ENUMERATING COMBINATORIAL OBJECTS WITH LIMITED SUB-CONFIGURATIONS

BY

LINA LI

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, 2020

Urbana, Illinois

Doctoral Committee:

Professor Alexandr Kostochka, Chair
Professor József Balogh, Director of Research
Professor Kevin Ford
Research Assistant Professor Mikhail Lavrov

Abstract

Many well-studied problems in extremal combinatorics concern the number and the typical structure of discrete objects with forbidden substructures. Over the past decades, such problems have been extensively studied for various objects by many notable researchers. This thesis focuses on several problems of this type using various techniques.

In Chapter 2, we investigate the family of linear hypergraphs with forbidden linear cycles. A substantial part of the work indeed focuses on a closely related problem, the study of the family of graphs with limited short even cycles, which may be of independent interest. To attack this problem, we introduce a new variant of the graph container algorithm. Another application of it to additive combinatorics is presented in Chapter 3 on generalized Sidon sets.

In Chapter 4, we investigate an enumeration problem on Gallai colorings, i.e. rainbow triangle-free colorings. In particular, we describe the typical structure of Gallai r-colorings of complete graphs, and complete the characterization of the extremal graphs for Gallai colorings. This work heavily relies on the hypergraph container method, and some ad-hoc stability analysis.

Another closely related problem is the study of sparse analogue of classical extremal results in random graphs, for example, the Erdős-Stone theorem, as it can also be interpreted as counting graphs in the corresponding probability space. In Chapter 5, we show a random analogue of the famous Erdős-Gallai theorem on extremal functions of paths.

To my parents.

Acknowledgments

First, I would like to express my deepest appreciation to my advisor, Professor József Balogh, for all of his support and guidance throughout my Ph.D. study. His professionalism, enthusiasm, and patience inspired and encouraged me in all the time of research and writing of this thesis.

I would like to extend my sincere thanks to a number of professors who inspired, encouraged and helped me during the past years: Alexandr Kostochka, Theodore Molla, Yuliy Baryshnikov, Philipp Hieronymi, Mikhail Lavrov, Kevin Ford, and Partha Dey. I have learned so much from all of you and enjoyed many interesting conversations with you.

I am also grateful to many of my friends and fellow graduate students: Yang Bai, Linlin Bu, Lini Xu, Chenyu Liang, Heejoung Kim, Xujun Liu, Ruth Luo, Emily Heath, Adam Wagner, Felix Clemen, Maryam Sharifzadeh, Michelle Delcourt, Ping Hu, Yifan Jing. Thank you for listening, offering me advice, and supporting me through this entire process, and making my time in grad school more entertaining.

A very special word of thanks goes for my parents, who have been great over the years and standing beside me with their love and unconditional support. Finally, I would like to thank my partner, Gemeng Zhang, for all of his encouragement, patience, support and love.

Table of Contents

Symbols and Notation

$\emptyset$	the empty set
$[n]$	for $n \in \mathbb{N}$, $[n] := \{1, \ldots, n\}$
$\mathbb{N}$	the set of natural numbers
$\mathbb{Z}$	the set of integers
$\log$	logarithm base 2
$\log_r$	logarithm base r
$\ln$	logarithm base e
$V(G)$	the vertex set of a (hyper)graph G
$v(G)$	$v(G) := \|V(G)\|$
$E(G)$	the edge set of a (hyper)graph G
$e(G)$	$e(G) := \|E(G)\|$
$N_G(v),\ N(v)$	the neighborhood of a vertex
$d_G(v),\ d(v)$	the degree of a vertex
$N_G(v,S),\ N(v,S)$	the neighborhood of a vertex restricted to S
$d_G(v,S),\ d(v,S)$	the degree of a vertex restricted to S
$\delta(G)$	the minimum degree of a graph
$\Delta(G)$	the maximum degree of a graph
$G[A]$	the subgraph induced in a graph G by a set A
$G[A,B]$	the bipartite subgraph induced in a graph G by two disjoint sets A and B
K_n	the complete graph on n vertices
$K_{s,t}$	the complete bipartite graph with parts of size s and t
C_n	the cycle on n vertices
P_n	the path on n vertices

Chapter 1

Introduction

One of the central challenges in extremal combinatorics is to determine the extremal and typical properties of the family of combinatorial objects with certain forbidden configurations. Over the past decades, this problem has been extensively studied for various discrete objects, such as graphs, hypergraphs, sets, and Boolean lattices, by many notable researchers. Many advances in this area not only discovered some crucial extremal phenomena exhibited in combinatorial objects, but also promoted the development of classical and new techniques, including but not limited to the entropy method, the probabilistic method, the graph container algorithm, and the hypergraph container method. In this thesis, we study several problems on graphs, hypergraphs, and additive sets, which fit in this area.

1.1 Enumerating (hyper)graphs with limited substructures via graph containers

1.1.1 Graphs with limited even cycles

For a graph H, we say a graph G is *H-free* if it contains no subgraph isomorphic to H. The *Turán number* of H, denoted by $\mathrm{ex}(n,H)$, is the maximum number of edges among n-vertex H-free graphs. In the 1970s, Erdős, Kleitman and Rothschild [35] introduced the problem of determining the number of H-free graphs on n vertices. For non-bipartite H, the answer has been well-understood since 1986, when Erdős, Frankl and Rödl [39], extending a result of Erdős, Kleitman and Rothschild [35] on cliques, proved that there are $2^{(1+o(1))\mathrm{ex}(n,H)}$ such graphs. More than thirty years ago, Erdős made the following conjecture (see, e.g., [68]).

Conjecture 1.1 (Erdős)**.** *The number of H-free graphs on n vertices is at most $2^{O(\mathrm{ex}(n,H))}$ for every graph H.*

It turns out that this problem is significantly harder for bipartite graphs, especially for even cycles. Recall that $\mathrm{ex}(n,C_{2\ell}) = O(n^{1+1/\ell})$ for every $\ell \geq 2$. Erdős and Simonovits conjectured that this bound is sharp up to the implied constant factor, while matching lower bounds are known only for $\ell \in \{2,3,5\}$ (see, e.g., [48]). The first breakthrough was made by Kleitman and Winston [68] in 1982, who showed that there are at most $2^{2.17\mathrm{ex}(n,C_4)}$ n-vertex C_4-free graphs, using the so called *graph container method*. Kleitman and Wilson [67], and independently Kreuter [77], and Kohayakawa, Kreuter, and Steger [69] later proved that there are $2^{O(n^{1+1/\ell})}$ graphs with no even cycles of length at most 2ℓ. However, they were unable to resolve the case of a single forbidden long even cycle. It was not until 2016 that Morris and Saxton [86] proved Conjecture 1.1 for all even cycles, using the *hypergraph container method*.

Theorem 1.2 (Morris and Saxton [86])**.** *For every $\ell \geq 2$, there are at most $2^{O(n^{1+1/\ell})}$ $C_{2\ell}$-free graphs on n vertices.*

Note that the supersaturation phenomenon indicates that if the number of edges in a graph G exceeds the extremal number, then G would contain many even cycles. This leads to an interesting question.

Problem 1.3. *What is the maximum number of $C_{2\ell}$'s we could allow a graph to have so that the number of such graphs is still $2^{O(n^{1+1/\ell})}$?*

As outlined here and appearing in Chapter 2, in joint work with Balogh [10], we gave some partial answers to this question.

Theorem 1.4. *Let $a = \Theta(\log^5 n)$. The number of n-vertex graphs with at most n^2/a C_4's is $2^{O(n^{3/2})}$.*

A standard probabilistic argument shows that $a = \Theta(\log^4 n)$ would be the best possible in Theorem 1.4, and we believe that it should be the truth. For longer cycles, although the answer to Problem 1.3 is still unknown, we believe that limiting just one even cycle has essentially the same effect as limiting all smaller cycles as well. Therefore, a natural attempt is to determine how one can further restrict the number of other short cycles, thus obtaining a proof for the desired upper bound.

Theorem 1.5. *For an integer $\ell \geq 3$ and a constant $L > 0$, denote by $\mathcal{G}_n(\ell, L)$ the family of n-vertex graphs G such that for every $3 \leq k \leq \ell$ and $e \in E(G)$, the number of k-cycles containing e is at most L. For n sufficiently large, we have $|\mathcal{G}_n(2\ell, L)| \leq 2^{3(\ell+1)n^{1+1/\ell}}$.*

Despite its own interest, Theorems 1.4 and 1.5 also have applications on hypergraph enumeration problems, which will be presented in the next section. This is also one of our initial motivations to study Problem 1.3.

1.1.2 Linear hypergraphs with no linear cycles

For an r-graph H, the *Turán number* of H, denoted by $\mathrm{ex}_r(n, H)$, is the maximum number of edges among all r-graphs on n vertices which contain no copy of H as a subgraph. Mirroring the situation described earlier for graphs, for each $r \geq 3$, it is generally believed that the number of H-free r-graphs on n vertices is at most $2^{O(\mathrm{ex}_r(n,H))}$ for every r-graph H. Indeed, it follows from the work of Nagle, Rödl and Schacht [88] on hypergraph regularity that for any fixed r-graph H, there are $2^{O(\mathrm{ex}_r(n,H))+o(n^r)}$ such r-graphs, which gives a reasonably satisfactory solution in the case where H is not r-partite.

Unsurprisingly, the enumeration problem for a fixed forbidden r-partite r-graph H is much harder and less understood. In recent years, one such prototypical family of r-partite r-graphs, namely, the family of r-uniform linear (or loose) cycles, has received much attention in the literature. For integers $r \geq 2$ and $\ell \geq 3$, an r-uniform *linear cycle* of length ℓ, denoted by C_ℓ^r, is an r-graph with edges $e_1, \ldots, e_\ell$ such that for every $i \in [\ell-1]$, $|e_i \cap e_{i+1}| = 1$, $|e_\ell \cap e_1| = 1$ and $e_i \cap e_j = \emptyset$ for all other pairs $\{i,j\}$, $i \neq j$. Kostochka, Mubayi and Verstraëte [76], and independently, Füredi and Jiang [47] proved that for every $r, \ell \geq 3$, $\mathrm{ex}_r(n, C_\ell^r) = \Theta(n^{r-1})$. Continuing the work of Mubayi and Wang [87], Han and Kohayakawa [54], Balogh, Narayanan, and Skokan [13] proved the following result using the hypergraph container method.

Theorem 1.6 (Balogh, Narayanan, and Skokan [13])**.** *For every pair of integers $r, k \geq 3$, there exists $C = C(r,k) > 0$ such that the number of C_ℓ^r-free r-graphs is at most $2^{Cn^{r-1}}$ for all $n \in \mathbb{N}$.*

An r-graph H is said to be *linear* if for every $e, e' \in E(H)$, $|e \cap e'| \leq 1$. Since the above forbidden substructure is a linear hypergraph, it seems natural to switch the host hypergraphs to linear hypergraphs. For a linear r-graph H, the *linear Turán number* of H, denoted by $\mathrm{ex}_L(n,H)$, is the maximum number of edges among linear r-graphs on n vertices which contain no copy of H as a subgraph. In 1968, Erdős, Frankl and Rödl [40] showed that for every $r \geq 3$, $\mathrm{ex}_L(n, C_3^r) = o(n^2)$ and $\mathrm{ex}_L(n, C_3^r) = \Omega(n^c)$ for every $c < 2$. Collier-Cartaino, Graber and Jiang [27], resolving a conjecture of Kostochka, Mubayi, and Verstraëte [75], proved that $\mathrm{ex}_L(n, C_\ell^r) = O\left(n^{1+\frac{1}{\lfloor \ell/2 \rfloor}}\right)$ for $r \geq 3$ and $\ell \geq 4$. However, the matching lower bound is only known for C_4^3 and C_5^3.

Denote by $\mathrm{Forb}_r(n, C_\ell^r)$ the family of C_ℓ^r-free r-uniform linear hypergraphs. For $\ell = 3$, the work of Erdős, Frankl and Rödl [40] could be extended to show that $|\mathrm{Forb}_L(n, C_3^r)| = 2^{o(n^2)}$ for every $r \geq 3$. Similarly to all existing results in the area, it is natural for us to conjecture that $|\mathrm{Forb}_r(n, C_\ell^r)| = 2^{\Theta(n^{1+1/\lfloor \ell/2 \rfloor})}$, for $r \geq 3$ and $\ell \geq 4$. In [10] with Balogh, we confirmed this conjecture for any $r \geq 3$ and $\ell = 4$.

Theorem 1.7. *For every $r \geq 3$, we have $|\mathrm{Forb}_r(n, C_4^r)| = 2^{O(n^{3/2})}$.*

For longer linear cycles, we provided an upper bound for the girth version.

Theorem 1.8. *For every $r \geq 3$ and $\ell \geq 4$, let $\mathrm{Forb}_L(n, r, \ell)$ denote the set of all linear r-graphs on $[n]$ with girth at least ℓ. Then we have $|\mathrm{Forb}_L(n,r,\ell)| = 2^{O(n^{1+1/\lfloor \ell/2 \rfloor})}$.*

The upper bound for C_4^3 is sharp in order of magnitude given by $\mathrm{ex}_L(n, C_4^3) = \Theta(n^{3/2})$. In general, both upper bounds are possibly sharp, but we are not able to confirm it now, as the sharp bound for the corresponding linear Turán number remains open.

Theorems 1.7 and 1.8 are indeed consequences of Theorems 1.4 and 1.5 by considering the *shadow graphs* of hypergraphs, see Chapter 2 for details.

1.1.3 Generalized Sidon sets

A set A of nonnegative integers is a *Sidon set* if there is no *Sidon 4-tuple*, i.e. a 4-tuple (a, b, c, d) in A with $a + b = c + d$ and $\{a, b\} \cap \{c, d\} = \emptyset$. Denote by $\Phi(n)$ the maximum size of Sidon subsets of $[n]$. Studies of Erdős and Turán [37], Singer [96], Erdős [33], and Chowla [26], answering a famous problem of Sidon, have showed that $\Phi(n) = (1 + o(1))\sqrt{n}$. Cameron and Erdős [25] first proposed the problem of determining the number of Sidon subsets in $[n]$. The extremal result indicates that there is a trivial lower bound $2^{\Phi(n)}$ and a trivial upper bound $2^{O(\sqrt{n}\log n)}$. This problem has been studied by Kohayakawa, Lee, Rödl and Samotij [70] with the graph container method, and by Saxton and Thomason [94] with the hypergraph container method, showing that neither of the trivial bounds is tight.

Theorem 1.9 (Kohayakawa, Lee, Rödl and Samotij [70], Saxton and Thomason [94])**.** *For sufficiently large enough n, the number of Sidon subsets in $[n]$ is between $2^{(1.16+o(1))\sqrt{n}}$ and $2^{c\Phi(n)}$, where $c = \log(32e) \approx 6.442$.*

An *α-generalized Sidon set* in $[n]$ is a set with at most α Sidon 4-tuples. Motivated by Theorem 1.4 and the closed connection between Sidon sets and C_4-free graphs, in [9] we investigate the maximum value of α for which the number of α-generalized Sidon subset of $[n]$ is still $2^{O(\sqrt{n})}$.

Theorem 1.10. *For $\alpha = O(n/\log^5 n)$, the number of α-generalized Sidon sets in $[n]$ is $2^{\Theta(\sqrt{n})}$.*

See Chapter 3 for the proof of Theorem 1.10. A simple probabilistic argument shows that for $\alpha \gg \sqrt{n}/\log^4 n$, there are $2^{\Theta((\alpha n)^{\frac{1}{4}}\log n)} \gg 2^{\Theta(\sqrt{n})}$ subsets with $\Theta(\alpha)$ Sidon 4-tuples. Therefore, we made the following conjecture.

Conjecture 1.11. *For $\alpha = \Theta(n/\log^4 n)$, the number of α-generalized Sidon sets in $[n]$ is $2^{\Theta(\sqrt{n})}$.*

1.1.4 The graph container algorithm

In 1982, Kleitman and Winston [68] proved that the number of C_4-free graphs on n vertices is at most $2^{cn^{3/2}}$ for $c \approx 1.081919$. This seminal paper not only resolved a longstanding open question posed by Erdős, but also authored one of the first papers in the field whose main idea was to find small certificates of families of sets in order to prove that there are not many of them. These so-called *graph containers* have emerged as powerful tools for attacking problems of counting discrete objects with certain forbidden sub-configurations, for example, the number of C_4-free graphs.

Roughly speaking, the graph container method constructs a relatively simple algorithm which can be used to produce a 'small' number of subgraphs (referred to as *containers*), so that every C_4-free graph is contained in one of such containers, and each of these containers is an 'almost C_4-free graph'. For an intuitive explanation and more applications of this method, we refer readers to an excellent survey of Samotij [93].

Like many of these advances, our proofs of Theorems 1.4, 1.5 and 1.10 are built on the graph container algorithm. However, the previous applications address the problems for discrete objects with forbidden sub-configurations, while we concern the ones with a small amount of sub-configurations. Therefore, the means by which we apply this technique is quite non-standard, and requires some new ideas, see Sections 2.2.2, 2.3.3 and 3.3 for applications of this variant of the graph container algorithm to graph theory and additive combinatorics.

1.2 Enumerating Gallai colorings via the hypergraph container method

1.2.1 Background and main results

An interesting direction of combinatorics in recent years is the study of multicolored version of classical extremal results, whose origin can be traced back to a question of Erdős and Rothschild [34].

Problem 1.12 (Erdős-Rothchild problem, 1974)**.** *Which n-vertex graph has the maximum number of two-edge-colorings without monochromatic triangles?*

Erdős and Rothschild believed that the restrictions from the triangles would more than counteract the extra possibilities offered by the additional edges, and therefore conjectured that the maximal triangle-free graph is the only extremal graph. About twenty years later, Yuster [100] confirmed this conjecture for sufficiently large n.

There are many natural generalizations of the Erdős-Rothschild problem. The most obvious one may be to ask it for graphs other than the triangles, and one may also increase the number of colors used. A graph G on n vertices is called *(r, F)-extremal* if it admits the maximum number of r-edge-colorings without any monochromatic copies of F among all n-vertex graphs. Alon, Balogh, Keevash and Sudakov [2] greatly extended Yuster's result.

Theorem 1.13 (Alon, Balogh, Keevash and Sudakov [2]). *For $r = \{2, 3\}$ and $k \geq 3$, the Turán graph $T_k(n)$ is the unique (r, K_{k+1})-extremal graph.*

Interestingly, they also showed that Turán graphs $T_k(n)$ are no longer optimal for $r \geq 4$. Indeed, Pikhurko, and Yilma [91] later proved that $T_4(n)$ is the unique $(4, K_3)$-extremal graph, while $T_9(n)$ is the unique $(4, K_4)$-extremal graph. Determining the extremal configurations in general for $k \geq 2$ and $r \geq 4$ turned out to be a difficult problem. For further results along this line of research (when F is a non-complete graph or a hypergraph), we refer readers to [57, 58, 59, 82, 83, 84].

Another variant of this problem is to study edge-colorings of a graph avoiding a copy of F with a prescribed color pattern. For a r-colored graph $\hat{F}$, a graph G on n vertices is called $(r, \hat{F})$-extremal if it admits the maximum number of r-colorings which contain no subgraph whose color pattern is isomorphic to $\hat{F}$. This line of work was initiated by Balogh [5], who showed that the Turán graph $T_k(n)$ once again yields the maximum number of 2-colorings avoiding H_{k+1}, where H_{k+1} is any 2-coloring of K_{k+1} that uses both colors. For $r \geq 3$, the behavior of (r, H_{k+1})-extremal graphs was studied by Benevides, Hoppen, Sampaio, Lefmann, and Odermann, see [19, 60, 61, 62, 63]. In particular, the case when $\hat{F} = \hat{K}_3$ is a triangle with rainbow pattern has recently received a lot of attention.

An edge coloring of a graph G is a *Gallai coloring* if it contains no rainbow triangle. Improving the results of Falgas Ravry, O'Connell, and Uzzell [42], Benevides, Hoppen, and Sampaio [19], and Bastos, Benevides, Mota, and Sau [29], we give a sharp upper bound on the number of Gallai colorings of nearly complete graphs.

Theorem 1.14. *For every integer $r \geq 3$, there exists n_0 such that for all $n > n_0$, the number of Gallai r-colorings of the complete graph K_n is at most*

$$\left(\binom{r}{2} + 2^{-\frac{n}{4\log^2 n}}\right) 2^{\binom{n}{2}}.$$

Note that the number of Gallai r-colorings of the complete graph K_n with at most 2 colors is $\binom{r}{2}2^{\binom{n}{2}} - r(r-2)$. As a direct consequence, this theorem describes the typical structure of Gallai r-colorings for complete graphs.

Corollary 1.15. *For every integer $r \geq 3$, almost all Gallai r-colorings of the complete graph are 2-colorings.*

Now we turn to the extremal configurations of Gallai colorings. A n-vertex graph G is *Gallai r-extremal* if its number of Gallai r-colorings is the maximum over all n-vertex graphs. Hoppen, Lefmann and Odermann [62] determined the Gallai r-extremal graphs for $r \geq 5$.

Theorem 1.16 (Hoppen, Lefmann, and Odermann [62]). *For all $r \geq 5$, there exists n_0 such that for all $n > n_0$, the only Gallai r-extremal graph of order n is the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.*

We, confirming conjectures of Benevides, Hoppen, and Sampaio [19] and Hoppen, Lefmann, and Odermann [62], determined the extremal graphs for $r \in \{3, 4\}$.

Theorem 1.17. *For n sufficiently large, the graph K_n is the unique Gallai 3-extremal graph, while for $r \geq 4$, the graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ is the unique Gallai r-extremal graph.*

1.2.2 The hypergraph container method

Many important theorems and conjectures in extremal combinatorics, such as the sparse random analogue of Erdős-Stone theorem (see [28, 95]), the KŁR conjecture (see [11]) and the number of $C_{2\ell}$-free graphs (see [86]), can be phrased as statements about families of independent sets in certain uniform hypergraphs. In 2015, two independent groups, Balogh, Morris and Samotij [11] and Saxton and Thomason [94], introduced a new approach to the problem of understanding the family of independent sets in a hypergraph. This approach allows one to prove enumerative, structural, and extremal results in a wide variety of settings, and now is well-known as the *hypergraph container method*.

Roughly speaking, the hypergraph container method describes a clustering phenomenon exhibited by the independent sets of many hypergraphs whose edges are sufficiently evenly distributed. For a given hypergraph graph H, it builds machinery to produce a 'relatively small' amount of subsets of $V(H)$, referred to as *containers*, such that every independent set is contained in one of the containers, and each of these containers is 'almost independent'. For more details on the method, we refer readers to the original papers of Balogh, Morris and Samotij [11] and Saxton and Thomason [94], and also a recent survey written by Balogh, Morris and Samotij [12].

A substantial part of the proofs of Theorems 1.14 and 1.17 relies on this hypergraph container method. In Section 4.2, we present a version of the hypergraph container theorem from Balogh and Solymosi [14]), and show how to apply it in the context of edge colorings.

1.3 A random analogue of Erdős–Gallai theorem via the probabilistic method

A celebrated theorem of Erdős and Gallai [41] from 1959 determines the maximum number of edges in an n-vertex graph with no k-vertex path P_k.

Theorem 1.18 (Erdős and Gallai [41]). *For $n, k \geq 2$, if G is an n-vertex graph with no copy of P_k, then the number of edges of G satisfies $e(G) \leq \frac{1}{2}(k-2)n$. If n is divisible by $k-1$, then the maximum is achieved by a union of disjoint copies of K_{k-1}.*

An important direction of combinatorics in recent years is the study of sparse random analogues of classical extremal results. For graphs G and F, we write $\mathrm{ex}(G, F)$ for the maximum number of edges in an F-free subgraph of G. We write $G(n, p)$ for the standard binomial model of random graphs, where each edge in an n-vertex graph is chosen independently with probability p. The breakthrough papers of Conlon and Gowers [28] and Schacht [95], proved a sparse random version of the Erdős-Stone theorem, showing a *transference principle* of Turán function $\mathrm{ex}(K_n, F)$, i.e. the maximum number of edges in an F-free n-vertex graph. Here we present the graph version of their result.

