method, 3rd edition, *Wiley*, New York, 2008.

[6] J. Balogh, B. Bollobás and R. Morris. Graph bootstrap percolation. *Random Structures Algorithms*, (41), 2012, 413–440.

[7] J. Balogh, B. Bollobás, R. Morris, O. Riordan. Linear algebra and bootstrap percolation. *J. Combin. Theory, Ser. A*, (119), 2012, 1328–1335.

[8] J. Balogh, B. Bollobás and M. Simonovits, The number of graphs without forbidden subgraphs, *J. Combin. Theory, Ser. B*, (91), 2004, 1–24.

[9] J. Balogh, B. Bollobás and M. Simonovits, The typical structure of graphs without given excluded subgraphs, *Random Structures Algorithms*, (34), 2009, 305–318.

[10] J. Balogh and J. Butterfield, Excluding induced subgraphs: critical graphs, *Random Structures Algorithms*, (38), 2011, 100–120.

[11] J. Balogh and H. Liu, On the number of K_4-saturating edges, *J. Combin. Theory, Ser. B*, (109), 2014, 250–257.

[12] J. Balogh, H. Liu, S. Petříčková and M. Sharifzadeh, The typical structure of maximal triangle-free graphs, *Forum of Mathematics, Sigma*, to appear.

[13] J. Balogh, H. Liu, M. Sharifzadeh, The number of subsets of integers with no k-term arithmetic progression, submitted.

[14] J. Balogh, H. Liu, M. Sharifzadeh and A. Treglown, The number of maximal sum-free subsets of integers, *Proc. Amer. Math. Soc.*, (143) 2015, 4713–4721.

[15] J. Balogh, H. Liu, M. Sharifzadeh and A. Treglown, Sharp bound on the number of maximal sum-free subsets of integers, submitted.

[16] J. Balogh, N. Bushaw, M. Collares Neto, H. Liu, R. Morris and M. Sharifzadeh, The typical structure of graphs with no large cliques, *Combinatorica*, to appear.

[17] J. Balogh, R. Morris and W. Samotij, Independent sets in hypergraphs, *J. Amer. Math. Soc.*, (28) 2015, 669–709.

[18] J. Balogh, R. Morris, W. Samotij and L. Warnke, The typical structure of sparse K_{r+1}-free graphs, *Trans. Amer. Math. Soc*, to appear.

[19] J. Balogh and S. Petříčková, The number of the maximal triangle-free graphs, *Bull. London Math. Soc.*, (46) 2014, 1003–1006.

[20] F.A. Behrend, On sets of integers which contain no three terms in arithmetical progression, *Proc. Nat. Acad. Sci. U.S.A.*, (32) 1946, 331–332.

[21] B. Bollobás, Weakly k-saturated graphs. *Beiträlge zur Graphentheorie*, (H. Sachs, H.J. Voss, and H. Walther, Eds.), pp. 25–31, Leipzig, Barth, 1968.

[22] B. Bollobás and V. Nikiforov, The number of graphs with large forbidden subgraphs, *European J. Combin.*, (32), 2010, 1964–1968.

[23] B. Bollobás and A. Thomason, Projections of bodies and hereditary properties of hypergraphs, *Bull. London Math. Soc.*, (27), 1995, 417–424.

[24] J. Bourgain, On triples in arithmetic progression, *Geom. Funct. Anal.*, (9) 1999, 968–984.

[25] P. Cameron and P. Erdős, On the number of sets of integers with various properties, in Number Theory (R.A. Mollin, ed.), 61–79, Walter de Gruyter, Berlin, 1990.

[26] P. Cameron and P. Erdős, Notes on sum-free and related sets, *Combin. Probab. Comput.*, (8), 1999, 95–107.

[27] D. Conlon and W.T. Gowers, Combinatorial theorems in sparse random sets, submitted.

[28] E. Croot and O. Sisask, A new proof of Roth's Theorem on Arithmetic progressions, *Proc. Amer. Math. Soc.*, (137) 2009, 805–809.

[29] S. Das, W. Gan and B. Sudakov, Sperner's Theorem and a problem of Erdős-Katona-Kleitman. *Combin. Probab. Comput.*, (24), 2015, 585–608.

[30] J. Deshouillers, G. Freiman, V. Sós and M. Temkin, On the structure of sum-free sets II, *Astérisque*, (258), 1999, 149–161.

[31] P.H. Diananda and H.P. Yap, Maximal sum-free sets of elements of finite groups, *Proc. Japan Acad.*, (45), 1969, 1–5.