Theorem 1.19 (Conlon and Gowers [28], Schacht [95]). *For every graph F with at least one vertex contained in at least two edges and every $\varepsilon \in (0, 1 - \pi(F))$, there exists constants $C > c > 0$ such that*

$$\lim_{n\to\infty} \mathbb{P}(\mathrm{ex}(G(n,p), F) \leq (\pi(F) + \varepsilon)e(G(n,p)) = \begin{cases} 0, & \text{if } p \leq cn^{-1/m_2(F)}, \\ 1, & \text{if } p \geq Cn^{-1/m_2(F)}, \end{cases}$$

where

$$m_2(F) = \max_{F' \subset F, v_{F'} \geq 3} \frac{e_{F'} - 1}{v_{F'} - 2}, \qquad \pi(F) = \lim_{n\to\infty} \mathrm{ex}(K_N, F)/\binom{n}{2}.$$

Note that this question can also be stated in the language of hypergraphs. Roughly speaking, these transference theorems say that if the edges of a hypergraph H are sufficiently uniformly distributed, then the independence number of H is well-behaved with respect to taking subhypergraphs induced by (sufficiently dense) random subsets of the vertex set. Via the hypergraph container method ([11] and [94]), the same results were proved, even when $|F|$ is a reasonable large function of n.

However, when F is a k-vertex path P_k, this result only gives a weak random analogue of the famous Erdős-Gallai theorem for paths with a fixed size, as the Turán density is zero. In joint work with Balogh and Dudek [8], we determined the asymptotic behavior of random variable $\mathrm{ex}(G(N, p), P_{n+1})$ as N and n go to infinity.

Theorem 1.20. *Let $3n \leq N \leq ne^{2n}$. The following hold a.a.s. as n approaches infinity. Let $\omega = \left(\log \frac{N}{n}\right)/(np)$.*

(i) For $p \geq \left(\log \frac{N}{n}\right)/(6n)$, we have $\mathrm{ex}(G(N, p), P_{n+1}) = \Theta(pnN)$.

(ii) For $N^{-1} \leq p \leq \left(\log \frac{N}{n}\right)/(6n)$, we have $\mathrm{ex}(G(N, p), P_{n+1}) = \Theta\left(\frac{\omega}{\log \omega} pnN\right)$.

Theorem 1.21. *Let $n \geq 2$ and $N \geq ne^{2n}$. The following hold a.a.s. as n approaches infinity or as N approaches infinity if n is a constant. Let $\omega = (\log N)/(np)$.*

(i) For $p \geq N^{-\frac{2}{5n}}$, we have $\mathrm{ex}(G(N,p), P_{n+1}) = \Theta(nN)$.

(ii) For $N^{-1} \leq p \leq N^{-\frac{2}{5n}}$, we have $\mathrm{ex}(G(N,p), P_{n+1}) = \Theta\left(\frac{\omega}{\log \omega} pnN\right)$.

The proofs of the above theorems are based on the probabilistic method, and an application of the depth first search algorithm (DFS) in finding long paths in random graphs. The details will be presented in Chapter 5.

Our work was also motivated by the size-Ramsey problem. The size-Ramsey number $\hat{R}(F,r)$ is the smallest integer m such that there exists a graph G on m edges with the property that any r-edge-coloring of G yields a monochromatic F. Krivelevich [79] and Dudek-Prałat [32] showed that $\Omega(r^2n) \leq \hat{R}(P_n, r) \leq O((\log r)r^2n)$. Determining whether $\hat{R}(P_n, r) = \Theta(r^2n)$ is perhaps the most interesting problem regarding the size-Ramsey number of a path. Both upper bound proofs give a stronger *density-type* result, which shows that for $p = \Omega((\log r)/n)$, every $H \subseteq G \in G(crn, p)$ with $e(H) \geq e(G)/r$ contains a P_{n+1}, for a constant c. Our results implies that $(\log r)/n$ is the threshold function for this density-type statement.

1.4 Basic definitions and notation

A *graph* G is a pair $(V(G), E(G))$ consisting of a set $V(G)$ of *vertices* along with a set $E(G)$ of *edges* which consists of 2-element subsets of $V(G)$; the pair of vertices in each edge are unordered. The *order* of a graph G is the cardinality of the vertex set $|V(G)|$ denoted here as $v(G)$. Similarly the *size* of a graph G is the cardinality of the edge set $|E(G)|$ denoted here as $e(G)$. Two vertices $u, v \in V(G)$ are said to be *adjacent*, denoted by $u \sim v$, uv, or vu, if $\{u, v\} \in E(G)$. An edge and a vertex on that edge are said to be incident. A graph with no loops (a loop is an edge $u \sim u$) or multiple edges (several edges $u \sim v$) is referred to in the literature as a *simple graph*. A graph allowing loops or multiple edges is referred to as a *multigraph*.

The *neighborhood* of a vertex v in a graph G, denoted by $N_G(v)$, is the set of vertices adjacent to v in G. The *degree* of a vertex v in a graph G, denoted by $d_G(v)$, is the number of edges incident to v in G. For a simple graph G, we always have $d_G(v) = |N_G(v)|$ for every vertex v. For a set $S \subseteq V(G)$, the *neighborhood of v restricted to S*, denoted by $N_G(v, S)$, is the number of vertices adjacent to v, which are contained in S ; the *degree of v restricted to S*, denoted by $d_G(v, S)$, is the number of edges incident to v with another endpoint in S. When the underlying graph is clear, we simply write $N(v)$, $d(v)$, $N(v, S)$ and $d(v, S)$ instead. The *minimum degree* of a graph G, denote by $\delta(G)$, is the degree of the vertex with the least number of edges incident to it. The *maximum degree* of a graph G, denote by $\delta(G)$, is the degree of the vertex with the most number of edges incident to it.

A graph H is a *subgraph* of a graph G, denoted $H \subseteq G$, if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. For a graph G and a set $A \subseteq V(G)$, the *induced subgraph* $G[A]$ is the subgraph of G whose vertex set is A and whose edge set consists of all of the edges with both endpoints in A. For two disjoint subsets $A, B \subseteq V(G)$, the *induced bipartite subgraph* $G[A, B]$ is the subgraph of G whose vertex set is $A \cup B$ and whose edge set consists of all of the edges with one endpoint in A and the other endpoint in B.

An *independent set* I in a graph G is a subset of $V(G)$ that forms no edges. A *edge coloring* of a graph G is an assignment of labels, traditionally called *colors*, to the edges of G. A k-coloring of a graph G is a coloring using k at most colors.

The *complete graph* on n vertices, denoted K_n, is the graph where every pair of distinct vertices is connected by exactly one edge. A *complete bipartite graph* is a bipartite graph such that every pair of graph vertices in the two parts of the partition are adjacent; this graph is denoted $K_{s,t}$ where s and t are the number of vertices in the two disjoint parts. A *path* on n vertices, denoted P_n, is a graph whose vertices can be linearly ordered so that two vertices are adjacent if and only if they appear consecutively in the ordering. A *cycle* on n vertices, denoted C_n, is a graph with n edges and there is a cyclic order of the vertices so that two vertices are adjacent if and only if they appear consecutively in this ordering.

A *r-uniform hypergraph*, or r-graph, H is a pair $(V(H), E(H))$ consisting of a set $V(H)$ of vertices along with a set $E(H)$ of hyperedges which consists of r-element subsets of $V(H)$. Similarly as for the graphs, the *order* of a hypergraph H is the cardinality of the vertex set $|V(H)|$ denoted here as $v(H)$, and the *size* of a hypergraph H is the cardinality of the edge set $|E(H)|$ denoted here as $e(H)$. As before, an *independent set* I in a hypergraph H is a subset of $V(H)$ that forms no edges.

For a positive integer n, we write $[n] = \{1, 2, \ldots, n\}$. Throughout the paper, we omit all floor and ceiling signs whenever these are not crucial.

Chapter 2

On the number of linear hypergraphs of large girth

2.1 Introduction

For a family of r-graphs $\mathcal{H}$, the *Turán number* (*function*) of $\mathcal{H}$, denoted by $\mathrm{ex}_r(n, \mathcal{H})$, is the maximum number of edges among all r-graphs on n vertices which contain no r-graph from $\mathcal{H}$ as a subgraph. Write $\mathrm{Forb}_r(n, \mathcal{H})$ for the set of r-graphs with vertex set $[n]$ which contain no r-graph from $\mathcal{H}$ as a subgraph. When $\mathcal{H}$ consists of a single graph H, we simply write $\mathrm{ex}_r(n, H)$ and $\mathrm{Forb}_r(n, H)$ instead. Since every subgraph of an H-free graph is also H-free, we have a trivial bound

$$2^{\mathrm{ex}(n,H)} \le |\mathrm{Forb}(n, H)| \le \sum_{i \le \mathrm{ex}(n,H)} \binom{\binom{n}{2}}{i} \le n^{2\cdot \mathrm{ex}(n,H)}. \tag{2.1}$$

The study on determination of $|\mathrm{Forb}_r(n, H)|$ has a very rich history. Recently, the case when H is a linear cycle received more attention. For integers $r \ge 2$ and $\ell \ge 3$, an r-uniform *linear cycle* of length ℓ, denoted by C_ℓ^r, is an r-graph with edges $e_1, \ldots, e_\ell$ such that for every $i \in [\ell - 1]$, $|e_i \cap e_{i+1}| = 1$, $|e_\ell \cap e_1| = 1$ and $e_i \cap e_j = \emptyset$ for all other pairs $\{i, j\}$, $i \ne j$. Kostochka, Mubayi and Verstraëte [76], and independently, Füredi and Jiang [47] proved that for every $r, \ell \ge 3$, $\mathrm{ex}_r(n, C_\ell^r) = \Theta(n^{r-1})$. Then by (2.1), we trivially have

$$|\mathrm{Forb}_r(n, C_\ell^r)| = 2^{\Omega(n^{r-1})} \text{ and } |\mathrm{Forb}_r(n, C_\ell^r)| = 2^{O(n^{r-1}\log n)} \tag{2.2}$$

for every $r, \ell \ge 3$. Guided and motivated by this development on the extremal numbers of linear cycles, Mubayi and Wang [87] showed that $|\mathrm{Forb}_3(n, C_\ell^3)| = 2^{O(n^2)}$ for all even ℓ and improved the trivial upper bound in (2.2) for $r > 3$. Inspired by Mubayi and Wang [87]'s method, Han and Kohayakawa [54] subsequently improved the general upper bound to $2^{O(n^{r-1}\log\log n)}$. Later, Balogh, Narayanan and Skokan [13] studied the balanced supersaturation phenomena of linear cycles, and proved $|\mathrm{Forb}_r(n, C_\ell^r)| = 2^{O(n^{r-1})}$ for every $r, \ell \ge 3$, using the hypergraph container method [11, 94].

In this chapter, we study the number of linear hypergraphs containing no linear cycle of fixed length.

An r-graph H is said to be *linear* if for every $e, e' \in E(H)$, $|e \cap e'| \leq 1$. For a family of linear r-graphs $\mathcal{H}$, the *linear Turán number* of $\mathcal{H}$, denoted by $\mathrm{ex}_L(n, \mathcal{H})$, is the maximum number of edges among linear r-graphs on n vertices which contain no r-graph from $\mathcal{H}$ as a subgraph. Write $\mathrm{Forb}_L(n, \mathcal{H})$ for the set of linear r-graphs with vertex set $[n]$ which contain no r-graph from $\mathcal{H}$ as a subgraph. Again, when $\mathcal{H}$ consists of a single graph H, we simply write $\mathrm{ex}_L(n, H)$ and $\mathrm{Forb}_L(n, H)$ instead. Similarly to (2.1), a trivial bound on the size of $\mathrm{Forb}_L(n, H)$ is given as follows.

$$2^{\mathrm{ex}_L(n,\mathcal{H})} \leq |\mathrm{Forb}_L(n, \mathcal{H})| \leq \sum_{i \leq \mathrm{ex}_L(n,\mathcal{H})} \binom{\binom{n}{r}}{i} \leq 2n^{r \cdot \mathrm{ex}_L(n,\mathcal{H})}. \tag{2.3}$$

It is known from the famous $(6,3)$-problem that $n^{2-c\sqrt{\log n}} < \mathrm{ex}_L(n, C_3^3) = o(n^2)$, where the lower bound is given by Behrend [17] and the upper bound is given by Ruzsa and Szemerédi [92]. In 1968, Erdős, Frankl and Rödl [40] showed that for every $r \geq 3$, $\mathrm{ex}_L(n, C_3^r) = o(n^2)$ and $\mathrm{ex}_L(n, C_3^r) = \Omega(n^c)$ for every $c < 2$. Using the so-called 2-fold Sidon sets, Lazebnik and Verstraëte [81] constructed linear 3-graphs with girth 5 and $\Omega(n^{3/2})$ edges. On the other hand, it is not hard to show that $\mathrm{ex}_L(n, C_4^3) = O(n^{3/2})$. Hence, $\mathrm{ex}_L(n, C_4^3) = \Theta(n^{3/2})$. Kostochka, Mubayi, and Verstraëte [75] proved $\mathrm{ex}_L(n, C_5^3) = \Theta(n^{3/2})$ and conjectured that

$$\mathrm{ex}_L(n, C_\ell^r) = \Theta\left(n^{1+\frac{1}{\lfloor \ell/2 \rfloor}}\right)$$

for every $r \geq 3$ and $\ell \geq 4$. Later, Collier-Cartaino, Graber and Jiang [27] proved that $\mathrm{ex}_L(n, C_\ell^r) = O\left(n^{1+\frac{1}{\lfloor \ell/2 \rfloor}}\right)$ for $r \geq 3$ and $\ell \geq 4$. Although the lower bound on the linear Turán number of linear cycles is still far from what is conjectured, following the same logic with the usual Turán problem of cycles, it is natural to conjecture that

$$|\mathrm{Forb}_L(n, C_\ell^r)| = 2^{\Theta\left(n^{1+\frac{1}{\lfloor \ell/2 \rfloor}}\right)} \tag{2.4}$$

for every $r \geq 3$ and $\ell \geq 4$. We first confirm the above conjecture for $\ell = 4$.

Theorem 2.1. *For every $r \geq 3$ there exists $C = C(r) > 0$ such that*

$$|\mathrm{Forb}_L(n, C_4^r)| \leq 2^{Cn^{3/2}}.$$

The upper bound for C_4^3 is sharp in order of magnitude given by $\mathrm{ex}_L(n, C_4^3) = \Theta(n^{3/2})$ and (2.3). In general, since the sharp bound of related linear Turán number remains open, we are not able to confirm the sharpness now.

For $\ell = 3$, the work of Erdős, Frankl and Rödl [40] could be extended to show that $|\mathrm{Forb}_L(n, C_3^r)| = 2^{o(n^2)}$

for every $r \geq 3$. For $\ell > 4$, although we are not ready to prove (2.4), we provide a result on the girth version. Recall that the girth of a graph is the length of a shortest cycle contained in the graph. Kleitman and Wilson [67], and independently Kreuter [77], and Kohayakawa, Kreuter, and Steger [69] proved that there are $2^{O(n^{1+1/\ell})}$ graphs with no even cycles of length 2ℓ, which made a step towards proving a longstanding conjecture of Erdős, who asked for determining the number of $C_{2\ell}$-free graphs. Motivated by the above work, we introduce an analogous girth problem on linear hypergraphs. For a linear r-graph H, the *girth* of H is the smallest integer k such that H contains a C_k^r. We remark that for linear r-graphs, our girth definition is equivalent to a more classical girth definition, *Berge girth*, i.e. the smallest number k such that the r-graph contains a Berge-C_k^r, as a linear Berge-C_k^r must contain a linear cycle of length i for some $3 \leq i \leq k$. For every $r \geq 3$ and $\ell \geq 4$, let $\mathrm{Forb}_L(n, r, \ell)$ denote the set of all linear r-graphs on $[n]$ with girth larger than ℓ. Our second main result of this chapter is as follows.

Theorem 2.2. *For every $r \geq 3$ and $\ell \geq 4$, there exists a constant $C = C(r, \ell) > 0$ such that*

$$|\mathrm{Forb}_L(n, r, \ell)| \leq 2^{Cn^{1+1/\lfloor \ell/2 \rfloor}}.$$

Palmer, Tait, Timmons and Wagner [89] considered such extremal problems for Berge-hypergraphs and proved a special case of Theorem 2.2 for $\ell = 4$. Note that for every $\ell \geq 4$, we have $\mathrm{Forb}_L(n, r, \ell + 1) \subseteq \mathrm{Forb}_L(n, r, \ell)$. Therefore, it is sufficient to prove Theorem 2.2 for all even ℓ and we provide the following equivalent theorem instead.

Theorem 2.3. *For every $r \geq 3$ and $\ell \geq 2$, there exists a constant $C = C(r, \ell) > 0$ such that*

$$|\mathrm{Forb}_L(n, r, 2\ell)| \leq 2^{Cn^{1+1/\ell}}.$$

Once again, the above upper bounds are possibly sharp, but we are not able to confirm it now.

The proofs of Theorems 2.1 and 2.3 are based on two graph enumeration results related to even cycles. A classical result of Bondy and Simonovits [23] yields $\mathrm{ex}_2(n, C_{2\ell}) = O(n^{1+1/\ell})$ for all $\ell \geq 2$. By a series of papers of Kleitman and Winston [68], Kleitman and Wilson [67], Kreuter [77], Kohayakawa, Kreuter, and Steger [69], and Morris and Saxton [86], we now know that the number of $C_{2\ell}$-free graphs is at most $2^{O(n^{1+1/\ell})}$. Inspired by these works, we prove that the number of graphs containing some but not many short cycles is still at most $2^{O(n^{1+1/\ell})}$, which may be of independent interest. We state our results as follows.

Theorem 2.4. *Let n be a sufficiently large integer and $a = 32\log^6 n$. The number of n-vertex graphs with at most n^2/a 4-cycles is at most $2^{11n^{3/2}}$.*

Given a graph G on $[n]$, for every integer $k \geq 3$ and every edge $uv \in E(G)$, denote by $c_k(u, v; G)$, the number of k-cycles in G containing edge uv. When the underlying graph is clear, we simply write $c_k(u, v)$. For an integer $\ell \geq 3$ and a constant $L > 0$, write $\mathcal{G}_n(\ell, L)$ for the family of graphs G on $[n]$ such that for every $3 \leq k \leq \ell$ and $uv \in E(G)$, $c_k(u, v; G) \leq L$.

Theorem 2.5. *For an integer $\ell \geq 3$ and a constant $L > 0$, let n be a sufficiently large integer and then we have*

$$|\mathcal{G}_n(2\ell, L)| \leq 2^{3(\ell+1)n^{1+1/\ell}}.$$

Like many of these advances, our approach to proving Theorems 2.4 and 2.5 relies on the graph container method developed in [68], in which one assigns a *certificate* for each target graph. The certificate should be able to uniquely determine the target graph, and then we can estimate the number of certificates instead of graphs. However, the previous applications of the graph container method address the problems for graphs forbidding short cycles, while we concern with the graphs with sparse short cycles. Therefore, the means by which we apply this technique is quite non-standard, and requires some new ideas.

Remark 2.6. *It is not hard to extend Theorem 2.4 to $a = \Theta(\log^5 n)$ by proving a similar statement for $\mathcal{G}_n(4, \sqrt{n}/\log^4 n)$ as in Theorem 2.5. We choose to present the current proof of Theorem 2.4 since it contains some ideas which may bring more insights of this method to readers. Let $p = \omega/(\sqrt{n}\log n)$. Note that the number of graphs on $[n]$ with $p\binom{n}{2}$ edges is about $2^{\omega n^{3/2}}$ and they typically contain $\Theta(n^4p^4) = \Theta(\omega^4 n^2/\log^4 n)$ 4-cycles. Therefore, $a = \Theta(\log^4 n)$ would be the best possible in Theorem 2.4 and we believe that it should be the truth. Given by the connection between Sidon sets and graphs without 4-cycles, this problem is closely related with the number of generalized Sidon sets, which will be studied in Chapter 3.*

2.2 Graphs with limited C_4's

2.2.1 Preliminary results

Definition 2.7 (Min-degree ordering, Min-degree sequence)**.** *For a graph G on $[n]$, a min-degree ordering is an ordering $v_n < v_{n-1} < \ldots < v_1$, such that v_i is a vertex of minimum degree in the graph $G_i = G[v_i, \ldots, v_1]$, for every $i \in [n]$ (if there are more than one vertices of the minimum degree, choose the one with the largest label). Let $d_i = d_{G_i}(v_i)$, then $d_n, d_{n-1}, \ldots, d_1$ is called the min-degree sequence.*

Lemma 2.8. *Let G be an n-vertex graph with average degree d. If $d \geq 2\sqrt{n}$, then G contains at least $d^4/36$ copies of 4-cycles.*

Proof. Let $v_1, v_2, \ldots, v_n$ be the vertices in G and $b_i = d_G(v_i)$ for every $i \in [n]$. Let S be the set of paths of length 2 (or 3-paths) in G. We will count 3-paths in two ways.

First, for a vertex v_i, the number of 3-paths containing v_i as the middle point is exactly $\binom{b_i}{2}$. Therefore, we have

$$|S| = \sum_{i=1}^{n} \binom{b_i}{2} \geq n \binom{(\sum_{i=1}^{n} b_i)/n}{2} = n\binom{d}{2} \geq \frac{1}{3} d^2 n.$$

On the other hand, for $1 \leq i < j \leq n$, let c_{ij} be the number of common neighbors of v_i and v_j. Then $|S| = \sum_{1\leq i<j\leq n} c_{ij}$. Therefore, the number of 4-cycles in G is equal to

$$\frac{1}{2} \sum_{1\leq i<j\leq n} \binom{c_{ij}}{2} \geq \frac{1}{2}\binom{n}{2}\binom{(\sum_{i<j} c_{ij})/\binom{n}{2}}{2} = \frac{1}{2}\binom{n}{2}\binom{|S|/\binom{n}{2}}{2} \geq \frac{|S|^2}{4n^2} \geq \frac{d^4}{36}.$$

□

From Lemma 2.8, we immediately obtain the following corollary.

Corollary 2.9. *Let G be a n-vertex graph which contains at most $4n^2/9$ 4-cycles, and $d_n, \ldots, d_1$ be the min-degree sequence of G. Then for every $i \in [n]$,*

$$d_i \leq 2\sqrt{n}.$$

Proof. Suppose that there exists $k \in [n]$, such that $d_k > 2\sqrt{n}$. Then by Lemma 2.8, the number of 4-cycles in G_k is at least $d_k^4/36 > \frac{4}{9}n^2$, which contradicts our assumption. □

We also provide an estimation for the following binomial coefficients, which will be used repeatly later.

Lemma 2.10. *For integers n, k, ℓ and a constant c satisfying $cn/k^\ell \geq k$,*

$$\binom{cn/k^\ell}{k} \leq 2^{\frac{\ell+1}{2^{1/\ln 2}\ln 2}(cen)^{\frac{1}{\ell+1}}},$$

where $2^{1/\ln 2}\ln 2 \approx 1.88$.

Proof. Let $f(x) = (\log cen - (\ell+1)\log x)\,x$ on $(0, +\infty)$. Since $f(x)$ is a concave function, it is maximized at the point x^*, where $f'(x^*) = \log cen - \frac{\ell+1}{\ln 2} - (\ell+1)\log x^* = 0$, i.e. $\log x^* = \frac{\log cen}{\ell+1} - \frac{1}{\ln 2}$. Therefore, we have

$$f(k) \leq f(x^*) = \left(\log cen - (\ell+1)\left(\frac{\log cen}{\ell+1} - \frac{1}{\ln 2}\right)\right) 2^{\frac{\log cen}{\ell+1} - \frac{1}{\ln 2}} = \frac{\ell+1}{2^{1/\ln 2}\ln 2}(cen)^{\frac{1}{\ell+1}}.$$

Since $\binom{n}{k} \le \left(\frac{ne}{k}\right)^k$ for every $1 \le k \le n$, we obtain that

$$\binom{cn/k^\ell}{k} \le \left(\frac{cen}{k^{\ell+1}}\right)^k = 2^{f(k)} \le 2^{\frac{\ell+1}{2^{1/\ln 2}\ln 2}(cen)^{\frac{1}{\ell+1}}}.$$

$\square$

2.2.2 Certificate lemma

This section is devoted to prove our main lemma, which is a key step to build the certificates for graphs with sparse 4-cycles. This lemma can be viewed as a generalization of the Kleitman-Winston algorithm [68], which builds certificates for graphs without 4-cycles. Before we proceed, we first need a counting lemma, which will be used later in the proof.

For a graph F, denote by F^2 the multigraph defined on $V(F)$ such that for every distinct $u, v \in V(F^2)$, the multiplicity of uv in F^2 is the number of (u,v)-paths of length 2 in F.

Lemma 2.11. *For integers $n > m \ge d \ge 8$, let F be an m-vertex graph with $\delta(F) \ge d-1$ and $H = F^2$. Then for every $J \subseteq V(H)$ of size at least $4n/d$, we have*

$$e(H[J]) \ge \frac{d^2|J|^2}{4n}.$$

Proof. Write $V(F) = \{v_1, \dots, v_m\}$. For every $j \in [m]$, let $b_j = d_F(v_j, J)$. Then we have $\sum_{j=1}^m b_j = \sum_{v\in J} d_F(v) \ge |J|(d-1) \ge \frac{4(d-1)}{d}n > 3n > 3m$. Therefore, we obtain that

$$e(H[J]) = \sum_{j=1}^m \binom{b_j}{2} \ge m\binom{\frac{\sum b_j}{m}}{2} \ge m\binom{\frac{|J|(d-1)}{m}}{2} \ge \frac{|J|^2(d-1)^2}{3m} \ge \frac{d^2|J|^2}{4n}.$$

$\square$

Lemma 2.12 (Certificate lemma)**.** *For a sufficiently large integer n, define $b = 16\log^4 n$ and $g = 32\log^5 n$. Let m and d be the integers satisfying $m \le n-1$ and $\frac{\sqrt{n}}{\log n} \le d \le 2\sqrt{n}$. Suppose that F is an m-vertex graph with $\delta(F) \ge d-1$ and $H = F^2$. Additionally, assume that for every $u, v \in V(F)$, $|N_F(u) \cap N_F(v)| \le \sqrt{n}/b$. Then for every set $I \subseteq V(F)$ of size d which satisfies $e(H[I]) \le n/g$, there exist a set T and a set $C(T)$ depending only on T, not on I, such that*

(i) $T \subseteq I \subseteq C(T)$,

(ii) $|T| \le 2\sqrt{n}/\log n$,

(iii) $|C(T)| \le 5n/d$.

Proof. Let I be a subset of $V(F)$ of size d which satisfies $e(H[I]) \le n/g$. Following the ideas of Kleitman and Winston [68], we describe a deterministic algorithm that associates to the set I a pair of sets T and $C(T)$, which shall be treated as the 'fingerprint' and the 'container' respectively.

Let $I_h = \{v \in I : d_H(v, I) > \sqrt{n}/b\}$ and $I_l = \{v \in I : d_H(v, I) \le \sqrt{n}/b\}$. Since $e(H[I]) \le n/g$, the size of I_h is at most

$$\frac{2e(H[I])}{\sqrt{n}/b} \le \frac{2\sqrt{n} \cdot b}{g} = \frac{\sqrt{n}}{\log n},$$

which is sufficiently small. Therefore, we only need to concern the vertices in I_l.

The core algorithm. We start the algorithm with sets $A_0 = V(H) - I_h$, $T_0 = \emptyset$ and the function $t_0(v) = 0$, for every $v \in V(H) - I_h$. As the algorithm proceeds, one should view A_i as the set of 'candidate' vertices, T_i as the set of 'representive' vertices, and $t_i(v)$ as a 'state' function which is used to control the process. In the i-th iteration step, we pick a vertex $u_i \in A_i$ of maximum degree in $H[A_i]$. In case there are multiple choices, we give preference to vertices that come earlier in some arbitrary predefined ordering of $V(H)$ as we always do, even if it is not pointed out at each time. If $u_i \in I_l$, we define

$$t_{i+1}(v) = \begin{cases} t_i(v) + d_H(v, u_i) & \text{if } v \in A_i, \\ t_i(v) & \text{if } v \notin A_i, \end{cases}$$

and $Q = \{v \mid t_{i+1}(v) > \sqrt{n}/b\}$, and let $T_{i+1} = T_i + u_i$, $A_{i+1} = A_i - u_i - Q$. Otherwise, let $T_{i+1} = T_i$, $A_{i+1} = A_i - u_i$ and $t_{i+1}(v) = t_i(v)$, for every $v \in V(H) - I_h$. The algorithm terminates at step K once we get a set A_K of size at most $4n/d$. We also assume that $u_{K-1} \in T_K$ as otherwise we can continue the algorithm until it is satisfied.

The algorithm outputs a vertex sequence $\{u_1, u_2, \ldots, u_{K-1}\}$, a set of 'representive' vertices T_K and a strictly decreasing set sequence $\{A_0, A_1, A_2, A_3, \ldots, A_K\}$. Let

$$T = T_K \cup I_h, \quad \text{and} \quad C(T) = A_K \cup T.$$

From the algorithm, we have $T_K \subseteq I_l$ and therefore $T \subseteq I$. Furthermore, if a vertex v satisfies $t_i(v) > \sqrt{n}/b$ for some i, then we have $d_H(v, I) \ge t_i(v) > \sqrt{n}/b$, which implies $v \notin I_l$. Therefore, we maintain $I_l \subseteq A_i \cup T_i$ for every $i \le K$ and in particular we have $I \subseteq A_K \cup T_K \cup I_h = A_K \cup T = C(T)$. Hence, Condition (i) is satisfied. Another crucial fact is that $C(T)$ depends only on T, not on I. The reason is that for a given underlying graph, its max degree sequence is fixed once we break the tie by some predefined ordering on vertices. Therefore, for two sets I_1, I_2 with the same 'fingerprint' T, the algorithm outputs the same vertex sequence $\{u_1, u_2, \ldots, u_{K-1}\}$, which uniquely determines the set $C(T)$ by the mechanics of the algorithm.

To verify Conditions (ii) and (iii), it is sufficient to show that $|T_K| \leq \sqrt{n}/\log n$. Once we prove it, we immediately obtain

$$|T| = |T_K| + |I_h| \leq \frac{\sqrt{n}}{\log n} + \frac{\sqrt{n}}{\log n} = \frac{2\sqrt{n}}{\log n},$$

and

$$|C(T)| = |A_K| + |T| \leq \frac{4n}{d} + \frac{2\sqrt{n}}{\log n} \leq \frac{5n}{d},$$

completing the proof.

Denote q the integer such that $n/2^q \leq |A_K| < n/2^{q-1}$. By the choice of A_K, we have $q < \log n$. For every integer $1 \leq l \leq q$, define A^l to be the first A-set satisfying

$$\frac{n}{2^l} \leq |A^l| < \frac{n}{2^{l-1}},$$

if it exists, and let T^l be the corresponding T-set and $t^l(v)$ be the corresponding t-function of A^l . Note that A^l may not exist for every l, but A^q always exists and it could be that $A^q = A_K$. Suppose that

$$A^{l_1} \supset A^{l_2} \supset \ldots \supset A^{l_p}$$

are all the well-defined A^l, where $p \leq q$. By the above definition, we have $A^{l_1} = A_0$, $T^{l_1} = T_0$ and $l_p = q$. Define $A^{l_{p+1}} = A_K, T^{l_{p+1}} = T_K$. Now, we have

$$T_K = \bigcup_{j=2}^{p+1} (T^{l_j} - T^{l_{j-1}}). \tag{2.5}$$

To achieve our goal, we are going to estimate the size of $T^{l_j} - T^{l_{j-1}}$ for every $2 \leq j \leq p+1$.

From the algorithm, we have $t^{l_j}(v) \leq \sqrt{n}/b$, for every $v \in A^{l_j} \cup T^{l_j}$. Moreover, for $v \in A^{l_{j-1}} - A^{l_j} - T^{l_j}$, suppose that v is removed in step i, then we have

$$t^{l_j}(v) \leq t_{i-1}(v) + d_H(v, u_i) \leq \frac{\sqrt{n}}{b} + |N_F(u_i) \cap N_F(v)| \leq \frac{2\sqrt{n}}{b},$$

where u_i is the selected vertex in step i. Therefore, we obtain

$$\sum_{v \in A^{l_{j-1}}} t^{l_j}(v) \leq \frac{2\sqrt{n}}{b} |A^{l_{j-1}}| \leq \frac{2n^{3/2}}{2^{l_{j-1}-1} b}. \tag{2.6}$$

Let $2 \leq j \leq p$. For every $u_i \in T^{l_j} - T^{l_{j-1}}$, u_i is chosen of maximum degree in $H[A_i]$, where A_i is a set between $A^{l_{j-1}}$ and A^{l_j}. By the choice of A^{l_j}, we have $|A_i| \geq n/2^{l_{j-1}}$. By Lemma 2.11, we have

$$d_H(u_i, A_i) \geq \frac{d^2|A_i|}{4n} \geq \frac{d^2}{2^{l_{j-1}+2}}.$$

Note that $d_H(u_i, A_i)$ only contributes to $t^{l_j}(v)$ for $v \in A_i \subseteq A^{l_{j-1}}$. Then we obtain

$$\left|T^{l_j} - T^{l_{j-1}}\right| \frac{d^2}{2^{l_{j-1}+2}} \leq \sum_{u_i \in T^{l_j} - T^{l_{j-1}}} d_H(u_i, A_i) \leq \sum_{v \in A^{l_{j-1}}} t^{l_j}(v). \tag{2.7}$$

Combining (2.6) and (2.7), we have

$$\left|T^{l_j} - T^{l_{j-1}}\right| \frac{d^2}{2^{l_{j-1}+2}} \leq \frac{2n^{3/2}}{2^{l_{j-1}-1} b},$$

which implies

$$\left|T^{l_j} - T^{l_{j-1}}\right| \leq \frac{16n^{3/2}}{bd^2} \leq \frac{16\sqrt{n}\log^2 n}{b} = \frac{\sqrt{n}}{\log^2 n}$$

for $2 \leq j \leq p$. For $j = p+1$, since we have $\frac{n}{2^q} \leq |A^{l_{p+1}}| \leq |A^{l_p}| \leq \frac{n}{2^{q-1}}$, by a similar argument, we obtain that

$$\left|T^{l_{p+1}} - T^{l_p}\right| \frac{d^2|A^{l_{p+1}}|}{4n} \leq \sum_{u_i \in T^{l_{p+1}} - T^{l_p}} d(u_i, A_i) \leq \sum_{v \in A^{l_p}} t^{l_{p+1}}(v) \leq \frac{2\sqrt{n}}{b}|A^{l_p}|,$$

which gives

$$\left|T^{l_{p+1}} - T^{l_p}\right| \leq \frac{16n^{3/2}}{bd^2} \leq \frac{16\sqrt{n}\log^2 n}{b} = \frac{\sqrt{n}}{\log^2 n}.$$

Finally, by (2.5), we get

$$|T_K| = \bigcup_{j=2}^{p+1} |T^{l_j} - T^{l_{j-1}}| \leq p \cdot \frac{\sqrt{n}}{\log^2 n} \leq q \cdot \frac{\sqrt{n}}{\log^2 n} \leq \frac{\sqrt{n}}{\log n}.$$

$\square$

2.2.3 Proof of Theorem 2.4

In this section, we give an upper bound on the number of graphs containing only 'few' 4-cycles. Before we proceed to prove Theorem 2.4, we need to do a cleaning process for the target graphs in order to apply Lemma 2.12.

Let $a = 32\log^6 n$, $g = 32\log^5 n$ and $b = 16\log^4 n$. Given a graph G on $[n]$, for every $1 \le i < j \le n$, define $N_G(i,j)$ to be the set of common neighbors of i and j in G. Let

$$m_G(i,j) = \begin{cases} |N_G(i,j)| & \text{when } |N_G(i,j)| > \frac{\sqrt{n}}{b}, \\ 0 & \text{when } |N_G(i,j)| \le \frac{\sqrt{n}}{b}. \end{cases}$$

We delete all edges from i to $N_G(i,j)$, for all $1 \le i < j \le n$ with $m_G(i,j) \ne 0$. Then the resulting subgraph, denoted by $\widehat{G}$, satisfies $|N_{\widehat{G}}(i,j)| \le \sqrt{n}/b$, for every $1 \le i < j \le n$. Let $\mathcal{G}_n$ be the family of graphs on $[n]$ with at most n^2/a 4-cycles and $\widehat{\mathcal{G}}_n = \{\widehat{G} : G \in \mathcal{G}_n\}$.

Lemma 2.13. *Let n be a sufficiently large integer. Then for every $G \in \mathcal{G}_n$, we have*

$$|E(G) - E(\widehat{G})| \le \frac{4n^{3/2}}{\log^2 n}.$$

Proof. By counting 4-cycles in G, we obtain that

$$\frac{1}{2}\sum_{i<j}\binom{m_G(i,j)}{2} \le \frac{n^2}{a},$$

which gives

$$\sum_{i<j} m_G(i,j)^2 \le 8\frac{n^2}{a}. \tag{2.8}$$

Let $B = \{(i,j) : 1 \le i < j \le n \text{ and } m_G(i,j) \ne 0\}$. By the definition of $m_G(i,j)$ and (2.8), we have $|B| \le 8\frac{n^2}{a}/(\frac{\sqrt{n}}{b})^2 = 8b^2n/a$. Therefore, by the convexity, we get

$$\sum_{(i,j)\in B} m_G(i,j)^2 \ge \frac{(\sum_{(i,j)\in B} m_G(i,j))^2}{|B|} = \frac{(\sum_{i<j} m_G(i,j))^2}{|B|} \ge \frac{(\sum_{i<j} m_G(i,j))^2}{8b^2n/a}. \tag{2.9}$$

Combining (2.8) and (2.9), we obtain

$$\sum_{i<j} m_G(i,j) \le \frac{8n^{3/2}b}{a} = \frac{4n^{3/2}}{\log^2 n}.$$

Finally, by the definition of $\widehat{G}$, we have $|E(G) - E(\widehat{G})| = \sum_{i<j} m_G(i,j) \leq \frac{4n^{3/2}}{\log^2 n}$. $\square$

Lemma 2.14. *Let n be a sufficiently large integer. Then $|\mathcal{G}_n| \leq |\widehat{\mathcal{G}}_n| \cdot 2^{\frac{4n^{3/2}}{\log n}}$.*

Proof. For every $F \in \widehat{\mathcal{G}}_n$, let $\mathcal{S}_F = \{G \in \mathcal{G}_n \mid \widehat{G} = F\}$. By Lemma 2.13, for every $G \in \mathcal{S}_F$, we have $|E(G) - E(F)| \leq \frac{4n^{3/2}}{\log^2 n}$. Therefore, the size of $\mathcal{S}_F$ is bounded by

$$|\mathcal{S}_F| \leq \binom{\binom{n}{2}}{0} + \binom{\binom{n}{2}}{1} + \ldots + \binom{\binom{n}{2}}{\lfloor \frac{4n^{3/2}}{\log^2 n} \rfloor} \leq 2\binom{\binom{n}{2}}{\lfloor \frac{4n^{3/2}}{\log^2 n} \rfloor} \leq 2^{\frac{4n^{3/2}}{\log n}}.$$

Finally, we obtain that

$$|\mathcal{G}_n| \leq \sum_{F \in \widehat{\mathcal{G}}_n} |\mathcal{S}_F| \leq |\widehat{\mathcal{G}}_n| \cdot 2^{\frac{4n^{3/2}}{\log n}}.$$

$\square$

Theorem 2.15. *Let n be a sufficiently large integer. Then $|\widehat{\mathcal{G}}_n| \leq 2^{10n^{3/2}}$.*

Proof. We construct the certificate of a graph G in the following way. Let $Y_G := v_n < v_{n-1} < \ldots < v_1$ be the min-degree ordering of G and $D_G := \{d_n, d_{n-1}, \ldots, d_1\}$ be the min-degree sequence of G. Let $G_i = G[v_i, \ldots, v_1]$, for every $i \in [n]$. Define the set sequence $S_G := \{S_n, S_{n-1}, \ldots, S_2\}$, where $S_i = N_G(v_i, G_{i-1})$. Then $S_i \subseteq \{v_{i-1}, \ldots, v_1\}$, and $|S_i| = d_i$. By the construction, $[Y_G, D_G, S_G]$ uniquely determines the graph G and so we build a certificate $[Y_G, D_G, S_G]$ for G. Therefore, instead of counting graphs, it is equivalent to estimate the number of their certificates.

For a graph $G \in \widehat{\mathcal{G}}_n$, its certificate has some important properties which would help us to achieve the desired bound. First, by Corollary 2.9, its min-degree ordering $D_G = \{d_n, d_{n-1}, \ldots, d_1\}$ satisfying $d_i \leq 2\sqrt{n}$. Let f_i be the number of 4-cycles in G_i containing vertex v_i. Since each 4-cycle contributes exactly to one of f_i's, we have $\sum_{i=1}^n f_i \leq n^2/a$. We call v_i a *heavy* vertex if $f_i > n/g$; otherwise, v_i is a *light* vertex. Another crucial fact about graphs in $\widehat{\mathcal{G}}_n$ is that the number of heavy vertices is at most

$$\frac{\sum_{v_i \in V_h} f_i}{n/g} \leq \frac{n^2/a}{n/g} = \frac{n}{\log n}. \tag{2.10}$$

Now we start to estimate the number of certificates which would generate graphs in $\widehat{\mathcal{G}}_n$. By the above discussion, we first observe that the number of ways to choose the min-degree orderings and the min-degree sequences is at most

$$n!(2\sqrt{n})^n. \tag{2.11}$$

Then we fix a min-degree ordering $Y^* = v_n < v_{n-1} < \ldots < v_1$, and a min-degree sequence $D^* = \{d_n, \ldots, d_1\}$.

Next, we fix the positions of heavy vertices and by (2.10) the number of ways is at most

$$\sum_{i \le \frac{n}{\log n}} \binom{n}{i}. \tag{2.12}$$

A major part of the proof is to count set sequences $S = \{S_n, S_{n-1}, \ldots, S_2\}$, where $S_i \subseteq \{v_{i-1}, \ldots, v_1\}$ and $|S_i| = d_i$, such that the graph reconstructed by $[Y^*, D^*, S]$, denoted by G_S, are in $\widehat{\mathcal{G}}$. For every $2 \le i \le n$, let M_i be the number of choices for S_i with fixed sets $S_{i-1}, \ldots, S_2$. Define

$$\mathcal{I}_1 = \{i : v_i \text{ is a heavy vertex}\}, \quad \mathcal{I}_2 = \{i : d_i < \frac{\sqrt{n}}{\log n}\},$$

and

$$\mathcal{I}_3 = \{i : v_i \text{ is a light vertex and } d_i \ge \frac{\sqrt{n}}{\log n}\}.$$

For every $i \in \mathcal{I}_1$, since $|S_i| = d_i \le 2\sqrt{n}$, we have a trivial upper bound

$$M_i \le \binom{i-1}{d_i} \le \binom{n}{2\sqrt{n}} \le n^{2\sqrt{n}} = 2^{2\sqrt{n}\log n}. \tag{2.13}$$

Similarly, for every $i \in \mathcal{I}_2$, we have

$$M_i \le \binom{i-1}{d_i} \le \binom{n}{\sqrt{n}/\log n} \le n^{\sqrt{n}/\log n} = 2^{\sqrt{n}}. \tag{2.14}$$

It remains to estimate M_i for $i \in \mathcal{I}_3$. With fixed sets $S_{i-1}, \ldots, S_2$, the graph $G_{i-1} = G_S[v_{i-1}, \ldots, v_1]$ is uniquely determined. Since $G_{i-1} \subseteq G_S$ and $G_S \in \widehat{\mathcal{G}}$, for every $u, v \in V(G_{i-1})$, we have $|N_{G_{i-1}}(u) \cap N_{G_{i-1}}(v)| \le \sqrt{n}/b$. Applying Lemma 2.12 on G_{i-1}, we obtain that every eligible S_i contains a subset T of size at most $2\sqrt{n}/\log n$, which determines a set $C(T) \supseteq S_i$ of size at most $5n/d_i$. Since the number of choices for T is at most

$$\sum_{0 \le j \le 2\sqrt{n}/\log n} \binom{i-1}{j} \le 2\binom{i-1}{2\sqrt{n}/\log n} \le 2\binom{n}{2\sqrt{n}/\log n} \le 2^{2\sqrt{n}},$$

we then have

$$M_i \le \sum_T \binom{C(T)}{d_i} \le \sum_T \binom{5n/d_i}{d_i} \le \sum_T 2^{\frac{2}{2^{1/\ln 2}\ln 2}\sqrt{5en}} \le \sum_T 2^{4\sqrt{n}} \le 2^{6\sqrt{n}} \tag{2.15}$$

for every $i \in \mathcal{I}_3$, where the third inequality is given by Lemma 2.10.

Combining (2.13), (2.14) and (2.15), we obtain that the number of choices for S is

$$\prod_{i=2}^{n} M_i \leq \prod_{i\in\mathcal{I}_1} M_i \prod_{i\in\mathcal{I}_2} M_i \prod_{i\in\mathcal{I}_3} M_i \leq (2^{2\sqrt{n}\log n})^{\frac{n}{\log n}} (2^{\sqrt{n}})^n (2^{6\sqrt{n}})^n \leq 2^{9n^{3/2}}.$$

Finally, together with (2.11) and (2.12), the total number of certificates is at most

$$n!(2\sqrt{n})^n \sum_{i\leq\frac{n}{\log n}} \binom{n}{i} \prod_{i=2}^{n} M_i \leq n!(2\sqrt{n})^n 2^n 2^{9n^{3/2}} \leq 2^{10n^{3/2}},$$

which leads to $|\widehat{\mathcal{G}}_n| \leq 2^{10n^{3/2}}$. □

Proof of Theorem 2.4. Lemma 2.14 and Theorem 2.15 imply Theorem 2.4. □

2.3 Graphs with limited short cycles

In the previous section, we estimated the number of graphs containing a few 4-cycles. Unfortunately, we are not ready to provide a similar result for longer cycles due to the failure of getting an appropriate counting lemma, like Lemma 2.11. However, this method still works when the target graph has a sparse structure on short cycles. More specially, for $\ell \geq 4$, we are going to consider the family of graphs such that each of its edges is contained in only $O(1)$ cycles of length at most 2ℓ. Following the idea from [69], we construct a proper auxiliary graph and provide a suitable counting lemma on it.

2.3.1 Expansion properties of graphs with limited short cycles

Given a graph G, a vertex $v \in V(G)$ and an integer $k \geq 1$, let $\Gamma_k(v)$ be the set of vertices of G at distance exactly k from v. Recall that for an edge $uv \in E(G)$, $c_k(u, v; G)$ is the number of k-cycles in G containing edge uv.

Lemma 2.16. *For integers $\ell \leq m$ and a constant $L > 0$, let F be an m-vertex graph such that for every $uv \in E(F)$ and $3 \leq i \leq 2\ell$, $c_i(u, v) \leq L$. Then for every $1 \leq k \leq \ell - 1$ and $v \in V(F)$, we have*

$$d(u, \Gamma_k(v)) \leq Lk$$

for all $u \in \Gamma_k(v)$.