[32] M. Elkin, An improved construction of progression-free sets, *Israel J. Math.*, (184) 2011, 93–128.

[33] P. Erdős, On a theorem of Rademacher-Turán. *Illinois Journal of Math*, (6), 1962, 122–127.

[34] P. Erdős, On the number of complete subgraphs contained in certain graphs. *Magy. Tud. Acad. Mat. Kut. Int. Közl.*, (7), 1962, 459–474.

[35] P. Erdős, Some of my old and new combinatorial problems. *Paths, flows, and VLSI-layout (Bonn, 1988), Algorithms Combin.*, (9), 35–45, Springer, Berlin, 1990.

[36] P. Erdős, On extremal problems of graphs and generalized graphs, *Israel J. Math.*, (2), 1964, 183–190.

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

[38] P. Erdős, D.J. Kleitman and B.L. Rothschild, Asymptotic enumeration of K_n-free graphs, Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II, 19–27. Atti dei Convegni Lincei, No. 17, Accad. Naz. Lincei, Rome, 1976.

[39] P. Erdős and A. Rényi, On Random Graphs, *Publicationes Mathematicae*, (6), 1959, 290–297.

[40] P. Erdős and M. Simonovits, A limit theorem in graph theory, *Studia Sci. Math. Hungar.*, (1), 1966, 51–57.

[41] P. Erdős and A.H. Stone, On the structure of linear graphs, *Bull. Amer. Math. Soc.*, (52) 1946, 1087–1091.

[42] P. Erdős and G. Szekeres, A combinatorial problem in geometry, *Compos. Math.*, (2), 1935, 463–470.

[43] P. Frankl, An extremal problem for two families of sets. *Europ. J. Math.*, (3), 1982, 125–127.

[44] Z. Füredi, A proof of the stability of extremal graphs, Simonovits' stability from Szemerédi's regularity, *J. Combin. Theory, Ser. B*, (115), 2015, 66–71.

[45] Z. Füredi, The number of maximal independent sets in connected graphs, *J. Graph Theory*, (11), 1987, 463–470.

[46] W.T. Gowers, A new proof of Szemerédi's theorem for progressions of length four, *Geom. Func. Anal.*, (8) 1998, 529–551.

[47] W.T. Gowers, A new proof of Szemerédi's theorem, *Geom. Func. Anal.*, (11) 2001, 465–588.

[48] B. Green, A Szemerédi-type regularity lemma in abelian groups, with applications, *Geom. Funct. Anal.*, (15) 2005, 340–376.

[49] B. Green, The Cameron-Erdős conjecture, *Bull. London Math. Soc.*, (36), 2004, 769–778.

[50] B. Green, R. Morris, Counting sets with small sumset and applications, *Combinatorica*, to appear.

[51] B. Green and I. Ruzsa, Counting sumsets and sum-free sets modulo a prime, *Studia Sci. Math. Hungarica*, (41), 2004, 285–293.

[52] B. Green and I. Ruzsa, Sum-free sets in abelian groups, *Israel J. Math.*, (147), 2005, 157–189.

[53] B. Green and J. Wolf, A note on Elkin's improvement of Behrend's construction, *Additive number theory*, Springer, New York, 2010, 141–144.

[54] D.R. Heath-Brown, Integer sets containing no arithmetic progressions, *J. London Math. Soc. (2)*, (35) 1987, 385–394.

[55] M. Hujter and Z. Tuza, The number of maximal independent sets in triangle-free graphs, *SIAM J. Discrete Math.*, (6), 1993, 284–288.

[56] L. Ilinca and J. Kahn, Counting maximal antichains and independent sets, *Order*, (30), 2013, 427–435.

[57] G. Kalai, Weakly saturated graphs are rigid, in Convexity and Graph Theory, M. Rosenfeld and J. Zaks (eds). *Ann. Discrete Math.*, (20), 1984, 189–190.

[58] N. Khadžiivanov, V. Nikiforov, Solution of a problem of P. Erdős about the maximum number of triangles with a common edge in a graph (Russian). *C. R. Acad. Bulgare Sci.*, (32), 1979, 1315–1318.

[59] D. Kleitman, A conjecture of Erdős-Katona on commensurable pairs among subsets of an n-set. *Theory of Graphs, Proc. Colloq., Tihany*, 1966, 215–218.

[60] D. Kleitman, On Dedekind's problem: The number of monotone Boolean functions, *Proc. Amer. Math. Soc.*, (21) 1969, 677–682.