Proof. Suppose there exists a vertex $u \in \Gamma_k(v)$ such that $d(u, \Gamma_k(v)) \geq Lk + 1$. Since $u \in \Gamma_k(v)$, there exists a (u, v)-path P_u of length k. Let u' be the neighbor of u in P_u. Similarly, for every vertex $w \in N(u, \Gamma_k(v))$, there is a (w,v)-path P_w of length k. Note that every $P_u + P_w + \{uw\}$ forms a closed walk of length $2k+1$, which contains an odd cycle of length at most $2k+1$ containing edges uu' and uw. Since $d(u,\Gamma_k(v)) \geq Lk+1$, we have at least $Lk+1$ distinct odd cycles of length at most $2k+1$ containing uu'. However, since $c_h(u,u') \leq L$ for every odd $h \leq 2k+1$, there are at most Lk odd cycles of length at most $2k+1$ containing uu', which is a contradiction. $\square$

Lemma 2.17. *For integers $\ell \leq m$ and a constant $L > 0$, let F be an m-vertex graph such that for every $uv \in E(F)$ and $3 \leq i \leq 2\ell$, $c_i(u,v) \leq L$. Then for every $2 \leq k \leq \ell$ and $v \in V(F)$, we have*

$$d(u,\Gamma_{k-1}(v)) \leq L(k-1)+1$$

for all $u \in \Gamma_k(v)$.

Proof. Suppose there exists a vertex $u \in \Gamma_k(v)$ such that $d(u,\Gamma_{k-1}(v)) \geq L(k-1)+2$. Let u' be a vertex in $N(u,\Gamma_{k-1}(v))$. Since $u' \in \Gamma_{k-1}(v)$, there exists a (u',v)-path $P_{u'}$ of length $k-1$. Similarly, for every vertex $w \in N(u,\Gamma_{k-1}(v)) \setminus \{u'\}$, there is a (w,v)-path P_w of length $k-1$. Note that every $P_{u'} + P_w + \{uu'\} + \{uw\}$ forms a closed walk of length $2k$, which contains an even cycle of length at most $2k$ containing edges uu' and uw. Since $|N(u,\Gamma_{k-1}(v)) \setminus \{u'\}| \geq L(k-1)+1$, we have at least $L(k-1)+1$ distinct even cycles of length at most $2k$ containing uu'. However, since $c_h(u,u') \leq L$ for every even $4 \leq h \leq 2k$, there are at most $L(k-1)$ even cycles of length at most $2k$ containing uu', which is a contradiction. $\square$

Now, we give a lemma on the expansion of graphs with sparse short cycles. This lemma can be viewed as a generalization of Lemma 11 in [69].

Lemma 2.18. *For integers $\ell, d \leq m$ and a constant $L \ll d$, let F be an m-vertex graph with minimum degree at least $d-1$, such that for every $uv \in E(F)$ and $3 \leq i \leq 2\ell$, $c_i(u,v) \leq L$. Suppose v is a vertex in F with degree $d(v)$. Then for every $1 \leq k \leq \ell$, we have*

$$|\Gamma_k(v)| \geq \frac{d(v)d^{k-1}}{g_k(L)}$$

for some constants $g_k(L)$ which only depend on k and L.

Proof. The case $k=1$ is trivially true with $g_1(L)=1$. Suppose that the lemma is true for $k<\ell$, i.e. $|\Gamma_k(L)| \geq d(v)d^{k-1}/g_k(L)$ for some constant $g_k(L)$.

For every vertex $u \in \Gamma_k(v)$, neighbors of u only appear in $\Gamma_{k-1}(v)$, $\Gamma_k(v)$ and $\Gamma_{k+1}(v)$. By Lemmas 2.16

and 2.17, we have

$$d(u, \Gamma_{k+1}(v)) \geq (d-1) - d(u, \Gamma_{k-1}(v)) - d(u, \Gamma_k(v)) \geq d - 2(Lk+1) + L \geq \frac{d}{2} \tag{2.16}$$

for all $u \in \Gamma_k(v)$ and this gives

$$e(\Gamma_k(v), \Gamma_{k+1}(v)) \geq \frac{d|\Gamma_k(v)|}{2}.$$

Again by Lemma 2.17, we know that for every $u \in \Gamma_{k+1}(v)$, $d(u, \Gamma_k(v)) \leq Lk+1$. Therefore, we have

$$|\Gamma_{k+1}(v)| \geq \frac{e(\Gamma_k(v), \Gamma_{k+1}(v))}{Lk+1} \geq \frac{d|\Gamma_k(v)|}{2(Lk+1)} \geq \frac{d(v)d^k}{2(Lk+1)g_k(L)} = \frac{d(v)d^k}{g_{k+1}(L)}$$

for $g_{k+1}(L) = 2(Lk+1)g_k(L)$ and the lemma follows by induction. $\square$

Lemma 2.18 gives an upper bound on the maximum degree of the graph with sparse short cycles.

Corollary 2.19. *For integers $\ell, d \leq m$ and a constant $L \ll d$, let F be an m-vertex graph with minimum degree $d-1$, such that for every $uv \in E(F)$ and $3 \leq i \leq 2\ell$, $c_i(u,v) \leq L$. Then*

$$\Delta(F) \leq \frac{m}{d^{\ell-1}} \cdot g_\ell(L),$$

where $g_\ell(L)$ is the constant defined in Lemma 2.18.

Proof. By Lemma 2.18, for every $v \in V(F)$, we have

$$|\Gamma_\ell(v)| \geq \frac{d(v)d^{\ell-1}}{g_\ell(L)},$$

which gives

$$d(v) \leq \frac{|\Gamma_\ell(v)|}{d^{\ell-1}} g_\ell(L) \leq \frac{m}{d^{\ell-1}} g_\ell(L),$$

This implies the corollary. $\square$

2.3.2 Construction of the auxiliary graph

In this section, we aim to give a generalization of Lemma 2.11 for longer cycles. We use a definition of *composed walk* from [69]. For every integer $k \geq 1$, call a $2k$-walk $x_0x_1 \ldots x_{2k}$ a *composed walk* if $x_0 \ldots x_k$ and $x_k \ldots x_{2k}$ are two shortest paths and they are different but not necessarily vertex-disjoint or edge-disjoint. A composed walk is said to be *closed* if its endpoints are the same.

Lemma 2.20. *For integers $\ell, \Delta \leq m$ and a constant $L \ll \Delta$, let F be an m-vertex graph with maximum degree Δ, such that for every $uv \in E(F)$ and $3 \leq k \leq 2\ell$, $c_k(u,v) \leq L$. Then for every vertex $u \in V(F)$ and every integer $2 \leq s \leq \ell - 1$, the number of closed composed walks of length $2s$ with endpoints u is at most*

$$\Delta^{s-1}\alpha_s(L)$$

for some constants $\alpha_s(L)$ which only depends on s and L.

Proof. For every vertex $u \in V(F)$ and every integer $2 \leq s \leq \ell - 1$, let $\mathcal{W}_s(u)$ be the set of closed composed walks of length $2s$ with endpoints u. For the case $s = 2$, the lemma is true with $\alpha_2(L) = L$. This is because that a closed composed walk of length 4 with endpoint u is exactly a 4-cycle containing u and then we have $|\mathcal{W}_2(u)| \leq \sum_{v \in N(u)} c_4(u,v) \leq \Delta L$.

Suppose for $s - 1 < \ell - 1$, the lemma is true for all integers $k \leq s - 1$, i.e. for every $v \in V(F)$, $|\mathcal{W}_k(v)| \leq \Delta^{k-1}\alpha_k(L)$ with some constants $\alpha_k(L)$. Fix an arbitrary vertex $u \in V(F)$, and let

$$\mathcal{W}_s^i(u) = \{ux_1x_2 \ldots x_{2s-1}u \in \mathcal{W}_s(u) \mid i \text{ is the first integer such that } x_i = x_{2s-i}\}$$

for every $1 \leq i \leq s$. Then we have $\mathcal{W}_s(u) = \bigcup_{i=1}^{s} \mathcal{W}_s^i(u)$.

First, every composed walk $W \in \mathcal{W}_s^1(u)$ consists of an edge ux_1 and a closed composed walk of length $2s - 2$ with endpoints x_1. Therefore, we have

$$|\mathcal{W}_s^1(u)| \leq \sum_{x_i \in N(u)} |\mathcal{W}_{s-1}(x_1)| \leq \Delta^{s-1}\alpha_{s-1}(L).$$

Let $2 \leq i \leq s - 1$. For every composed walk

$$W = ux_1x_2 \ldots x_{2s-1}u \in \mathcal{W}_s^i(u),$$

$\{ux_1 \ldots x_i \; x_{2s-(i-1)} \ldots x_{2s-1}u\}$ forms a cycle C of length $2i$ containing u. Since for every $x_1 \in N(u)$, $c_{2i}(u, x_1) \leq L$, then the number of choices for C is at most ΔL. For a fixed C and $x_i \in C$, $W - C$ forms a path of length $(s - i)$ with endpoints x_i or a closed composed walks of length $2(s - i)$ with endpoints x_i. In the first case there are at most Δ^{s-i} choices, while in the later case there are at most $|\mathcal{W}_{s-i}(x_i)|$ choices.

Therefore, we have

$$\begin{aligned}|\mathcal{W}_s^i(u)| &\le \Delta L \cdot (\Delta^{s-i} + |\mathcal{W}_{s-i}(x_i)|)\\ &\le \Delta L \cdot (\Delta^{s-i} + \Delta^{s-i-1}\alpha_{s-i}(L))\\ &\le 2\Delta^{s-i+1}L \le 2\Delta^{s-1}L.\end{aligned}$$

Finally, every composed walk $W \in \mathcal{W}_s^s(u)$ is a cycle of length $2s$ containing u, and then we have

$$|\mathcal{W}_s^s(u)| \le \sum_{v\in N(u)} c_{2s}(u,v) \le \Delta L.$$

Hence, we have

$$|\mathcal{W}_s(u)| = \bigcup_{i=1}^{s} |\mathcal{W}_s^i(u)| \le \Delta^{s-1}\alpha_{s-1}(L) + 2(s-2)\Delta^{s-1}L + \Delta L \le \Delta^{s-1}\alpha_s(L)$$

for $\alpha_s(L) = \alpha_{s-1}(L) + 2(s-2)L + 1$, and the lemma follows by induction. □

For an integer $\ell \ge 3$ and a graph F, denote by F^ℓ the multigraph defined on $V(F)$ such that for every distinct $u,v \in V(F^\ell)$, the multiplicity of uv in F^ℓ is the number of composed (u,v)-walks of length $2(\ell-1)$ in F.

Lemma 2.21. *For an integer $\ell \ge 3$ and a constant $L > 0$, let n be a sufficiently large integer. Let m and d be the integers satisfying $m \le n$ and $d \ge \frac{n^{1/\ell}}{\log n}$. Suppose F is an m-vertex graph with minimum degree $d-1$, such that for every $uv \in E(F)$ and $3 \le k \le 2\ell$, $c_k(u,v) \le L$. Then for every set $J \subseteq V(F)$ of size at least $2^\ell n/d^{\ell-1}$, we have*

$$e(F^\ell[J]) \ge \frac{d^{2\ell-2}|J|^2}{2^{2\ell+1}n}.$$

Proof. Let $\mathcal{W}$ be the set of composed walks of length $2(\ell-1)$ with endpoints in J, and $\mathcal{W}_c$ be the set of closed composed walks of length $2(\ell-1)$ with endpoint in J. By the definition of F^ℓ, we have

$$e(F^\ell[J]) = |\mathcal{W}| - |\mathcal{W}_c|.$$

By Lemma 2.20, we know that

$$|\mathcal{W}_c| \le \Delta^{\ell-2}\alpha_{\ell-1}(L)\cdot |J|,$$

where Δ is the maximum degree of F, which, by Corollary 2.19, satisfies

$$\Delta \le \frac{m}{d^{\ell-1}} \cdot g_\ell(L) \le \frac{n}{d^{\ell-1}} \cdot g_\ell(L) \le \frac{d^\ell \log^\ell n}{d^{\ell-1}} \cdot g_\ell(L) = d \log^\ell n \cdot g_\ell(L). \tag{2.17}$$

Now, it remains to estimate the lower bound of $\mathcal{W}$. For every $v \in J$, let a_v be the number of shortest paths of length $\ell - 1$ such that v is one of the endpoints. For every $u \in V(F)$, let $\mathcal{P}_u$ be the set of shortest paths of length $\ell - 1$ such that one endpoint is u and another endpoint is in J. Let $b_u = |\mathcal{P}_u|$ and then we have $\sum_{u\in V(F)} b_u = \sum_{v\in J} a_v$. By (2.16), we have

$$\sum_{u\in V(F)} b_u = \sum_{v\in J} a_v \ge \sum_{v\in J} (d/2)^{\ell-1} = \frac{d^{\ell-1}|J|}{2^{\ell-1}}.$$

Note that for every vertex $u \in V(F)$ and $P_1, P_2 \in \mathcal{P}_u$, $P_1 + P_2$ forms a composed walk in $\mathcal{W}$ and vice versa. Therefore, we have

$$|\mathcal{W}| = \sum_{u\in V(F)} \binom{b_u}{2} \ge m\binom{\frac{\sum_u b_u}{m}}{2} \ge m\binom{\frac{d^{\ell-1}|J|}{2^{\ell-1}\cdot m}}{2} \ge \frac{d^{2\ell-2}|J|^2}{2^{2\ell}m} \ge \frac{d^{2\ell-2}|J|^2}{2^{2\ell}n}$$

for $|J| \ge 2^\ell n/d^{\ell-1} \ge 2^\ell m/d^{\ell-1}$. Note that

$$\begin{aligned}
\frac{|\mathcal{W}_c|}{|\mathcal{W}|} &\le \frac{\Delta^{\ell-2}\alpha_{\ell-1}(L)\cdot|J|}{\frac{d^{2\ell-2}|J|^2}{2^{2\ell}n}} \le \frac{d^{\ell-2}\log^{\ell(\ell-2)} n}{d^{2\ell-2}} \cdot \frac{n}{|J|} \cdot g_\ell^{\ell-2}(L)\alpha_{\ell-1}(L)2^{2\ell} \\
&\le \frac{d^{\ell-2}\log^{\ell(\ell-2)} n}{d^{2\ell-2}} \cdot \frac{d^{\ell-1}}{2^\ell} \cdot g_\ell^{\ell-2}(L)\alpha_{\ell-1}(L)2^{2\ell} \\
&\le \frac{\log^{\ell(\ell-2)} n}{d} \cdot g_\ell^{\ell-2}(L)\alpha_{\ell-1}(L)2^{\ell} \ll 1,
\end{aligned}$$

when n is sufficiently large. Hence, we have

$$e(F^\ell[J]) = |\mathcal{W}| - |\mathcal{W}_c| \ge \frac{1}{2}|\mathcal{W}| \ge \frac{d^{2\ell-2}|J|^2}{2^{2\ell+1}n}.$$

□

Now, we start to define the auxiliary graph, which will be used in Lemma 2.25 in the next section. For every integer $k \ge 1$, call a path $x_0x_1\dots x_{2k}$ a *composed path* if $x_0\dots x_k$ and $x_k\dots x_{2k}$ are both shortest paths of length k. For an integer $\ell \ge 3$ and a graph F, denote by F_*^ℓ the simple graph defined on $V(F)$ such that for every distinct $u, v \in V(F_*^\ell)$, $uv \in E(F_*^\ell)$ if there is a composed (u, v)-path of length at most $2(\ell - 1)$ in F. To estimate the number of edges in F_*^ℓ, we need the following lemma.

Lemma 2.22. *For integers $\ell, \Delta \leq m$ and a constant $L \ll \Delta$, let F be an m-vertex graph with maximum degree Δ, such that for every $uv \in E(F)$ and $3 \leq k \leq 2\ell$, $c_k(u,v) \leq L$. For every $1 \leq s \leq \ell - 1$ and every distinct $u, v \in V(F)$, the number of composed paths of length $2s$ with endpoints u, v is at most*

$$\Delta^{s-1}\left((sL+1)^s + 1\right).$$

Proof. Let $\mathcal{P}$ be the set of composed paths of length $2s$ with endpoints u, v. For given vertices $a_1, \ldots, a_{s-1}$, let

$$\mathcal{P}(a_1, \ldots, a_{s-1}) = \{ux_1 \ldots x_{2s-1}v \in \mathcal{P} \mid x_1 = a_1, \ldots, x_{s-1} = a_{s-1}\}.$$

Note that the number of non-empty $\mathcal{P}(a_1, \ldots, a_{s-1})$ is at most Δ^{s-1}, since $ua_1 \ldots a_{s-1}$ is a path.

Suppose that $P_0 = ua_1 \ldots a_{2s-1}v$ is a composed path in $\mathcal{P}(a_1, \ldots, a_{s-1})$. For every composed path $P = ua_1 \ldots a_{s-1}x_s \ldots x_{2s-1}v \in \mathcal{P}(a_1, \ldots, a_{s-1}) \setminus \{P_0\}$, $a_{s-1} \ldots a_{2s-1}v$ and $a_{s-1}x_s \ldots x_{2s-1}v$ form a closed walk W of length $2(s+1)$. For every $s \leq i \leq 2s-1$, if $x_i = a_i$, the number of choices for x_i is 1. Otherwise, W contains an even cycle of length at most $2(s+1)$, which contains the edge $a_{i-1}a_i$ and vertex x_i. Since $c_{2k}(a_{i-1}, a_i) \leq L$ for every $2 \leq k \leq s+1$, the number of choices for $x_i \neq a_i$ is at most sL. Therefore, we have

$$|\mathcal{P}(a_1, \ldots, a_{s-1})| \leq (sL+1)^s + 1.$$

Finally, we obtain

$$|\mathcal{P}| = \sum_{a_1, \ldots, a_{s-1}} |\mathcal{P}(a_1, \ldots, a_{s-1})| \leq \Delta^{s-1}\left((sL+1)^s + 1\right).$$

□

Now, we give an upper bound on the multiplicity of F^ℓ.

Lemma 2.23. *For integers $\ell, \Delta \leq m$ and a constant $L \ll \Delta$, let F be an m-vertex graph with maximum degree Δ, such that for every $uv \in E(F)$ and $3 \leq k \leq 2\ell$, $c_k(u,v) \leq L$. For every distinct $u, v \in V(F)$, the number of composed walks of length $2(\ell-1)$ with endpoints u, v is at most*

$$\Delta^{\ell-2}\beta_\ell(L),$$

for a constant $\beta_\ell(L)$ which only depends on ℓ and L.

Proof. Let $\mathcal{W}$ be the number of composed walks of length $2(\ell-1)$ in F with endpoints u, v. For every

$1 \leq i \leq \ell - 1$, let

$$\mathcal{W}_i = \{ux_1 \dots x_{2(\ell-1)-1}v \in \mathcal{W} \mid i \text{ is the first integer such that } x_i = x_{2(\ell-1)-i}\},$$

and then we have $\mathcal{W} = \bigcup_{i=1}^{\ell-1} \mathcal{W}_i$.

Let $1 \leq i \leq \ell - 2$. For every composed walk

$$W = ux_1 \dots x_{\ell-1} \dots x_{2(\ell-1)-1}v \in \mathcal{W}_i,$$

the vertices $\{ux_1 \dots x_i x_{2(\ell-1)-(i-1)} \dots x_{2(\ell-1)-1}v\}$ forms a composed path P of length $2i$. By Lemma 2.22, there are at most $\Delta^{i-1}[(iL+1)^i + 1]$ choices for P. For a fixed P, $W - P$ forms a path of length $\ell - i - 1$ with endpoint x_i or a close composed walk of length $2(\ell - i - 1)$ with endpoint x_i. In the first case, there are at most $\Delta^{\ell-i-1}$ choices, while in the later case, by Lemma 2.20, there are at most $\Delta^{\ell-i-2}\alpha_{\ell-i-1}(L)$ choices. Therefore, we have

$$|\mathcal{W}_i| \leq \Delta^{i-1}\left((iL+1)^i + 1\right) \cdot (\Delta^{\ell-i-1} + \Delta^{\ell-i-2}\alpha_{\ell-i-1}(L)) \leq 2\Delta^{\ell-2}\left((iL+1)^i + 1\right).$$

Moreover, every walk $W \in \mathcal{W}_{\ell-1}$ is a composed path of length $2(\ell - 1)$ with endpoints u and v. By Lemma 2.22, we have

$$|\mathcal{W}_{\ell-1}| \leq \Delta^{\ell-2}\left((\ell L - L + 1)^{\ell-1} + 1\right).$$

Hence, we have

$$|\mathcal{W}| = \sum_{i=1}^{\ell-1} |\mathcal{W}_i| \leq \sum_{i=1}^{\ell-2} 2\Delta^{\ell-2}\left((iL+1)^i + 1\right) + \Delta^{\ell-2}\left((\ell L - L + 1)^{\ell-1} + 1\right) = \Delta^{\ell-2}\beta_\ell(L)$$

for $\beta_\ell(L) = \sum_{i=1}^{\ell-2} 2\left((iL+1)^i + 1\right) + \left((\ell L - L + 1)^{\ell-1} + 1\right)$. $\square$

We have all the ingredients to give a lower bound on the number of edges in auxiliary graph F_*^l. This lemma will play the same role as Lemma 2.11 in the case of 4-cycles.

Lemma 2.24. *For an integer $\ell \geq 3$ and a constant $L > 0$, let n be a sufficiently large integer. Let m and d be the integers satisfying $m \leq n$ and $d \geq \frac{n^{1/\ell}}{\log n}$. Suppose F is an m-vertex graph with minimum degree $d-1$, such that for every $uv \in E(F)$ and $3 \leq k \leq 2\ell$, $c_k(u,v) \leq L$. Then for every set $J \subseteq V(F)$ of size at least $2^\ell n/d^{\ell-1}$, we have*

$$e(F_*^\ell[J]) \geq \frac{d^\ell |J|^2}{n \log^{\ell(\ell-2)} n} f_\ell(L)$$

for a constant $f_\ell(L)$ which only depends on ℓ and L.

Proof. Note that every composed walk of length $2(\ell-1)$ with endpoints in J contains a composed path of length at most $2(\ell-1)$ with endpoints in J. Therefore, by Lemma 2.23, we have

$$e(F_*^\ell[J]) \geq \frac{e(F^\ell[J])}{\Delta^{\ell-2}\beta_\ell(L)},$$

where Δ is the maximum degree of F, which by (2.17), satisfies

$$\Delta \leq d \log^\ell n \cdot g_\ell(L).$$

Hence, we have

$$e(F_*^\ell[J]) \geq \frac{e(F^\ell[J])}{d^{\ell-2}\log^{\ell(\ell-2)} n \cdot g_\ell^{\ell-2}(L)\beta_\ell(L)} \geq \frac{d^\ell|J|^2}{n\log^{\ell(\ell-2)} n} f_\ell(L),$$

where $f_\ell(L) = \frac{1}{2^{2\ell+1} g_\ell^{\ell-2}(L)}\beta_\ell(L)$. □

2.3.3 Certificate lemma

In this section, we give our second main lemma, which will be used to build certificates for graphs with sparse short cycles. This lemma is a generalization of Lemma 2.12 for longer cycle, although the condition is slightly different. The idea of proof is also similar to Lemma 2.12, which originally comes from Kleitman and Winston [68] and Kohayakawa, Kreuter and Steger [69].

Lemma 2.25. *For an integer $\ell \geq 3$ and constants $L, \alpha > 0$, let n be a sufficiently large integer. Let m and d be the integers satisfying $m \leq n$ and $\frac{n^{1/\ell}}{\log n} \leq d \leq \alpha n^{1/\ell}$. Suppose F is an m-vertex graph with minimum degree $d-1$, such that for every $uv \in E(F)$ and $3 \leq k \leq 2\ell$, $c_k(u,v) \leq L$. Let $H = F_*^\ell$. Then for every set $I \subseteq V(F)$ of size d, such that $d_H(v, I) \leq (\ell-1)L$ for all $v \in I$, there exist a set T and a set $C(T)$ depending only on T, not on I, such that*

(i) $T \subseteq I \subseteq C(T)$,

(ii) $|T| \leq n^{1/\ell}/\log n$,

(iii) $|C(T)| \leq (2^\ell + 1)n/d^{\ell-1}$.

Proof. This proof is similar to the proof of Lemma 2.12. We will describe a deterministic algorithm that associates to the set I a pair of sets T and $C(T)$.

We start the algorithm with sets $A_0 = V(H)$, $T_0 = \emptyset$ and a function $t_0(v) = 0$, for every $v \in V(H)$. In the i-th iteration step, we pick a vertex $u_i \in A_i$ of maximum degree in $H[A_i]$. If $u_i \in I$, we define

$$t_{i+1}(v) = \begin{cases} t_i(v) + d_H(v, u_i) & \text{if } v \in A_i, \\ t_i(v) & \text{if } v \notin A_i, \end{cases}$$

and $Q = \{v \mid t_{i+1}(v) > (\ell - 1)L\}$, and let $T_{i+1} = T_i + u_i$, $A_{i+1} = A_i - u_i - Q$. Otherwise, let $T_{i+1} = T_i$, $A_{i+1} = A_i - u_i$ and $t_{i+1}(v) = t_i(v)$, for every $v \in V(H)$. The algorithm terminates at step K when we get a set A_K of size at most $2^\ell n/d^{\ell-1}$. We also assume that $u_{K-1} \in T_K$ as otherwise we can continue the algorithm until it is satisfied.

The algorithm outputs a vertex sequence $\{u_1, u_2, \ldots, u_{K-1}\}$, a set of 'representative' vertices T_K and a strictly decreasing set sequence $\{A_0, A_1, A_2, A_3, \ldots A_K\}$. Let $T = T_K$ and $C(T) = A_K \cup T$. From the algorithm, we have $T \subseteq I$. Furthermore, if a vertex v satisfies $t_i(v) > (\ell - 1)L$ for some i, then we have $d_H(v, I) \geq t_i(v) > (\ell - 1)L$, which implies $v \notin I$. Therefore, we maintain $I \subseteq A_i \cup T_i$ for every $i \leq K$ and especially get $I \subseteq A_K \cup T_K = C(T)$. Hence, Condition (i) is satisfied. Similarly as in Lemma 2.12, the set $C(T)$ only depends on T, not on I.

To finish the proof, it is sufficient to show that $|T_K| \leq n^{1/\ell}/\log n$. Once we prove it, we immediately obtain

$$|T| = |T_K| \leq \frac{n^{1/\ell}}{\log n},$$

and

$$|C(T)| = |A_K| + |T| \leq \frac{2^\ell n}{d^{\ell-1}} + \frac{n^{1/\ell}}{\log n} \leq \frac{(2^\ell + 1)n}{d^{\ell-1}},$$

which completes the proof.

In the rest of proof, we apply the same technique used in the proof of Lemma 2.12. We repeat the process as follows. Denote q the integer such that $n/2^q \leq |A_K| < n/2^{q-1}$. By the choice of A_K, we have $q \leq \log n$. For every integer $1 \leq l \leq q$, define A^l to be the first A-set satisfying

$$\frac{n}{2^l} \leq |A^l| < \frac{n}{2^{l-1}},$$

if it exists, and let T^l be the corresponding T-set and $t^l(v)$ be the corresponding t-function of A^l. Note that

A^l may not exist for every l, but A^q always exists and it could be that $A^q = A_K$. Suppose

$$A^{l_1} \supset A^{l_2} \supset \ldots \supset A^{l_p}$$

are all the defined A^l, where $p \leq q$. By the above definition, we have $A^{l_1} = A_0$, $T^{l_1} = T_0$ and $l_p = q$. Define $A^{l_{p+1}} = A_K, T^{l_{p+1}} = T_K$. Now, we have

$$T_K = \bigcup_{j=2}^{p+1} (T^{l_j} - T^{l_{j-1}}). \tag{2.18}$$

To achieve our goal, we are going to estimate the size of $T^{l_j} - T^{l_{j-1}}$, for every $2 \leq j \leq p+1$.

From the algorithm, we have $t^{l_j}(v) \leq (\ell-1)L$, for every $v \in A^{l_j} \cup T^{l_j}$. Moreover, for $v \in A^{l_{j-1}} - A^{l_j} - T^{l_j}$, suppose v is removed in step i, then we have

$$t^{l_j}(v) = t_{i-1}(v) + d_H(v, u_{i-1}) \leq (\ell - 1)L + 1,$$

where u_i is the selected vertex in step i. Therefore, we obtain

$$\sum_{v \in A^{l_{j-1}}} t^{l_j}(v) \leq ((\ell-1)L+1)\,|A^{l_{j-1}}| \leq ((\ell-1)L+1)\,\frac{n}{2^{l_{j-1}-1}}. \tag{2.19}$$

Let $2 \leq j \leq p$. For every $u_i \in T^{l_j} - T^{l_{j-1}}$, u_i is chosen of maximum degree in $H[A_i]$, where A_i is a set between $A^{l_{j-1}}$ and A^{l_j}. By the choice of A^{l_j}, we have $|A_i| \geq n/2^{l_{j-1}}$. From Lemma 2.24, we obtain that

$$d_H(u_i, A_i) \geq \frac{d^\ell |A_i|}{n \log^{\ell(\ell-2)} n} f_\ell(L) \geq \frac{d^\ell}{2^{l_{j-1}} \log^{\ell(\ell-2)} n} f_\ell(L).$$

Note that $d_H(u_i, A_i)$ only contributes to $t^{l_j}(v)$, for $v \in A_i \subseteq A^{l_{j-1}}$. Then we obtain

$$\left|T^{l_j} - T^{l_{j-1}}\right| \frac{d^\ell}{2^{l_{j-1}} \log^{\ell(\ell-2)} n} f_\ell(L) \leq \sum_{u_i \in T^{l_j} - T^{l_{j-1}}} d_H(u_i, A_i) \leq \sum_{v \in A^{l_{j-1}}} t^{l_j}(v). \tag{2.20}$$

Combining (2.19) and (2.20), we have

$$\left|T^{l_j} - T^{l_{j-1}}\right| \frac{d^\ell}{2^{l_{j-1}} \log^{\ell(\ell-2)} n} f_\ell(L) \leq ((\ell-1)L+1)\,\frac{n}{2^{l_{j-1}-1}},$$

which implies

$$\left|T^{l_j} - T^{l_{j-1}}\right| \le \frac{2\,(\ell-1)L+1)}{f_\ell(L)} \cdot \frac{n\log^{\ell(\ell-2)} n}{d^\ell} \le \frac{2\,((\ell-1)L+1)}{f_\ell(L)} \log^{\ell(\ell-1)} n \le \frac{n^{1/\ell}}{\log^2 n},$$

for $2 \le j \le p$. For $j = p+1$, since we have $\frac{n}{2^q} \le |A^{l_{p+1}}| \le |A^{l_p}| \le \frac{n}{2^{q-1}}$, by a similar argument, we obtain that

$$\begin{aligned}\left|T^{l_{p+1}} - T^{l_p}\right| \frac{d^\ell |A_{l_{p+1}}|}{n\log^{\ell(\ell-2)} n} f_\ell(L) &\le \sum_{u_i \in T^{l_{p+1}} - T^{l_p}} d(u_i, A_i) \\ &\le \sum_{v \in A^{l_p}} t^{l_{p+1}}(v) \le ((\ell-1)L+1)\,|A^{l_p}|,\end{aligned}$$

which gives

$$\left|T^{l_{p+1}} - T^{l_p}\right| \le \frac{2\,((\ell-1)L+1)}{f_\ell(L)} \cdot \frac{n\log^{\ell(\ell-2)} n}{d^\ell} \le \frac{n^{1/\ell}}{\log^2 n}.$$

Finally, by (2.18), we have

$$|T_K| = \bigcup_{j=2}^{p+1} |T^{l_j} - T^{l_{j-1}}| \le p \cdot \frac{n^{1/\ell}}{\log^2 n} \le q \cdot \frac{n^{1/\ell}}{\log^2 n} \le \frac{n^{1/\ell}}{\log n}.$$

□

2.3.4 Proof of Theorem 2.5

This section is entirely devoted to the proof of Theorem 2.5. The idea is the same as the proof of Theorem 2.15: we will build a certificate for each graph in $\mathcal{G}_n(2\ell, L)$ and estimate the number of such certificates. Before we proceed, we first need the supersaturation result for $C_{2\ell}$ to give a bound on the min-degree sequence of graphs in $\mathcal{G}_n(2\ell, L)$. It was mentioned in [36] that Simonovits first proved the supersaturation for the even cycles, but the proof has not been published yet and it might appear in Faudree and Simonovits [43]. Morris and Saxton [86] recently provided a stronger version of supersaturation for even cycles. Very recently, Jiang and Yepremyan [65] give a supersaturation result of even linear cycles in linear hypergraphs, which includes the graph case. We use the graph version of their result and rephrase it in terms of the average degree.

Theorem 2.26. *[65] For an integer $\ell \ge 2$, there exist constants C, c such that if G is an n-vertex graph with the average degree $d \ge 2Cn^{1/\ell}$, then G contains at least $c(\frac{d}{2})^{2\ell}$ copies of $C_{2\ell}$.*

Corollary 2.27. *Let G be a n-vertex graph in $\mathcal{G}_n(2\ell, L)$, and $d_n, \ldots, d_1$ be the min-degree sequence of G.*

Then for every $i \in [n]$, we have

$$d_i \leq \alpha n^{1/\ell}$$

for some constant $\alpha = \max\{2C, 2(\frac{L}{2c})^{1/2\ell}\}$, where C, c are constants given in Theorem 2.26.

Proof. Suppose that there exists $k \in [n]$, such that $d_k > \alpha n^{1/\ell}$. Then by Theorem 2.26, the number of $C_{2\ell}$'s in G_k is at least

$$c\left(\frac{d_k}{2}\right)^{2\ell} > c\left(\frac{\alpha n^{1/\ell}}{2}\right)^{2\ell} \geq c\frac{L}{2c}n^2 \geq L\binom{k}{2},$$

which contradicts the fact that $G \in \mathcal{G}_n(2\ell, L)$. $\square$

Proof of Theorem 2.5. The way to construct the certificate is exactly same with in the proof of Theorem 2.15. Here we restate the process. For a graph $G \in \mathcal{G}_n(2\ell, L)$, let $Y_G := v_n < v_{n-1} < \ldots < v_1$ be the min-degree ordering of G and $D_G := \{d_n, d_{n-1}, \ldots, d_1\}$ be the min-degree sequence of G. Note that by Corollary 2.27, there exists a constant α such that $d_i \leq \alpha n^{1/\ell}$, for every $i \in [n]$. For every $i \in [n]$, let $G_i = G[v_i, \ldots, v_1]$. Define the set sequence $S_G := \{S_n, S_{n-1}, \ldots, S_2\}$, where $S_i = N_G(v_i, G_{i-1})$. Note that $S_i \subseteq \{v_{i-1}, \ldots, v_1\}$ and $|S_i| = d_i$. By the construction, $[Y_G, D_G, S_G]$ uniquely determines the graph G and so we build a certificate $[Y_G, D_G, S_G]$ for G. To complete the proof, it is sufficient to estimate the number of such certificates.

We first choose a min-degree ordering $Y^* = v_n < v_{n-1} < \ldots < v_1$, and a min-degree sequence $D^* = \{d_n, \ldots, d_1\}$; the number of options is at most

$$n!(\alpha n^{1/\ell})^n. \tag{2.21}$$

Next, we count set sequences $S = \{S_n, S_{n-1}, \ldots, S_2\}$, where $S_i \subseteq \{v_{i-1}, \ldots, v_1\}$ and $|S_i| = d_i$, such that the graph reconstructed by $[Y^*, D^*, S]$, denoted by G_S, are in $\mathcal{G}_n(2\ell, L)$. For every $2 \leq i \leq n$, let M_i be the number of choices for S_i with fixed sets $S_{i-1}, \ldots, S_2$. Define

$$\mathcal{I}_1 = \{i : d_i < n^{1/\ell}/\log n\}, \qquad \mathcal{I}_2 = \{i : d_i \geq n^{1/\ell}/\log n\}.$$

For every $i \in \mathcal{I}_1$, since $|S_i| = d_i < n^{1/\ell}/\log n$, we have a trivial bound

$$M_i \leq \binom{i-1}{d_i} \leq \binom{n}{n^{1/\ell}/\log n} \leq n^{n^{1/\ell}/\log n} = 2^{n^{1/\ell}}. \tag{2.22}$$

It remains to consider the upper bound on M_i for $i \in \mathcal{I}_2$. With fixed sets $S_{i-1}, \ldots, S_2$, the graph $G_{i-1} = G_S[v_{i-1}, \ldots, v_1]$ is uniquely determined. Since $G_{i-1} \subseteq G_S$ and $G_S \in \mathcal{G}_n(2\ell, L)$, for every $uv \in E(G_{i-1})$ and every $3 \leq k \leq 2\ell$, we know that $c_k(u, v; G_{i-1}) \leq L$. Note that every eligible S_i should satisfy

$d_H(u,S_i) \le (l-1)L$ for all $u \in S_i$, where $H = (G_{i-1})^l_*$. Otherwise, there exists a vertex $u \in S_i$, such that $\sum_{k=2}^{\ell} c_{2k}(v_i,u;G_i) \ge d_H(u,S_i) > (l-1)L$, which is a contradiction. Applying Lemma 2.25 on G_{i-1}, we obtain that every eligible S_i contains a subset T of size at most $n^{1/\ell}/\log n$, which uniquely determines a set $C(T) \supseteq S_i$ of size at most $(2^\ell+1)n/d_i^{\ell-1}$. Since the number of choices for T is at most

$$\sum_{0\le j\le n^{1/\ell}/\log n} \binom{i-1}{j} \le 2\binom{i-1}{n^{1/\ell}/\log n} \le 2\binom{n}{n^{1/\ell}/\log n} \le 2^{n^{1/\ell}},$$

we then have

$$M_i \le \sum_T \binom{C(T)}{d_i} \le \sum_T \binom{\frac{(2^\ell+1)n}{d_i^{\ell-1}}}{d_i} \le \sum_T 2^{\frac{\ell}{1.88}\left((2^\ell+1)en\right)^{1/\ell}} \le \sum_T 2^{3\ell n^{1/\ell}} \le 2^{(3\ell+1)n^{1/\ell}} \tag{2.23}$$

for every $i \in \mathcal{I}_2$, where the third inequality is given by Lemma 2.10.

Combining (2.22) and (2.23), we obtain that the number of choices for S is

$$\prod_{i=2}^n M_i \le \prod_{i\in\mathcal{I}_1} M_i \prod_{i\in\mathcal{I}_2} M_i \le 2^{n^{1+1/\ell}} 2^{(3\ell+1)n^{1+1/\ell}} \le 2^{(3\ell+2)n^{1+1/\ell}}.$$

Hence, together with (2.21), the total number of certificates is at most

$$n!(\alpha n^{1/\ell})^n \prod_{i=2}^n M_i \le 2^{n^{1+1/\ell}} 2^{(3\ell+2)n^{1+1/\ell}} \le 2^{3(\ell+1)n^{1+1/\ell}}$$

for n sufficiently large, which leads to $|\mathcal{G}_n(2\ell,L)| \le 2^{3(\ell+1)n^{1+1/\ell}}$. □

2.4 Linear hypergraphs of large girth

In this section, we study the enumeration problems of r-graphs with given girth and r-graphs without C_4^r's. To prove it, we need a result on the linear Turán number of linear cycles given by Collier-Cartaino, Graber and Jiang [27].

Theorem 2.28. *[27] For every $r,\ell \ge 3$, there exists a constant $\alpha_{r,\ell} > 0$, depending on r and ℓ, such that*

$$\mathrm{ex}_L(n, C_\ell^r) \le \alpha_{r,\ell} n^{1+\frac{1}{\lfloor \ell/2 \rfloor}}.$$

2.4.1 Proof of Theorem 2.3

Once we have Theorems 2.4 and 2.5, it is natural to think about reducing the hypergraph problems to problems on graphs and then apply our graph counting theorems.

Definition 2.29 (Shadow graph)**.** *Given a hypergraph H, the shadow graph of H, denoted by $\partial_2(H)$, is defined as*

$$\partial_2(H) = \{D : |D| = 2, \exists e \in H, D \subseteq e\}.$$

Proposition 2.30. *Let $r \geq 3$, $\ell \geq 2$ and $H \in \mathrm{Forb}_L(n, r, 2\ell)$. For every r-element subset $S \in V(H)$, S forms an r-clique in $\partial_2(H)$ if and only if S is a hyperedge in H.*

Proof. Assume that there exists a r-clique with vertex set S in $\partial_2(H)$ and two edges e_1, e_2 such that e_1, e_2 lie on two different hyperedges f_1, f_2. Without loss of generality, we can assume that e_1 and e_2 share a common vertex, as otherwise, we let $e_1 = ab$ and $e_2 = cd$ and one of the edge pairs $\{ab, ac\}$ or $\{ac, cd\}$ is contained in different hyperedges.

Let $e_1 = ab \subset f_1$ and $e_2 = ac \subset f_2$. Note that $c \notin f_1$ and $b \notin f_2$, as otherwise we have $f_1 = f_2$ by the linearity of H. Let f_3 be the hyperedge which includes bc. Then f_1, f_2, f_3 are distinct, and form a C_3^r by the linearity of H. This contradicts the fact that $H \in \mathrm{Forb}_L(n, r, 2\ell)$. $\square$

We also need the following short lemma on 4-cycles of the shadow graphs of hypergraphs in $\mathrm{Forb}_L(n, r, 4)$.

Lemma 2.31. *For every $r \geq 3$, there exists a constant $\beta = \beta(r)$ such that for every $H \in \mathrm{Forb}_L(n, r, 4)$, the shadow graph $\partial_2(H)$ contains at most $\beta n^{3/2}$ 4-cycles.*

Proof. Let $G = \partial_2(H)$. Since the girth of H is larger than 4, every 4-cycle in G must be contained in a hyperedge of H. By Theorem 2.28, we have $e(H) \leq \alpha_{r,4} n^{3/2}$. Hence, the number of 4-cycles in G is at most

$$\binom{r}{4} e(H) \leq \binom{r}{4} \alpha_{r,4} n^{3/2} = \beta n^{3/2}$$

for $\beta = \binom{r}{4}\alpha_{r,4}$. $\square$

Proof of Theorem 2.3 for $\ell = 2$. Define a map $\varphi : \mathrm{Forb}_L(n, r, 4) \to \mathcal{G} = \{\partial_2(H) : H \in \mathrm{Forb}_L(n, r, 4)\}$ given by $\varphi(H) = \partial_2(H)$. By Proposition 2.30, φ is a bijection. Note that by Lemma 2.31, every graph in $\mathcal{G}$ has at most $\beta n^{3/2}$ 4-cycles, where β is a constant depending on r. Applying Theorem 2.4, when n is sufficiently large, we have

$$|\mathcal{G}| \leq 2^{11 n^{3/2}}.$$

Hence, we obtain that $|\mathrm{Forb}_L(n,r,4)| = |\mathcal{G}| \le 2^{11n^{3/2}}$ for n sufficiently large, which completes the proof. $\square$

Proof of Theorem 2.3 for $\ell \ge 3$. Define a map $\varphi : \mathrm{Forb}_L(n,r,2\ell) \to \mathcal{G} = \{\partial_2(H) : H \in \mathrm{Forb}_L(n,r,2\ell)\}$ given by $\varphi(H) = \partial_2(H)$. By Proposition 2.30, φ is a bijection. For a graph $G = \partial_2(H) \in \mathcal{G}$ and an edge $uv \in E(G)$, since the girth of H is larger than 2ℓ, each k-cycle in G, which contains edge uv, must be contained in a hyperedge of H, for all $3 \le k \le 2\ell$. Indeed, this hyperedge is unique by the linearity of H. Therefore, we have

$$c_k(u,v;G) \le \binom{r-2}{k-2}$$

for all $3 \le k \le 2\ell$. Applying Theorem 2.5, when n is sufficiently large, we have

$$|\mathcal{G}| \le 2^{3(\ell+1)n^{1+1/\ell}}.$$

Hence, we obtain that $|\mathrm{Forb}_L(n,r,2\ell)| = |\mathcal{G}| \le 2^{3(\ell+1)n^{1+1/\ell}}$ for n sufficiently large, which completes the proof. $\square$

2.4.2 Proof of Theorem 2.1

We now estimate the number of r-graphs without C_4^r. The main idea is the same as in the previous section: we convert the hypergraph enumeration problem to a graph enumeration problem and then apply Theorem 2.4. However, because of the possible existence of C_3^r's, some facts we used before is no longer trivial and even not true. The first difficulty is to give an upper bound on the number of 4-cycles in shadow graphs, and we need the following lemma on the number of C_3^r's.

Lemma 2.32. *Let $r \ge 3$. For every $H \in \mathrm{Forb}_L(n, C_4^r)$ and every edge $e \in E(H)$, the number of C_3^r's in H containing e as an edge is at most*

$$\binom{r}{2}(4r^2 - 10r + 7).$$

Proof. For every distinct $u, v \in e$, let

$$\mathcal{C}_{u,v} = \{\{e, f_i, g_i\} \subseteq H : e \cap f_i = \{u\}, e \cap g_i = \{v\}, |f_i \cap g_i| = 1\}.$$

Suppose $\mathcal{C}_{u,v}$ is nonempty, and fix a $C_0 = \{e, f_0, g_0\} \in \mathcal{C}_{u,v}$. For every $C = \{e, f_i, g_i\} \in \mathcal{C}_{u,v} \setminus \{C_0\}$, we know that

$$(f_0 \cup g_0) \cap (f_i \cup g_i) - \{u, v\} \ne \emptyset,$$

otherwise, $\{f_0, g_0, f_i, g_i\}$ would form a C_4^r. Let w be a vertex in $(f_0 \cup g_0) \cap (f_i \cup g_i) - \{u, v\}$. Since $w \in f_0 \cup g_0 - \{u, v\}$, there are at most $2r - 3$ choices for w. By linearity of H, the number of linear 3-cycles in $\mathcal{C}_{u,v}$ containing w is at most $2(r-1)$. Therefore, we get

$$|\mathcal{C}_{u,v}| \leq 1 + 2(r-1)(2r-3) = 4r^2 - 10r + 7. \tag{2.24}$$

Hence, the number of C_3^r's in H containing e as an edge is equal to

$$\sum_{u,v \in e} |\mathcal{C}_{u,v}| \leq \binom{r}{2}(4r^2 - 10r + 7).$$

□

Proposition 2.33. *For $H \in \mathrm{Forb}_L(n, C_4^r)$, every 4-cycle in $\partial_2(H)$ must be contained in a hyperedge or a C_3^r of H.*

Proof. Assume that a 4-cycle $abcd$ is not contained in any hyperedge of H. Then there exist two edges e_1 and e_2 which lie on two different hyperedges f_1 and f_2. Without loss of generality, we can assume that $e_1 = ab \subset f_1$, and $e_2 = ad \subset f_2$. Note that $d \notin f_1$ and $b \notin f_2$, as otherwise we have $f_1 = f_2$ by the linearity of H. Let f_3 be the hyperedge which includes bd. Then f_1, f_2, f_3 are distinct, and form a C_3^r by the linearity of H. This contradicts the fact that $H \in \mathrm{Forb}_L(n, C_4^r)$. □

Lemma 2.34. *For every $r \geq 3$, there exists a constant $\beta = \beta(r)$ such that for every $H \in \mathrm{Forb}_L(n, C_4^r)$, the shadow graph $\partial_2(H)$ contains at most $\beta n^{3/2}$ 4-cycles.*

Proof. Let $G = \partial_2(H)$. We first claim that every 4-cycle in G is contained in a hyperedge or a C_3^r of H. By Lemma 2.32, there are at most

$$\frac{1}{3}\binom{r}{2}\left(4r^2 - 10r + 7\right) e(H)$$

C_3^r's in H. Since H is linear and contains no C_4^r, every 4-cycle in G must be contained in a hyperedge or a C_3^r of H. Moreover, by Theorem 2.28, we have

$$e(H) \leq \alpha_{r,4} n^{3/2}.$$

Hence, the number of 4-cycles in G is at most

$$3\binom{r}{4}e(H) + 3\binom{3r-3}{4} \cdot \frac{1}{3}\binom{r}{2}(4r^2 - 10r + 7)e(H) \leq \beta n^{3/2}$$

for

$$\beta = \left[3\binom{r}{4} + \binom{3r-3}{4}\binom{r}{2}(4r^2-10r+7)\right]\alpha_{r,4},$$

where $\alpha_{r,4}$ is a constant defined in Theorem 2.28. $\square$

Another difficulty is that the map we defined in the proof of Theorem 2.3 might be no longer injective. To overcome it, we have the following lemma to measure how far the map is from the injection.