[61] D. Kleitman and D. Wilson, On the number of graphs which lack small cycles, manuscript, 1996.

[62] Y. Kohayakawa, S. Lee, V. Rödl and W. Samotij, The number of Sidon sets and the maximum size of Sidon sets contained in a sparse random set of integers, *Random Structures and Algorithms*, (46) 2015, 1–25.

[63] P. G. Kolaitis, H. J. Prömel and B. L. Rothschild, $K_{\ell+1}$-free graphs: asymptotic structure and a 0-1 law, *Trans. Amer. Math. Soc.*, (303), 1987, 637–671.

[64] D. Král', O. Serra and L. Vena, A combinatorial proof of the Removal Lemma for groups, *J. Combin. Theory A*, (116), 2009, 971–978.

[65] L. Lovász and M. Simonovits, On the number of complete subgraphs of a graph, II. *Studies in pure mathematics*, 459–495, Birkhuser, Basel, 1983.

[66] T. Łuczak and T. Schoen, On the number of maximal sum-free sets, *Proc. Amer. Math. Soc.*, (129), 2001, 2205–2207.

[67] W. Mantel, Problem 28. *Winkundige Opgaven*, (10), 1907, 60–61.

[68] J. W. Moon and L. Moser, On cliques in graphs, *Israel J. Math.*, (3), 1965, 23–28.

[69] R. Morris and D. Saxton, The number of C_{2k}-free graphs, *Advances in Math.*, to appear.

[70] F. Mousset, R. Nenadov and A. Steger, On the number of graphs without large cliques, *Siam J. Discrete Math.*, (28), 2014, 1980–1986.

[71] D. Mubayi, Counting substructures I: color critical graphs. *Advances in Mathematics*, (225), 2010, 2731–2740.

[72] D. Mubayi, Counting substructures II: hypergraphs. *Combinatorica*, (33), 2013, 591–612.

[73] D. Osthus, H. J. Prömel and A. Taraz, For which densities are random triangle-free graphs almost surely bipartite?, *Combinatorica*, (23), 2003, 105–150.

[74] H. J. Prömel and A. Steger, The asymptotic number of graphs not containing a fixed color-critical subgraph, *Combinatorica*, (12), 1992, 463–473.

[75] F.P. Ramsey, On a problem of formal logic, *Proc. London Math. Soc.*, (30), 1930, 264–286.

[76] R.A. Rankin, Sets of integers containing not more than a given number of terms in arithmetical progression, *Proc. Roy. Soc. Edinburgh Sect. A*, (65) 1960/1961, 332–344.

[77] K.F. Roth, On certain sets of integers, *J. London Math. Soc.*, (28) 1953, 104–109.

[78] T. Sanders, On Roth's theorem on progressions, *Ann. of Math. (2)*, (174) 2011, 619–636.

[79] A. A. Sapozhenko, The Cameron-Erdős conjecture, (Russian) *Dokl. Akad. Nauk.*, (393), 2003, 749–752.

[80] A. A. Sapozhenko, The number of independent sets in graphs, *Moscow Univ. Math. Bull.*, (62), 2004, 116–118.

[81] D. Saxton and A. Thomason, Hypergraph containers, *Invent. Math.*, (201), 2015, 925–992.

[82] M. Schacht, Extremal results for random discrete structures, submitted.

[83] I. Schur, Uber die Kongruenz $x^m + y^m \equiv z^m \pmod{p}$, *ber. Deutsch. Mat. Verein.*, (25), 1916, 114–117.

[84] A. Selberg, An elementary proof of the prime-number theorem, *Ann. of Math. (2)*, (50) 1949, 305–313.

[85] E. Sperner, Ein Satz über Untermengen einer endlichen Menge, *Mathematische Zeitschrift*, (27), 1928, 544–548.

[86] E. Szemerédi, Integer sets containing no arithmetic progressions, *Acta Math. Hungar.*, (56), 1990, 155–158.

[87] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.*, (27), 1975, 199–245.

[88] P. Turán, Eine Extremalaufgabe aus der Graphentheorie, *Mat. Fiz. Lapok*, (48), 1941, 436–452.

[89] P. Varnavides, On certain sets of positive density, *J. London Math. Soc.*, (34), 1959, 358–360.

[90] G. Wolfovitz, Bounds on the number of maximal sum-free sets, *European J. Combin.*, (30), 2009, 1718–1723.