Lemma 2.35. *For every $r \geq 3$, there exists a constant $\alpha = \alpha(r)$ such that for every $H \in \mathrm{Forb}_L(n, C_4^r)$, there are at most $\alpha n^{3/2}$ r-cliques in $\partial_2(H)$.*

Proof. Let $G = \partial_2(H)$ and $\mathcal{F}$ be the set of r-cliques in G. For every $e \in E(H)$, let

$$\mathcal{F}_e = \{F \in \mathcal{F} : |F \cap e| = \max_{f \in E(H)} |F \cap f|\}.$$

Then we have $\mathcal{F} = \bigcup_{e \in H} \mathcal{F}_e$. Fix an arbitrary hyperedge $e \in H$. For every $2 \leq q \leq r$, let

$$\mathcal{R}_q = \{F \in \mathcal{F}_e : |F \cap e| = q\},$$

then we have $\mathcal{F}_e = \bigcup_{q=2}^{r} \mathcal{R}_q$.

First, it is trivial to get $|R_r| = 1$. Let $2 \leq q \leq r-1$ and F be an r-clique in $\mathcal{R}_q$. Since $|F \cap e| = q$, the number of choices for $F \cap e$ is at most $\binom{r}{q}$. Given $F \cap e$, let u, v be two distinct vertices in $F \cap e$. For every $w \in F - e$, by the definition of the shadow graph and the linearity of H, there exist hyperedges f, g such that $\{e, f, g\}$ forms a C_3^r with $e \cap f = u, e \cap g = v$ and $f \cap g = w$. By (2.24), the number of such C_3^r's is at most $4r^2 - 10r + 7$. Therefore, the choices of w is at most $4r^2 - 10r + 7$. Hence, we have

$$\mathcal{R}_q \leq \binom{r}{q}(4r^2 - 10r + 7)^{r-q}.$$

Then, we obtain

$$|\mathcal{F}_e| = \sum_{q=2}^{r} |\mathcal{R}_q| \leq \sum_{r=2}^{r-1} \binom{r}{q}(4r^2-10r+7)^{r-q} + 1 \leq 2^r(4r^2)^r.$$

Finally, we get

$$|\mathcal{F}| = \sum_{e \in E(H)} |\mathcal{F}_e| \leq 2^r(4r^2)^r e(\mathcal{H}) \leq \alpha n^{3/2}$$

for $\alpha = 2^r(4r^2)^r \alpha_{r,4}$, where $\alpha_{r,4}$ is the constant defined in Theorem 2.28. $\square$

Proof of Theorem 2.1. Define a map $\varphi : \mathrm{Forb}_L(n, C_4^r) \to \mathcal{G} = \{\partial_2(H) : H \in \mathrm{Forb}_L(n, C_4^r)\}$ given by

$\varphi(H) = \partial_2(H)$. By Lemma 2.34, every graph $G \in \mathcal{G}$ has at most $\beta n^{3/2}$ 4-cycles, where β is a constant depending on r. By Theorem 2.4, when n is sufficiently large, we have

$$|\mathcal{G}| \leq 2^{11n^{3/2}}.$$

By Lemma 2.35, for every $G \in \mathcal{G}$, the number of r-cliques in G is at most $\alpha n^{3/2}$, where α is a constant depending on r. Since every hyperedge corresponds to an r-clique in its shadow graph, we have

$$|\varphi^{-1}(G)| \leq 2^{\alpha n^{3/2}}.$$

Finally, we obtain

$$|\mathrm{Forb}_L(n, C_4^r)| \leq \sum_{G \in \mathcal{G}} \left|\varphi^{-1}(G)\right| \leq |\mathcal{G}| 2^{\alpha n^{3/2}} \leq 2^{(11+\alpha)n^{3/2}}$$

for n sufficiently large, which completes the proof. $\square$

Chapter 3

On the number of generalized Sidon sets

3.1 Introduction

A set A of nonnegative integers is called a *Sidon set* if there is no 4-tuple (a, b, c, d) in A with $a + b = c + d$ and $\{a, b\} \cap \{c, d\} = \emptyset$. Such a tuple (a, b, c, d) is referred to as a *Sidon 4-tuple*. A famous problem raised by Sidon asks the maximum size $\Phi(n)$ of Sidon subsets of $[n]$. Previous studies of Erdős and Turán [37], Singer [96], Erdős [33], and Chowla [26], have showed that $\Phi(n) = (1 + o(1))\sqrt{n}$. We denote by $\mathcal{Z}_n$ the family of Sidon subsets in $[n]$. Cameron and Erdős [25] first proposed the problem of determining $|\mathcal{Z}_n|$. The extremal result indicates a trivial bound

$$2^{\Phi(n)} \leq |\mathcal{Z}_n| \leq \sum_{1 \leq i \leq \Phi(n)} \binom{n}{i} \leq n^{(1/2+o(1))\sqrt{n}}. \tag{3.1}$$

Cameron and Erdős [25] improved the lower bound by showing $\limsup_n |\mathcal{Z}_n| 2^{-\Phi(n)} = \infty$ and asked if the upper bound could also be improved. Based on the method introduced by Kleitman and Winston [68], Kohayakawa, Lee, Rödl and Samotij [70] strengthened the upper bound to $2^{c\Phi(n)}$, where c is a constant arbitrarily close to $\log(32e) \approx 6.442$ for sufficiently large enough n. Using the hypergraph container method [11, 94], Saxton and Thomason [94] showed that there are between $2^{(1.16+o(1))\sqrt{n}}$ and $2^{(55+o(1))\sqrt{n}}$ Sidon subsets of $[n]$, which indicates that neither of the bounds in (3.1) is tight.

We consider counting sets in which a positive upper bound is imposed on the number of Sidon 4-tuples. An α*-generalized Sidon set* in $[n]$ is a set with at most α Sidon 4-tuples. One way to extend the Cameron and Erdős problem is to estimate the number of α-generalized Sidon sets. Clearly, a trivial lower bound of $2^{\Omega(\sqrt{n})}$ can be given by the number of Sidon sets. In this chapter, we focus on the case when α is small. In particular, we are interested in determining how large can α be such that the number of α-generalized Sidon subsets in $[n]$ is still $2^{\Theta(\sqrt{n})}$.

For a set $I \subseteq [n]$ and a vertex $v \in [n]$, let $S_I(v)$ be the set of Sidon 4-tuples in I containing v and write $s_I(v) = |S_I(v)|$. Denote by $\mathcal{I}_n(\alpha)$ the family of α-generalized Sidon sets I in $[n]$ with $|I| \leq \sqrt{n}/\log n$

or $|I| \geq \sqrt{n}/\sqrt{\log n}$, and $\mathcal{J}_n(\alpha)$ the family of α-generalized Sidon sets I in $[n]$ with $|\{v \in I : s_I(v) \geq \sqrt{n}/\log^4 n\}| \leq \sqrt{n}/\log n$. Our main results are the following.

Theorem 3.1. *Let $\alpha = n/\log^4 n$. For n sufficiently large, we have $|\mathcal{I}_n(\alpha)| \leq 2^{180\sqrt{n}}$.*

Theorem 3.2. *Let $\alpha = n/\log^4 n$. For n sufficiently large, we have $|\mathcal{J}_n(\alpha)| \leq 2^{180\sqrt{n}}$.*

One can indeed run the same proofs and show that for any given number $c > 0$, both theorems hold for $\alpha = cn/\log^4 n$ with the upper bound $2^{C\sqrt{n}}$ for some constant C, depending on c.

Theorem 3.2 immediately implies the following.

Corollary 3.3. *For $\alpha = O(n/\log^5 n)$, the number of α-generalized Sidon sets in $[n]$ is $2^{\Theta(\sqrt{n})}$.*

A simple probabilistic argument can be used to give a lower bound on the number of α-generalized Sidon sets in $[n]$: let $m = (\alpha n)^{\frac{1}{4}}$; a typical m-element subset on $[n]$ contains about $\Theta(m^4/n) = \Theta(\alpha)$ Sidon 4-tuples, and there are $2^{\Theta(m \log n)} = 2^{\Theta((\alpha n)^{\frac{1}{4}} \log n)}$ of them. In particular, for $\alpha \gg \sqrt{n}/\log^4 n$, there are $2^{\Theta((\alpha n)^{\frac{1}{4}} \log n)} = 2^{\omega(\sqrt{n})}$ subsets with $\Theta(\alpha)$ Sidon 4-tuples. Therefore, if the number of α-generalized Sidon subsets in $[n]$ has magnitude $2^{\Theta(\sqrt{n})}$, then the order of α cannot be greater than $n/\log^4 n$. We believe that 4 in the exponent is the best possible.

Conjecture 3.4. *For $\alpha = \Theta(n/\log^4 n)$, the number of α-generalized Sidon sets in $[n]$ is $2^{\Theta(\sqrt{n})}$.*

The main idea of our proofs is based on the graph container method, in which we assign a *cerfiticate* to each set I in $\mathcal{I}_n(\alpha)$ (or $\mathcal{J}_n(\alpha)$) such that I is contained in a unique 'container' determined by its certificate. The certificate should be sufficiently small so that the total number of certificates is properly bounded. Moreover, for each certificate, the number of sets I assigned to it should not be large. Then we can estimate the size of $\mathcal{I}_n(\alpha)$ (or $\mathcal{J}_n(\alpha)$) by counting their certificates. Again, the classical graph container method only applies for the independent sets while we study on the sets with sparse structure. Therefore, similarly as in Chapter 2, we need to make some modifications of the argument.

Although we did not manage to achieve our goal, i.e., to prove Conjecture 3.4, our proof still contains a few new ideas which might be useful to attack some other problems. This chapter is organized as follows. In Section 3.2, we present a supersaturation lemma and some probabilistic results to be used in Section 3.3. In Section 3.3, we introduce our certificate lemmas, Lemmas 3.12 and 3.14, which are used to prove Theorem 3.1 and 3.2 respectively. The proofs of Theorems 3.1 and 3.2 are given in Section 3.4. Finally, we have some concluding remarks in Section 3.5.

3.2 Supersaturation and probabilistic tools

3.2.1 Supersaturation

For two sets $A, U \subseteq [n]$, define a multigraph $H^U(A)$ on vertex set A such that for every $a_1, a_2 \in A$ with $a_1 < a_2$, the multiplicity of the edge a_1a_2 in $H^U(A)$ is the number of ordered pairs (u_1, u_2) in U such that (a_1, u_1, u_2, a_2) is a Sidon 4-tuple. We shall use the following simple supersaturation result.

Lemma 3.5. *Let $A, U \subseteq [n]$. If $|A| \cdot |U| \geq 6n$, then $e(H^U(A)) > \frac{|A|^2|U|^2}{12n}$.*

Proof. Let F be a simple bipartite graph defined on the set $A \cup [2n]$ satisfying that for every $a \in A$ and $m \in [2n]$, a is adjacent to m if and only if there is an element $u \in U$ such that $a + u = m$. Clearly, for every vertex $a \in A$, we have $d_F(a) = |U|$.

Let $\mathcal{P}$ be the set of paths of length 2 (or 3-paths) in F with endpoints in A. Then we have

$$|\mathcal{P}| = \sum_{m\in[2n]} \binom{d_F(m)}{2} \geq 2n\binom{\frac{\sum_{m\in[2n]} d_F(m)}{2n}}{2} = 2n\binom{\frac{|A|\cdot|U|}{2n}}{2} > \frac{|A|^2|U|^2}{6n}.$$

A path $P = \{xyz\} \in \mathcal{P}$ is called *trivial* if $x + z = y$; otherwise, P is *nontrivial*. Note that P is trivial if and only if both x and y belong to $A \cap U$. Thus, the number of trivial paths in $\mathcal{P}$ is exactly $\binom{|A\cap U|}{2}$. Let P' be the set of nontrivial paths in P. Every 3-path in P' corresponds to an edge in $H^U(A)$ and vice versa. Therefore, we obtain

$$e(H^U(A)) = |P'| = |P| - \binom{|A \cap U|}{2} > \frac{|A|^2|U|^2}{6n} - \frac{|A| \cdot |U|}{2} \geq \frac{|A|^2|U|^2}{12n},$$

where the first inequality is given by $|A \cap U| \leq \min\{|A|, |U|\} \leq \sqrt{|A| \cdot |U|}$ and the second inequality follows from the assumption $|A| \cdot |U| \geq 6n$. □

Corollary 3.6. *Let $A \subseteq [n]$ be a set with at most $3n$ Sidon 4-tuples. Then $|A| < \sqrt{6n}$.*

Proof. Apply Lemma 3.5 with $U = A$. Then we obtain that the number of Sidon 4-tuples in A is more than $|A|^4/12n$. On the other hand, the assumption states that there are at most $3n$ Sidon 4-tuples, which indicates that $|A|^4/12n < 3n$, i.e., $|A| < \sqrt{6n}$. □

Lemma 3.7. *Suppose $I, A \subseteq [n]$ and $g \leq n$. For every set $U \subseteq \{v \in I \mid s_I(v) < g\}$ and edge $ab \in H^U(A)$, the multiplicity of ab in $H^U(A)$ is at most g.*

Proof. Let m be the multiplicity of the edge ab in $H^U(A)$. By the definition of $H^U(A)$, there exist $u_1, u_2, \ldots, u_m$, $v_1, v_2, \ldots, v_m \in U$ such that $a + u_i = v_i + b$, for every $i \in [m]$. Then for every $i \in [m] \setminus \{1\}$, we have $u_i - v_i = b - a = u_1 - v_1$, i.e., $u_1 + v_i = u_i + v_1$. Since $u_1 \in U \subseteq \{v \in I \mid s_I(v) < g\}$, we must have $m - 1 \leq s_U(u_1) \leq s_I(u_1) < g$, that is, $m \leq g$. □

3.2.2 Large deviations for sum of partly dependent random variables

The classical Chernoff bound is a powerful tool, but it only applies to sums of random variables that are independent. Janson [64] extended a method of Hoeffding and obtained strong large deviation bounds for sums of dependent random variables with suitable dependency structure. For a family of random variables $\{Y_\alpha\}_{\alpha\in\mathcal{A}}$, a *dependency graph* is a graph Γ with vertex set $\mathcal{A}$ such that if $\mathcal{B} \subset \mathcal{A}$ and $\alpha \in \mathcal{A}$ is not connected by an edge to any vertex in $\mathcal{B}$, then Y_α is independent of $\{Y_\beta\}_{\beta\in\mathcal{B}}$. Recall that $\Delta(\Gamma)$ denotes the maximum degree of Γ and let (for convenience) $\Delta_1(\Gamma) := \Delta(\Gamma) + 1$.

Theorem 3.8 ([64], Corollary 2.2)**.** *Suppose that X is a random variable which can be written as a sum*

$$X = \sum_{\alpha\in\mathcal{A}} Y_\alpha,$$

where each Y_α is an indicator variable taking the values 0 and 1 only. Let Γ be the dependency graph for $\{Y_\alpha\}_{\alpha\in\mathcal{A}}$. Then for $t \geq 0$,

$$\mathbb{P}(X \geq \mathbb{E}[X] + t) \leq \exp\left(-2\frac{t^2}{\Delta_1(\Gamma)|\mathcal{A}|}\right).$$

3.2.3 Some probabilistic results

For this section, fix $\alpha = \frac{\sqrt{n}}{\log^4 n}$. Let I be an α-generalized Sidon set in $[n]$ such that for every $v \in I$, $s_I(v) < \frac{\sqrt{n}}{\log^3 n}$. Define $I_h = \{v \in I : s_I(v) \geq \frac{\sqrt{n}}{\log^4 n}\}$. We further assume that

$$|I| \geq \frac{\sqrt{n}}{\sqrt{\log n}} \quad \text{and} \quad |I_h| > \frac{\sqrt{n}}{\log n}. \tag{3.2}$$

From the Chernoff bound and (3.2), we instantly get the following.

Lemma 3.9. *Let W be a random subset of I obtained by choosing each $u \in I$ independently with probability $p = \frac{2}{\sqrt{\log n}}$. Then $\mathbb{P}\left(|W| < \frac{\sqrt{n}}{\log n}\right) = o(1)$.*

For two different numbers u, v and a set A, let $S(u, A, v) = \{(u, a, b, v) \mid a, b \in A \text{ and } u + a = b + v\}$ and write $s(u, A, v) = |S(u, A, v)|$.

Lemma 3.10. *Let W be a random subset of I obtained by choosing each $u \in I$ independently with probability $p = \frac{2}{\sqrt{\log n}}$. Then almost always $s(u, W, v) \leq 8\frac{\sqrt{n}}{\log^4 n}$, for all $u, v \in I$ simultaneously.*

Proof. It is sufficient to prove the inequality for all $u, v \in I$ with $s(u, I, v) > \frac{8\sqrt{n}}{\log^4 n}$. For a 4-tuple $r = (u, a, b, v) \in S(u, I, v)$, let X_r be the indicator random variable for the event $r \in S(u, W, v)$. Since a, b are always different, we have $\mathbb{P}(X_r = 1) = p^2 = \frac{4}{\log n}$. Then

$$\mu_{uv} = \mathbb{E}[s(u, W, v)] = \mathbb{E}\left[\sum_{r \in S(u,I,v)} X_r\right] = p^2 s(u, I, v) > \frac{32\sqrt{n}}{\log^5 n}.$$

For a given pair of numbers $u, v \in I$, let Γ be the dependency graph for $\{X_r : r \in S(u, I, v)\}$. Then we have $\Delta_1(\Gamma) = \Delta(\Gamma) + 1 \leq 3$. Using Theorem 3.8, we show that

$$\mathbb{P}(s(u, W, v) > 2\mu_{uv}) < \exp\left(-2\frac{\mu_{uv}^2}{3s(u, I, v)}\right) = \exp\left(-2\frac{\mu_{uv}}{3p^2}\right) < \exp\left(-\frac{16\sqrt{n}}{3\log^4 n}\right). \tag{3.3}$$

On the other hand, by Lemma 3.7, we obtain

$$\mu_{uv} = p^2 s(u, I, v) \leq p^2 \frac{\sqrt{n}}{\log^3 n} = \frac{4\sqrt{n}}{\log^4 n}. \tag{3.4}$$

Combining (3.3) and (3.4), we obtain

$$\mathbb{P}\left(s(u, W, v) > 8\frac{\sqrt{n}}{\log^4 n}\right) < \exp\left(-\frac{16\sqrt{n}}{3\log^4 n}\right).$$

Finally, using the union bound, we have

$$\mathbb{P}\left(\exists u, v \in I \text{ s.t. } s(u, W, v) > 8\frac{\sqrt{n}}{\log^4 n}\right) < n^2 \exp\left(-\frac{16\sqrt{n}}{3\log^4 n}\right) = o(1). \qquad \square$$

For two sets $B \subseteq A \subseteq [n]$ and a vertex $v \in A$, let $S_{A,B}(v) = \{(a, b, c, d) \in S_A(v) \mid v \in \{a, d\} \text{ and } b, c \in B\}$ and write $s_{A,B}(v) = |S_{A,B}(v)|$. Note that for a Sidon 4-tuple (a, b, c, d), we can switch the a, b and c, d and the resulting tuple is still a Sidon 4-tuple. Therefore, we have $s_{A,A}(v) = \frac{1}{2}s_A(v)$.

Lemma 3.11. *Let W be a random subset of I obtained by choosing each $u \in I$ independently with probability $p = \frac{2}{\sqrt{\log n}}$. Let $S(W) = \{v \in I \mid s_{I,W}(v) > \frac{\sqrt{n}}{\log^4 n}\}$. Then $|S(W)| \leq \frac{16\sqrt{n}}{\log n}$ almost always.*

Proof. Let $R = \bigcup_{v \in I} S_{I,I}(v)$. Then we have

$$\frac{n}{\log^4 n} \geq |R| = \frac{1}{2}\sum_{v \in I} s_{I,I}(v) = \frac{1}{4}\sum_{v \in I} s_I(v) \geq \frac{1}{4}\sum_{v \in I_h} \frac{\sqrt{n}}{\log^4 n} \geq \frac{n}{4\log^5 n}, \tag{3.5}$$

where the last inequality holds by (3.2). Let $R_W = \bigcup_{v \in I} S_{I,W}(v)$. For every $r \in R$, let X_r be the indicator random variable for the event $r \in R_W$. Note that $\mathbb{P}(X_r = 1) = p^2$. Then we obtain $\mathbb{E}[|R_W|] = \mathbb{E}\left[\sum_{r\in R} X_r\right] = p^2|R|$. Define a simple graph $\Gamma = (R, E)$ such that

$$E = \{r_1r_2 \in \binom{R}{2} \mid r_1 = (a_1, b_1, c_1, d_1),\ r_2 = (a_2, b_2, c_2, d_2) \text{ and } \{b_1, c_1\} \cap \{b_2, c_2\} \neq \emptyset\}.$$

For every $r = (a, b, c, d) \in R$, the number of its neighbors in Γ is at most $s_I(b)+s_I(c) < \frac{2\sqrt{n}}{\log^3 n}$, which implies

$$\Delta_1(\Gamma) = \Delta(\Gamma) + 1 \leq \frac{2\sqrt{n}}{\log^3 n}. \tag{3.6}$$

The graph Γ can be viewed as the dependency graph of $\{X_r\}_{r\in R}$, since X_{r_1}, X_{r_2} are dependent if and only if $r_1r_2 \in E$. By Theorem 3.8, we have

$$\begin{aligned} P(R_W \geq 2\mathbb{E}[R_W]) &\leq \exp\left(-2\frac{(\mathbb{E}[R_W])^2}{\Delta_1(\Gamma)|R|}\right) = \exp\left(-2\frac{p^4|R|}{\Delta_1(\Gamma)}\right) \\ &\leq \exp\left(-2\frac{\frac{2^4}{\log^2 n}\cdot\frac{n}{4\log^5 n}}{\frac{2\sqrt{n}}{\log^3 n}}\right) = \exp\left(-\frac{4\sqrt{n}}{\log^4 n}\right), \end{aligned}$$

i.e.,

$$|R_W| < 2\mathbb{E}[R_W] = 2p^2|R| \leq \frac{8n}{\log^5 n}$$

almost always. Finally, we obtain

$$|S(W)| \leq \frac{2|R_W|}{\frac{\sqrt{n}}{\log^4 n}} \leq \frac{16\sqrt{n}}{\log n}$$

almost always. □

3.3 Certificate lemmas

In this section, we aim to prove two lemmas which are used to define proper certificates for the desired sets. For the proof of Theorem 3.1, we introduce Lemma 3.12 as the certificate lemma. A minor modification of its proof gives Lemma 3.14, which is used to prove Theorem 3.2. The original proof idea comes from Kleitman and Winston [68], who estimated the number of C_4-free graphs. Kohayakawa, Lee, Rödl and Samotij [70] later applied this method to the Sidon problem and gave an upper bound on the number of Sidon sets in $[n]$.

Lemma 3.12. *For a sufficiently large integer n, let $\alpha = n/\log^4 n$ and I be an α-generalized Sidon set in $[n]$ such that for every $v \in I$, $s_I(v) < \sqrt{n}/\log^3 n$. Further assume that the size of I is at least $\sqrt{n}/\sqrt{\log n}$. Then* *there exist set sequences $R_0, R_1, \ldots, R_L$ and $U_0, U_1, \ldots, U_{L-1}$, where $0 \le L < \log\log n + 1$, which determine a unique set sequence $C_0 \supset C_1 \supset C_2 \supset \ldots \supset C_L$. Furthermore, the following are all satisfied:*

(i) $\bigcup_{i=0}^{L} R_i \subseteq I \subseteq C_L \cup \bigcup_{i=0}^{L} R_i$;

(ii) $|C_0| \le n$ *and* $12\sqrt{n} < |C_i| \le \frac{6\sqrt{n}\log n}{2^{i-1}}$, *for* $i = 1, 2, \ldots, L-1$;

(iii) $R_0 \subseteq [n]$ *and* $|R_0| \le \frac{16\sqrt{n}}{\log n}$;

(iv) $R_i \subseteq C_{i-1}$, $|R_1| \le \frac{108\sqrt{n}}{\log n}$ *and* $|R_i| \le \frac{1}{2^{2i-4}} \frac{12\sqrt{n}}{\log n}$, *for* $i = 2, \ldots, L$;

(v) $U_0 \subseteq [n]$ *and* $|U_0| = \frac{\sqrt{n}}{\log n}$;

(vi) $U_i \subseteq C_i$ *and* $|U_i| = 12\frac{n}{|C_i|}$, *for* $i = 1, \ldots, L-1$;

(vii) $L = 0$ *and* $|C_0 \cap I| < \frac{\sqrt{n}}{\log n}$ *or* $|C_L \cap I| < 12\frac{n}{|C_L|}$ *or* $|C_L| \le 12\sqrt{n}$.

We say the set sequences $R_0, R_1, \ldots, R_L$ and $U_0, U_1, \ldots, U_{L-1}$ founded in Lemma 3.12 give a *certificate* for I. Conditions (ii)–(v) guarantee that the number of such certificates is properly bounded. Condition (vii) guarantees that a fixed certificate is associated to small number of sets I. This follows from the fact that the most part of I is contained in C_L.

Proof of Lemma 3.12. Fix a sufficiently large integer n. Following the ideas of [68] and [70], we gave a deterministic algorithm that associates every set I to the desired set sequences.

The core algorithm. We start with sets $A \subseteq [n]$, $T = \emptyset$ and a function $t(v) = 0$, for every $v \in A$. Here, one can view A as the set of 'available' vertices, T as the set of 'selected' vertices, and $t(v)$ as a 'state' function which is used to control the process. As the algorithm proceeds, we add 'selected' vertices from A to T and remove 'ineligible' vertices from A, whose 'state' value exceed some predetermined threshold $t_{\text{threshold}}$. More formally, take the auxiliary graph H ($H = H^U(A)$ for some set U and we will discuss the choice of U later) and choose a vertex $u \in A$ of maximum degree in $H[A]$; we break ties arbitrarily by giving preference to vertices that come early in some arbitrarily predefined ordering. If $u \notin I$, then let $T = T$, $A = A - u$ and $t(v) = t(v)$, for every $v \in A$. Otherwise, let

$$t(v) = \begin{cases} t(v) + d_H(v, u) & \text{for } v \in A, \\ t(v) & \text{for } v \notin A, \end{cases}$$

and define $Q = \{v \in A \mid t(v) > t_{\text{threshold}}\}$; let $T = T \cup \{u\}$ and $A = A - u - Q$. We stop the algorithm when A is sufficiently small.

The goal of the algorithm is to obtain a small representative set T for a given set I such that the choice of T determines a set $A \supseteq I - T$. If A is sufficiently small, then it reduces the number of choices for $I - T$, and hence for I. Note that in each round T increases by at most 1. Therefore, a good algorithm should reduce the size of A rapidly so that we can keep T small in the end. Recall that in every step, we take a vertex u of maximum degree in the auxiliary graph $H^U(A)$ and add it to T when $u \in I$. After that, we delete 'ineligible' vertices, whose 'state' exceed the given threshold. The idea behind this is that if the degree of a vertex is larger than the threshold, then it does not belong to I, since for every $v \in I$, $s_I(v)$ is bounded. To speed up the process, we should take a large set U so that we could quickly accumulate the 'state' value and produce more 'ineligible' vertices in each step. However, the cost of using a larger set U is that the number of choices for U becomes larger and so for the certificates. Therefore, we need to find a balance between the demand for large U and the small number of choices for U. Moreover, ideally if we can find one proper set U through the whole algorithm, then the certificates would be much more concise than in our current lemma. Unfortunately, it turns out that U must vary as the set A shrinks in order to reach the condition of the supersaturation result.

For $i \geq 0$, let A_i, T_i and $t_i(v)$ be the state after running the algorithm i rounds. In the rest of the proof, we divide the iterations of the core algorithm into several phases and then choose a proper auxiliary set U for each phase. In Phase 1, we execute the algorithm from $A_0 = [n]$ to A_{ℓ_1}, which is the first set A_i of size smaller than $6\sqrt{n}\log n$. For $j \geq 2$, Phase j consists of the executions of the algorithm between $A_{\ell_{j-1}}$, the set produced at the end of Phase $j-1$, and A_{ℓ_j}, which is the first set A_i of size smaller than $|A_{\ell_{j-1}}|/2$.

Set-ups for initial certificate $\{\mathbf{R_0}, \mathbf{C_0}\}$. Let $I_l = \{v \in I : s_I(v) < \frac{\sqrt{n}}{\log^4 n}\}$ and $I_h = \{v \in I : s_I(v) \geq \frac{\sqrt{n}}{\log^4 n}\}$. Based on the size of I_h, we have two different set-ups for R_0 and C_0.

Case 1. If $|I_h| \leq \frac{\sqrt{n}}{\log n}$, then we define:

$$R_0 = I_h, \qquad C_0 = [n] - R_0.$$

Case 2. If $|I_h| > \frac{\sqrt{n}}{\log n}$, Lemmas 3.9, 3.10 and 3.11 indicate that there exists a set $W \subseteq I$ of size $\frac{\sqrt{n}}{\log n}$ such that

$$|S(W)| \leq \frac{16\sqrt{n}}{\log n} \tag{3.7}$$

and

$$s(u, W, v) \leq 8\frac{\sqrt{n}}{\log^4 n}, \quad \text{for all } u, v \in I. \tag{3.8}$$

Then we define:

$$R_0 = S(W), \qquad C_0 = [n] - R_0.$$

Phase 1. If $|C_0 \cap I| < \frac{\sqrt{n}}{\log n}$, then we stop the algorithm with $L = 0$. Otherwise, take a set $U_0 \subseteq [n]$ of size $\frac{\sqrt{n}}{\log n}$: for Case 1, let U_0 be an arbitrary subset of $C_0 \cap I$ of size $\frac{\sqrt{n}}{\log n}$; for Case 2, let $U_0 = W$. Denote $H_0 = H^{U_0}(A)$. We now use H_0 as an auxiliary graph and run the core algorithm with $t_{\text{threshold}} = \frac{\sqrt{n}}{\log^4 n}$ and initial state

$$A_0 = C_0, \quad T_0 = \emptyset \quad \text{and} \quad t_0(v) = 0, \text{ for every } v \in A_0,$$

until we obtain the set A_{ℓ_1}, the first set of size smaller than $6\sqrt{n}\log n$.

Let K be the integer such that $\frac{n}{2^K} \leq |A_{\ell_1}| < \frac{n}{2^{K-1}}$. By the choice of A_{ℓ_1}, we have $K \leq \frac{1}{2}\log n$. For every integer $1 \leq k \leq K$, let A^k be the first set satisfying $\frac{n}{2^k} \leq |A^k| < \frac{n}{2^{k-1}}$ if it exists, T^k be the corresponding T-set of A^k and $t^k(v)$ be the corresponding t-function. Note that A^k may not exist for every k. Moreover, A^K always exists and it could be A_{ℓ_1}. Suppose

$$A^{k_1} \supset A^{k_2} \supset \ldots \supset A^{k_p}, \quad p \leq K \leq \frac{1}{2}\log n$$

are all the well-defined A^k. From the definition, we obtain that $A^{k_1} = A_0$, $T^{k_1} = T_0 = \emptyset$ and $k_p = K$. We additionally define $A^{k_{p+1}} = A_{\ell_1}$ and $T^{k_{p+1}} = T_{\ell_1}$. Then we have

$$T_{\ell_1} = T^{k_{p+1}} = \bigcup_{j=2}^{p+1} (T^{k_j} - T^{k_{j-1}}).$$

Now we shall give an estimation on the size of each $T^{k_j} - T^{k_{j-1}}$.

During the process, the algorithm ensures that $t^{k_j}(v) \leq \frac{\sqrt{n}}{\log^4 n}$, for every $v \in A^{k_j} \cup T^{k_j}$. For every $v \in A^{k_{j-1}} - (A^{k_j} \cup T^{k_j})$, suppose v was removed from $A^{k_{j-1}}$ in the i-th round and let u_i denote the selected vertex in the round. Then we obtain that

$$t^{k_j}(v) \le t_{i-1}(v) + d_{H_0}(v, u_i) \le \frac{\sqrt{n}}{\log^4 n} + d_{H_0}(v, u_i) \le \frac{\sqrt{n}}{\log^4 n} + \frac{8\sqrt{n}}{\log^4 n} = \frac{9\sqrt{n}}{\log^4 n},$$

where the last inequality is given by Lemma 3.7 and (3.8). Therefore, we have

$$\sum_{v \in A^{k_{j-1}}} t^{k_j}(v) \le \frac{9\sqrt{n}}{\log^4 n} |A^{k_{j-1}}| < \frac{9\sqrt{n}}{\log^4 n} \frac{n}{2^{k_{j-1}-1}}. \tag{3.9}$$

On the other hand, we can also estimate $\sum_{v \in A^{k_{j-1}}} t^{k_j}(v)$ from the view of 'selected' vertices. Let $2 \le j \le p$. Take a vertex $u_i \in T^{k_j} - T^{k_{j-1}}$ and suppose that u_i is selected in the i-th round, i.e., from A_i. Since A^{k_j} is the first set of size smaller than $\frac{n}{2^{k_j-1}}$, we have $|A_i| \ge \frac{n}{2^{k_j-1}}$ and then $|A_i||U_0| \ge 6\sqrt{n}\log n \cdot \frac{\sqrt{n}}{\log n} = 6n$. By Lemma 3.5, we obtain that

$$d_{H_0[A_i]}(u_i) \ge \frac{|A_i||U_0|^2}{12n} \ge \frac{n}{12 \cdot 2^{k_{j-1}} \log^2 n}.$$

Since $d_{H_0[A_i]}(u_i)$ does not contribute to $t^{k_j}(v)$ for $v \notin A^{k_{j-1}}$, we have

$$\sum_{v \in A^{k_{j-1}}} t^{k_j}(v) \ge \sum_{u_i \in T^{k_j} - T^{k_{j-1}}} d_{H_0[A_i]}(u_i) \ge \left|T^{k_j} - T^{k_{j-1}}\right| \frac{n}{12 \cdot 2^{k_{j-1}} \log^2 n}. \tag{3.10}$$

Combining (3.9) and (3.10), we obtain

$$\left|T^{k_j} - T^{k_{j-1}}\right| \le \frac{216\sqrt{n}}{\log^2 n} \quad \text{for } 2 \le j \le p.$$

Let $j = p+1$, since $\frac{n}{2^K} \le |A^{k_{p+1}}| \le |A^{k_p}| \le \frac{n}{2^{K-1}}$, by a similar argument, we obtain that

$$\left|T^{k_{p+1}} - T^{k_p}\right| \frac{|A^{k_{p+1}}||U_0|^2}{12n} \le \sum_{u_i \in T^{k_{p+1}} - T^{k_p}} d_{H_0[A_i]}(u_i) \le \sum_{v \in A^{k_p}} t^{k_{p+1}}(v) \le 9\frac{\sqrt{n}}{\log^4 n}|A^{k_p}|,$$

which gives

$$\left|T^{k_{p+1}} - T^{k_p}\right| \le \frac{216 n^{3/2}}{\log^4 n |U_0|^2} = \frac{216\sqrt{n}}{\log^2 n}.$$

We eventually have

$$|T_{\ell_1}| = \bigcup_{j=2}^{p+1} |T^{k_j} - T^{k_{j-1}}| \le \frac{1}{2}\log n \cdot \frac{216\sqrt{n}}{\log^2 n} = \frac{108\sqrt{n}}{\log n}.$$

For Phase 1, we define:

$$R_1 = T_{\ell_1}, \qquad C_1 = A_{\ell_1}.$$

Phase 2. If $|C_1 \cap I| < 12\frac{n}{|C_1|}$ or $|C_1| \leq 12\sqrt{n}$, we stop the algorithm with $L = 1$. Otherwise, take an arbitrary set $U_1 \subseteq C_1 \cap I$ of size $12\frac{n}{|C_1|}$ and denote $H_1 = H^{U_1}(C_1)$. We will use H_1 as an auxiliary graph and run the core algorithm with $t_{\text{threshold}} = \frac{\sqrt{n}}{\log^3 n}$ and initial state

$$A_0 = C_1, \quad T_0 = \emptyset \quad \text{and} \quad t_0(v) = 0, \text{ for every } v \in A_0,$$

until we obtain the set A_{ℓ_2}, the first set of size smaller than $|C_1|/2$.

We use a similar argument as in Phase 1. For every $v \in A_{\ell_2} \cup T_{\ell_2}$, the algorithm ensures that $t_{\ell_2}(v) \leq \frac{\sqrt{n}}{\log^3 n}$. For every $v \in A_0 - (A_{\ell_2} \cup T_{\ell_2})$, suppose v was removed from A_0 in the i-th round and let u_i denote the selected vertex in the round. Then using Lemma 3.7, we obtain that

$$t_{\ell_2}(v) \leq t_{i-1}(v) + d_{H_1}(v, u_i) \leq 2\frac{\sqrt{n}}{\log^3 n}.$$

Therefore, we have

$$\sum_{v \in A_0} t_{\ell_2}(v) \leq 2\frac{\sqrt{n}}{\log^3 n}|A_0| = 2\frac{\sqrt{n}}{\log^3 n}|C_1|. \tag{3.11}$$

On the other hand, take a vertex $u_i \in T_{\ell_2}$ and suppose that u_i is selected in the i-th round, i.e., from A_i. Since A_{ℓ_2} is the first set of size smaller than $|C_1|/2$, we have $|A_i| \geq |C_1|/2$ and then $|A_i||U_1| \geq \frac{|C_1|}{2} \cdot 12\frac{n}{|C_1|} = 6n$. From Lemma 3.5, we obtain that

$$d_{H_1[A_i]}(u_i) \geq \frac{|A_i||U_1|^2}{12n} \geq \frac{|C_1||U_1|^2}{24n}.$$

Consequently, we have

$$\sum_{v \in A_0} t_{\ell_2}(v) = \sum_{u_i \in T_{\ell_2}} d_{H_1[A_i]}(u_i) \geq |T_{\ell_2}| \frac{|C_1||U_1|^2}{24n}. \tag{3.12}$$

Combining (3.11) and (3.12), we obtain

$$|T_{\ell_2}| \leq \frac{48n^{3/2}}{\log^3 n|U_1|^2} = \frac{48n^{3/2}|C_1|^2}{\log^3 n \cdot 12^2 n^2} \leq \frac{48n^{3/2}(6\sqrt{n}\log n)^2}{\log^3 n \cdot 12^2 n^2} = \frac{12\sqrt{n}\log^2 n}{\log^3 n} \leq \frac{12\sqrt{n}}{\log n}.$$

For Phase 2, we define:

$$R_2 = T_{\ell_2}, \qquad C_2 = A_{\ell_2}.$$

Phase j for j $\geq$ 3. In general, when the algorithm goes to Phase j, we first check if $|C_{j-1} \cap I| < 12\frac{n}{|C_{j-1}|}$ or $|C_{j-1}| \leq 12\sqrt{n}$. If one of these conditions holds, we stop the algorithm with $L = j-1$. Otherwise, take an arbitrary set $U_{j-1} \subseteq C_{j-1} \cap I$ of size $12\frac{n}{|C_{j-1}|}$ and denote $H_{j-1} = H^{U_{j-1}}(C_{j-1})$. We will use H_{j-1} as an auxiliary graph and run the core algorithm with $t_{\text{threshold}} = \frac{\sqrt{n}}{\log^3 n}$ and initial state

$$A_0 = C_{j-1}, \quad T_0 = \emptyset \quad \text{and} \quad t_0(v) = 0, \text{ for every } v \in C_{j-1},$$

until we obtain the set A_{ℓ_j}, the first set of size smaller than $|C_{j-1}|/2$. Using the exactly same argument as in Phase 2, in the end, we obtain

$$|T_{\ell_j}| \leq \frac{48n^{3/2}}{\log^3 n|U_{j-1}|^2} \leq \frac{48n^{3/2}|C_{j-1}|^2}{\log^3 n \cdot 12^2 n^2} \leq \frac{1}{2^{2j-4}} \cdot \frac{48n^{3/2}|C_1|^2}{\log^3 n \cdot 12^2 n^2} \leq \frac{1}{2^{2j-4}} \cdot \frac{12\sqrt{n}}{\log n}.$$

For Phase j, we define:

$$R_j = T_{\ell_j}, \qquad C_j = A_{\ell_j}.$$

The algorithm terminates if any of the stopping rules is satisfied. In the process, we obtain set sequences $\{R_0, R_1, R_2, \cdots, R_L\}$, $\{U_0, U_1, U_2, \ldots, U_{L-1}\}$ and $\{C_0, C_1, C_2, \ldots, C_L\}$, which satisfy Conditions (ii)–(vii). From the stopping rules, we know that $12\sqrt{n} < |C_{L-1}| \leq \frac{6\sqrt{n}\log n}{2^{L-2}}$, which implies $L < \log\log n + 1$.

It remains to check Condition (i). For every $j \geq 0$, if a vertex v was removed in Phase j, then there exists i such that $t_i(v) > t_{\text{threshold}}$. This implies that there are more than $t_{\text{threshold}}$ Sidon 4-tuples containing v in I. By the choices of $t_{\text{threshold}}$ and R_0, we know that v does not belong to I, and Condition (i) follows from it. $\square$

Remark 3.13. *In Case 2, we aim to find a set satisfying inequalities (3.7) and (3.8). For this reason, when we apply the probabilistic method, we need consider the random subset $W \subseteq I$ with the probability $2/\sqrt{\log n}$. On the other hand, the proof requires the size of W to be large enough, i.e., $\sqrt{n}/\log n$. Therefore, it is necessary to assume that $|I| \geq \sqrt{n}/\sqrt{\log n}$.*

Now, let us assume that the set I satisfies $|\{v \in I : s_I(v) \geq \sqrt{n}/\log^4 n\}| \leq \sqrt{n}/\log n$. In regard to this assumption, Case 1 always works for the initial certificate $\{R_0, C_0\}$. This means that when we go through the previous proof under the new assumption, we can actually skip Case 2, where the assumption '$|I| \geq \sqrt{n}/\sqrt{\log n}$' is needed, and let everything else follow in the same way. As a result, we obtain a lemma similar to Lemma 3.12.

Lemma 3.14. *For a sufficiently large integer n, let $\alpha = n/\log^4 n$ and I be an α-generalized Sidon subset of $[n]$ such that for every $v \in I$, $s_I(v) < \sqrt{n}/\log^3 n$. Further assume that $|\{v \in I : s_I(v) \geq \sqrt{n}/\log^4 n\}| \leq \sqrt{n}/\log n$. Then there exist set sequences $R_0, R_1, \ldots, R_L$ and $U_0, U_1, \ldots, U_{L-1}$, where $0 \leq L < \log\log n + 1$, which determine a unique set sequence $C_0 \supset C_1 \supset C_2 \supset \ldots \supset C_L$. Furthermore, Conditions (i)–(vii) from Lemma 3.12 are all satisfied.*

3.4 Counting generalized Sidon sets

Proof of Theorem 3.1. Since the number of sets in $[n]$ of size at most $\frac{\sqrt{n}}{\log n}$ is bounded by $2^{\sqrt{n}}$, it is sufficient to count the sets of size at least $\frac{\sqrt{n}}{\sqrt{\log n}}$. For every $I \in \mathcal{I}_n(\alpha)$, we iteratively remove a number v from I, which has $s_I(v) \geq \frac{\sqrt{n}}{\log^3 n}$. Denote by I' the set of remaining numbers. Since I contains at most $\frac{n}{\log^4 n}$ Sidon 4-tuples, the process stops after at most $\frac{\sqrt{n}}{\log n}$ steps, i.e.,

$$|I - I'| \leq \frac{\sqrt{n}}{\log n}. \tag{3.13}$$

This cleaning process ensures that $s_{I'}(v) < \frac{\sqrt{n}}{\log^3 n}$, for every $v \in I'$. By Lemma 3.12, I' can be associated to a certificate $\{\mathcal{R}, \mathcal{U}\}$, where $\mathcal{R} = \{R_0, R_1, \ldots, R_L\}$ and $\mathcal{U} = \{U_0, U_1, \ldots, U_{L-1}\}$ are two set sequences satisfying Conditions (i)–(vii) in Lemma 3.12. Thus, each $I \in \mathcal{I}_n(\alpha)$ can be assigned to a certificate

$$\mathcal{C}_I = [I - I', L, \mathcal{R}, \mathcal{U}].$$

Note that different sets could have the same certificate. Therefore, to estimate $|\mathcal{I}_n(\alpha)|$, we need to give upper bounds on the number of certificates and on the number of subsets assigned to one certificate.

Let $\mathcal{C} = \{\mathcal{C}_I = [I - I', L, \mathcal{R}, \mathcal{U}] \mid I \in \mathcal{I}_n(\alpha)\}$. For every integer $\ell \geq 0$, denote by $\mathcal{C}_\ell$ the set of certificates in $\mathcal{C}$ with $L = \ell$. By Lemma 3.12, we have

$$\mathcal{C} = \bigcup_{\ell=0}^{\log\log n+1} \mathcal{C}_\ell. \tag{3.14}$$

For $\ell = 0$ and a certificate $[I - I', 0, \mathcal{R}, \mathcal{U}] \in \mathcal{C}_0$, $\mathcal{U}$ is empty sequence and $\mathcal{R}$ only contains one set, i.e. $\mathcal{R} = \{R_0\}$. By Lemma 3.12 and (3.13), R_0 and $I - I'$ are subsets of $[n]$ satisfying $|R_0| \leq \frac{16\sqrt{n}}{\log n}$ and

$|I - I'| \le \frac{\sqrt{n}}{\log n}$ respectively. Therefore, the number of certificates in $\mathcal{C}_0$ is

$$|\mathcal{C}_0| \le \sum_{i=0}^{\frac{\sqrt{n}}{\log n}} \binom{n}{i} + \sum_{i=0}^{\frac{16\sqrt{n}}{\log n}} \binom{n}{i} \le 2^{16\sqrt{n}+1}. \tag{3.15}$$

For $1 \le \ell \le \log\log n+1$ and a certificate $[I-I', \ell, \mathcal{R}, \mathcal{U}] \in \mathcal{C}_\ell$, $\mathcal{R}, \mathcal{U}$ can be written as $\mathcal{R} = \{R_0, R_1, \ldots, R_\ell\}$ and $\mathcal{U} = \{U_0, U_1, \ldots, U_{\ell-1}\}$. Similarly, since $I - I' \subseteq [n]$ and $|I - I'| \le \frac{\sqrt{n}}{\log n}$, the number of ways to choose $I - I'$ is at most

$$\sum_{i=0}^{\frac{\sqrt{n}}{\log n}} \binom{n}{i} \le 2\binom{n}{\frac{\sqrt{n}}{\log n}} \le n^{\frac{\sqrt{n}}{\log n}} = 2^{\sqrt{n}}.$$

Now, we discuss the number of choices for sequences $\mathcal{U} = \{U_0, U_1, \ldots, U_{\ell-1}\}$ and $\mathcal{R} = \{R_0, R_1, \ldots, R_\ell\}$ iteratively. First, by Condition (iii) in Lemma 3.12, we have $R_0 \subseteq [n]$ and $|R_0| \le \frac{16\sqrt{n}}{\log n}$. Thus, the number of ways to choose R_0 is at most

$$\sum_{i=0}^{\frac{16\sqrt{n}}{\log n}} \binom{n}{i} \le 2\binom{n}{\frac{16\sqrt{n}}{\log n}} \le 2^{16\sqrt{n}}.$$

From the proof of Lemma 3.12, $I - I'$ and R_0 determines a unique set C_0 of size at most n. By Conditions (iv) and (v) in Lemma 3.12, we obtain that $U_0 \subseteq [n]$, $R_1 \subseteq C_0$, $|U_0| = \frac{\sqrt{n}}{\log n}$ and $|R_1| \le \frac{108\sqrt{n}}{\log n}$. Thus, the number of ways to choose U_0 and R_1 are at most

$$\binom{n}{\frac{\sqrt{n}}{\log n}} \le 2^{\sqrt{n}}$$

and

$$\sum_{i=0}^{\frac{108\sqrt{n}}{\log n}} \binom{n}{i} \le 2\binom{n}{\frac{108\sqrt{n}}{\log n}} \le 2^{108\sqrt{n}}$$

respectively. For every $1 \le i \le \ell-1$, suppose that sets $I - I'$, $R_0, \ldots, R_i$, and $U_0, \ldots, U_{i-1}$ are already fixed. The proof of Lemma 3.12 shows that there is a unique set C_i such that $U_i, R_{i+1} \subseteq C_i \subseteq [n]$. Moreover, there exists an integer z_i such that

$$\frac{6\sqrt{n}\log n}{2^{z_i}} < |C_i| \le \frac{6\sqrt{n}\log n}{2^{z_i-1}},$$

where $1 \le z_1 < \ldots < z_{i-1} < z_i < \log\log n$. By Conditions (iv) and (vi) in Lemma 3.12, we obtain that $|U_i| = 12\frac{n}{|C_i|}$ and $|R_{i+1}| \le \frac{1}{2^{2i-2}} \frac{12\sqrt{n}}{\log n}$. Thus, the number of ways to choose U_i and R_{i+1} are at most

$$\binom{|C_i|}{12n/|C_i|} \le \binom{\frac{6\sqrt{n}\log n}{2^{z_i-1}}}{\frac{12n\cdot 2^{z_i}}{6\sqrt{n}\log n}}$$

and

$$\sum_{i=0}^{\frac{1}{2^{2i-2}}\frac{12\sqrt{n}}{\log n}} \binom{n}{i} \leq 2\binom{n}{\frac{1}{2^{2i-2}}\frac{12\sqrt{n}}{\log n}} \leq 2^{\frac{12}{2^{2i-2}}\sqrt{n}}$$

respectively. We summarize the above discussion and obtain that

$$\begin{aligned} |\mathcal{C}_l| &\leq 2^{\sqrt{n}+16\sqrt{n}+\sqrt{n}+108\sqrt{n}} \cdot \prod_{i=1}^{\infty} 2^{\frac{12}{2^{2i-2}}\sqrt{n}} \sum_{z_1,\ldots,z_{\ell-1}} \binom{\frac{6\sqrt{n}\log n}{2^{z_1-1}}}{\frac{12n\cdot 2^{z_1}}{6\sqrt{n}\log n}} \binom{\frac{6\sqrt{n}\log n}{2^{z_2-1}}}{\frac{12n\cdot 2^{z_2}}{6\sqrt{n}\log n}} \cdots \binom{\frac{6\sqrt{n}\log n}{2^{z_{\ell-1}-1}}}{\frac{12n\cdot 2^{z_{\ell-1}}}{6\sqrt{n}\log n}} \\ &\leq 2^{142\sqrt{n}} \cdot \sum_{z_1,\ldots,z_{\ell-1}} \binom{\frac{6\sqrt{n}\log n}{2^{z_1-1}}}{\frac{12n\cdot 2^{z_1}}{6\sqrt{n}\log n}} \binom{\frac{6\sqrt{n}\log n}{2^{z_2-1}}}{\frac{12n\cdot 2^{z_2}}{6\sqrt{n}\log n}} \cdots \binom{\frac{6\sqrt{n}\log n}{2^{z_{\ell-1}-1}}}{\frac{12n\cdot 2^{z_{\ell-1}}}{6\sqrt{n}\log n}}, \end{aligned} \tag{3.16}$$

where $z_1 < z_2 < \ldots < z_{\ell-1}$ take over integers in $[1,\ \log\log n)$. To estimate the summation term in inequality (3.16), we provide the following claim.

Claim 3.15. *For sufficiently large n, we have*

$$\binom{6\sqrt{n}\log n}{\frac{24n}{6\sqrt{n}\log n}} \binom{\frac{6\sqrt{n}\log n}{2}}{\frac{24n\cdot 2}{6\sqrt{n}\log n}} \cdots \binom{\frac{6\sqrt{n}\log n}{2^{\log\log n-1}}}{\frac{24n\cdot 2^{\log\log n-1}}{6\sqrt{n}\log n}} \leq 2^{25\sqrt{n}}.$$

Proof. Let $x = \frac{6\sqrt{n}\log n}{2^{\log\log n-1}} = 12\sqrt{n}$. Then the left side is equal to

$$\begin{aligned} \prod_{i=0}^{\log\log n-1} \binom{2^i x}{\frac{24n}{2^i x}} &\leq \prod_{i=0}^{\log\log n-1} \left(\frac{e\cdot x^2 2^{2i}}{24n}\right)^{\frac{24n}{2^i x}} \leq \prod_{i=0}^{\log\log n-1} (6e2^{2i})^{\frac{2}{2^i}\sqrt{n}} \\ &\leq \prod_{i=0}^{\log\log n-1} 2^{\left[(\log 6e+2i)\frac{2}{2^i}\right]\sqrt{n}} \leq 2^{\left[\sum_{i=0}^{\infty}(\log 6e+2i)\frac{2}{2^i}\right]\sqrt{n}} \\ &= 2^{(4\log 6e+8)\sqrt{n}} \leq 2^{25\sqrt{n}}, \end{aligned}$$

where the first inequality follows from the Stirling's formula. $\square$

Using Claim 3.15, we can show that for every $1 \leq \ell \leq \log\log n + 1$,

$$|\mathcal{C}_\ell| \leq 2^{142\sqrt{n}} \cdot \binom{\log\log n}{\ell-1} 2^{25\sqrt{n}} \leq 2^{167\sqrt{n}} \log n. \tag{3.17}$$

Combining (3.15) and (3.17), we obtain

$$|\mathcal{C}| = \sum_{\ell=0}^{\log\log n+1} |\mathcal{C}_\ell| \leq 2^{16\sqrt{n}+1} + 2^{167\sqrt{n}}(\log\log n+1)\log n \leq 2^{168\sqrt{n}}. \tag{3.18}$$

It remains to give an upper bound on the number of subsets assigned to one certificate. For a certificate

$C = [I - I', L, \mathcal{R}, \mathcal{U}] \in \mathcal{C}$, let $\mathcal{I}_C = \{I \in \mathcal{I}_n(\alpha) \mid C_I = C\}$. For every $I \in \mathcal{I}_C$, by Lemma 3.12, we have

$$I \subseteq (I - I') \cup \bigcup_{i=0}^{L} R_i \cup C_L,$$

where C_L is uniquely determined. Note that the set $(I - I') \cup \bigcup_{i=0}^{L} R_i$ is given by the certificate C. Therefore, $\mathcal{I}_C$ is decided by the ways to choose $C_L \cap I = C_L \cap I'$. There are three cases:

Case 1: $|C_L| \leq 12\sqrt{n}$.

In the case, we have $|\mathcal{I}_C| \leq 2^{|C_L|} \leq 2^{12\sqrt{n}}$.

Case 2: $L = 0$ and $|C_L| > 12\sqrt{n}$.

By Condition (vii) in Lemma 3.12, for every $I \in \mathcal{I}_C$, I satisfies $|C_0 \cap I'| < \frac{\sqrt{n}}{\log n}$. In this case, we have

$$|\mathcal{I}_C| \leq \sum_{i=0}^{\frac{\sqrt{n}}{\log n}} \binom{|C_0|}{i} \leq \sum_{i=0}^{\frac{\sqrt{n}}{\log n}} \binom{n}{i} \leq 2\binom{n}{\frac{\sqrt{n}}{\log n}} \leq 2^{\sqrt{n}}.$$

Case 3: $L \geq 1$ and $|C_L| > 12\sqrt{n}$.

By Condition (vii) in Lemma 3.12, for every $I \in \mathcal{I}_C$, I satisfies $|C_L \cap I'| < 12\frac{n}{|C_L|}$. In this case, we have

$$|\mathcal{I}_C| \leq \sum_{i=0}^{12\frac{n}{|C_L|}} \binom{|C_L|}{i} \leq 2\binom{|C_L|}{12\frac{n}{|C_L|}}.$$

Let $x = 12\frac{n}{|C_L|}$. By convexity, we obtain that

$$|\mathcal{I}_C| \leq 2\binom{12n/x}{x} \leq 2\left(\frac{12en}{x^2}\right)^x \leq 2^{[\log(12en) - 2\log x]x + 1} \leq 2^{\sqrt{12en}+1} \leq 2^{6\sqrt{n}}.$$

From the above discussion, for every $C \in \mathcal{C}$, we have

$$|\mathcal{I}_C| \leq 2^{12\sqrt{n}}. \tag{3.19}$$

Eventually, combining (3.18) and (3.19), we obtain that

$$|\mathcal{I}_n(\alpha)| \leq |\mathcal{C}| \cdot 2^{12\sqrt{n}} \leq 2^{180\sqrt{n}}. \qquad \square$$

Proof of Theorem 3.2: For every set $J \in \mathcal{J}_n(\alpha)$, we apply the same cleaning process as in the proof of Theorem 3.1 and obtain a set J' satisfying $|J - J'| \le \frac{\sqrt{n}}{\log n}$ and $s_{J'}(v) < \frac{\sqrt{n}}{\log^3 n}$ for every $v \in J'$. Due to the definition of $\mathcal{J}_n(\alpha)$ and $J' \subseteq J$, we also have $|\{v \in J' : s_{J'}(v) \ge \frac{\sqrt{n}}{\log^4 n}\}| \le \frac{\sqrt{n}}{\log n}$. By Lemma 3.14, J' can be associated to a certificate $\{\mathcal{R},\mathcal{U}\}$, where $\mathcal{R} = \{R_0, R_1, \ldots, R_L\}$ and $\mathcal{U} = \{U_0, U_1, \ldots, U_{L-1}\}$ are two set sequences satisfying Conditions (i)–(vii) in Lemma 3.12. The rest of the proof is the same as that of Theorem 3.1. $\square$

3.5 Concluding remarks

In [94], Saxton and Thomason established the hypergraph container theorem not only covering independent sets but also for sufficiently sparse structures. One can use their result to estimate the number of α-generalized sets for some functions α; however, the estimates obtained from it are weaker than the ones from the graph container method. To be more specific, using the hypergraph container method, we would consider the 4-uniform hypergraph whose vertex set is $[n]$ and whose edges are all the Sidon 4-tuples; to generate small containers, we need to iterate Theorem 6.2 ([94]) repeatedly $\Theta(\log n)$ times. This produces $2^{O(n\tau \log(1/\tau)\log n)}$ containers of size at most $O(n\tau)$, for the sets with at most $O(\tau^4 n^3)$ Sidon 4-tuples. Since we are interested in obtaining a family of containers with $2^{\Theta(\sqrt{n})}$ elements, the order of τ should not be higher than $1/(\sqrt{n}\log^2 n)$. (One can easily check that $\tau = \Theta(1/\sqrt{n}\log^2 n)$ satisfies the conditions of Theorem 6.2.) Therefore, the hypergraph container theorem in [94] provides that the number of α-generalized Sidon is $2^{O(\sqrt{n})}$ for $\alpha = O(\tau^4 n^3) = O(n/\log^8 n)$, while the best result we have is for $\alpha = O(n/\log^5 n)$.

We also studied the family of α-generalized Sidon sets for some other functions α. Denote by $\mathcal{G}_n(\alpha)$ the family of α-generalized Sidon sets in $[n]$. The results we have is summarized in the following table.

α	Upper bound for $\lvert\mathcal{G}_n(\alpha)\rvert$	Lower bound for $\lvert\mathcal{G}_n(\alpha)\rvert$
$n/\log^5 n$	$2^{O(\sqrt{n})}$	$2^{\Omega(\sqrt{n})}$
$n/\log^4 n$	$2^{O(\sqrt{n}\log^{1/4} n))}$	$2^{\Omega(\sqrt{n})}$
$n/\log^3 n$	$2^{O(\sqrt{n}\sqrt{\log n})}$	$2^{\Omega(\sqrt{n}\log^{1/4} n)}$
$n/\log^2 n$	$2^{O(\sqrt{n}\log^{3/4} n)}$	$2^{\Omega(\sqrt{n}\sqrt{\log n})}$
$n/\log n$	$2^{O(\sqrt{n}\log n)}$	$2^{\Omega(\sqrt{n}\log^{3/4} n)}$
n	$2^{O(\sqrt{n}\log n)}$	$2^{\Omega(\sqrt{n}\log n)}$

Table 3.1: The number of α-generalized Sidon sets.

In Table 3.1, all the lower bounds come from the probabilistic argument discussed in Section 3.1, except for the case $\alpha \le n/\log^4 n$, where we use the number of Sidon sets as the lower bound; all the upper bounds follow from our graph container method, except for the case $\alpha = n$, where we use Corollary 3.6. For

$\alpha \in \{n/\log^5 n, n\}$, the current bounds are tight. For other α, the distance between the lower bound and the upper bound is a $\log^{1/4} n$ factor on the exponent. We believe that the lower bounds are the truth.

Chapter 4

The typical structure of Gallai colorings and their extremal graphs

4.1 Introduction

An edge coloring of a graph G is a *Gallai coloring* if it contains no rainbow triangle, that is, no triangle is colored with three distinct colors. The term *Gallai coloring* was first introduced by Gyárfás and Simonyi [53], but this concept had already occurred in an important result of Gallai [49] on comparability graphs, which can be reformulated in terms of Gallai colorings. It also turns out that Gallai colorings are relevant to generalizations of the perfect graph theorem [24], and some applications in information theory [74]. There are a variety of papers which consider structural and Ramsey-type problems on Gallai colorings, see, e.g., [44, 51, 52, 53, 99].

Two important themes in extremal combinatorics are to enumerate discrete structures that have certain properties and describe their typical properties. In this chapter, we shall be concerned with Gallai colorings from such an extremal perspective.

4.1.1 Gallai colorings of complete graphs

For an integer $r \geq 3$, an *r-coloring* is an edge coloring that uses at most r colors. By choosing two of the r colors and coloring the edges of K_n arbitrarily with these two colors, one can easily obtain that the number of Gallai r-colorings of K_n is at least

$$\binom{r}{2}\left(2^{\binom{n}{2}}-2\right)+r=\binom{r}{2}2^{\binom{n}{2}}-r(r-2). \tag{4.1}$$

If we further consider all Gallai r-colorings of K_n using exactly 3 colors, red, green, and blue, in which the red color is used only once, the number of them is exactly

$$\binom{n}{2}\left(2^{\binom{n}{2}-(n-1)}-2\right).$$

Combining with (4.1), for n sufficiently large, a trivial lower bound for the number of Gallai r-colorings of K_n is

$$\left(\binom{r}{2} + 2^{-n}\right) 2^{\binom{n}{2}}. \tag{4.2}$$

Motivated by a question of Erdős and Rothschild [34] and the resolution by Alon, Balogh, Keevash and Sudakov [2], Benevides, Hoppen and Sampaio [19] studied the general problem of counting the number of edge colorings of a graph that avoid a subgraph colored with a given pattern. In particular, they proved that the number of Gallai 3-colorings of K_n is at most $\frac{3}{2}(n-1)! \cdot 2^{\binom{n-1}{2}}$. At the same time, Falgas-Ravry, O'Connell, and Uzzell [42] provided a weaker upper bound of the form $2^{(1+o(1))\binom{n}{2}}$, which is a consequence of the multi-color container theory. Bastos, Benevides, Mota and Sau [29] later improved the upper bound to $7(n+1)2^{\binom{n}{2}}$. Note that the gap between the best upper bound and the trivial lower bound is a linear factor. We show that the lower bound is indeed closer to the truth, and this actually applies for any integer r. Our first main result is as follows.

Theorem 4.1. *For every integer $r \geq 3$, there exists n_0 such that for all $n > n_0$, the number of Gallai r-colorings of the complete graph K_n is at most*

$$\left(\binom{r}{2} + 2^{-\frac{n}{4\log^2 n}}\right) 2^{\binom{n}{2}}.$$

Given a class of graphs $\mathcal{A}$, we denote $\mathcal{A}_n$ the set of graphs in $\mathcal{A}$ of order n. We say that *almost all graphs in $\mathcal{A}$ has property $\mathcal{B}$* if

$$\lim_{n\to\infty} \frac{|\{G \in \mathcal{A}_n : G \text{ has property } \mathcal{B}\}|}{|\mathcal{A}_n|} = 1.$$

Recall that the number of Gallai r-colorings with at most 2 colors is $\binom{r}{2}2^{\binom{n}{2}} - r(r-2)$. Then the description of the typical structure of Gallai r-colorings immediately follows from Theorem 4.1.

Corollary 4.2. *For every integer $r \geq 3$, almost all Gallai r-colorings of the complete graph are 2-colorings.*

4.1.2 The extremal graphs of Gallai colorings

There have been considerable advances in edge coloring problems whose origin can be traced back to a question of Erdős and Rothschild [34], who asked which n-vertex graph admits the largest number of r-colorings avoiding a copy of F with a prescribed colored pattern, where r is a positive integer and F is a fixed graph. In particular, the study for the extremal graph of Gallai colorings, that is the case when F is a triangle with rainbow pattern, has received attention recently. A graph G on n vertices is *Gallai r-extremal* if the number of Gallai r-colorings of G is largest over all graphs on n vertices. For $r \geq 5$, the

Gallai r-extremal graph has been determined by Hoppen, Lefmann and Odermann [61, 62, 63].

Theorem 4.3. *[62] For all $r \geq 10$ and $n \geq 5$, the only Gallai r-extremal graph of order n is the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.*

Theorem 4.4. *[62] For all $r \geq 5$, there exists n_0 such that for all $n > n_0$, the only Gallai r-extremal graph of order n is the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.*

For the cases $r \in \{3, 4\}$, several approximate results were given.

Theorem 4.5. *[19] There exists n_0 such that the following hold for all $n > n_0$.*

(i) For all $\delta > 0$, if G is a graph of order n, then the number of Gallai 3-colorings of G is at most $2^{(1+\delta)n^2/2}$.

(ii) For all $\xi > 0$, if G is a graph of order n, and $e(G) \leq (1-\xi)\binom{n}{2}$, then the number of Gallai 3-colorings of G is at most $2^{\binom{n}{2}}$.

We remark that the part (i) of Theorem 4.5 was also proved in [62], and the authors further provided an upper bound for $r = 4$.

Theorem 4.6. *[62] There exists n_0 such that the following hold for all $n > n_0$. For all $\delta > 0$, if G is a graph of order n, then the number of Gallai 4-colorings of G is at most $4^{(1+\delta)n^2/4}$.*

The above theorems show that for $r \in \{3, 4\}$, the complete graph K_n is not far from being Gallai r-extremal, while for $r = 4$, the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ is also close to be Gallai r-extremal. Benevides, Hoppen and Sampaio [19] made the following conjecture.

Conjecture 4.7. *[19] The only Gallai 3-extremal graph of order n is the complete graph K_n.*

For the case $r = 4$, Hoppen, Lefmann and Odermann [62] believed that $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ should be the extremal graph.

Conjecture 4.8. *[62] The only Gallai 4-extremal graph of order n is the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.*

Using a similar technique as in Theorem 4.1, we prove an analogous result for dense non-complete graphs when $r = 3$.

Theorem 4.9. *For $0 < \xi \leq \frac{1}{64}$, there exists n_0 such that for all $n > n_0$ the following holds. If G is a graph of order n, and $e(G) \geq (1-\xi)\binom{n}{2}$, then the number of Gallai 3-colorings of G is at most*

$$3 \cdot 2^{e(G)} + 2^{-\frac{n}{4\log^2 n}} 2^{\binom{n}{2}}.$$

Together with Theorem 4.5 and the lower bound (4.2), Theorem 4.9 solves Conjecture 4.7 for sufficiently large n.

Theorem 4.10. *There exists n_0 such that for all $n > n_0$, among all graphs of order n, the complete graph K_n is the unique Gallai 3-extremal graph.*

Our third contribution is the following theorem.

Theorem 4.11. *For $n, r \in \mathbb{N}$ with $r \geq 4$, there exists n_0 such that for all $n > n_0$ the following holds. If G is a graph of order n, and $e(G) > \lfloor n^2/4 \rfloor$, then the number of Gallai r-colorings of G is less than $r^{\lfloor n^2/4 \rfloor}$.*

We remark that for a graph G with $e(G) = \lfloor n^2/4 \rfloor$, which is not $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$, G contains at least one triangle. Therefore, the number of Gallai r-colorings of G is at most $r(r + 2(r-1))r^{e(G)-3} < r^{\lfloor n^2/4 \rfloor}$. As a direct consequence of Theorem 4.11 and the above remark, we reprove Theorem 4.4, and in particular, we show that Conjecture 4.8 is true for sufficiently large n.

Theorem 4.12. *There exists n_0 such that for all $n > n_0$, among all graphs of order n, the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ is the unique Gallai 4-extremal graph.*

4.1.3 Organization of the chapter

Combining Szemerédi's Regularity Lemma and the stability method was used at many earlier works on extremal problems, including Erdős-Rothchild type problems, see, e.g., [2, 6, 19, 62]. However, our main approach relies on the method of hypergragh containers, developed independently by Balogh, Morris and Samotij [11] as well as by Saxton and Thomason [94], and some stability results for containers, which may be of independent interest to readers.

This chapter is organized as follows. First, in Section 4.2, we introduce some important definitions and then state a container theorem which is applicable to colorings. In Section 4.3, we present a key enumeration result on the number of colorings with special restrictions, which will be used repeatedly in the rest of the chapter. Then in Section 4.4, we study the stability behavior of the containers for the complete graph, and apply the multicolor container theorem to give an asymptotic upper bound for the number of Gallai r-colorings of the complete graph. In Section 4.5, we deal with the Gallai 3-colorings of dense non-complete graphs; the idea is the same as in Section 4.4 except that we need to provide a new stability result which is applicable to non-complete graphs.

In the second half of this chapter, that is, in Section 4.6, we study the Gallai r-colorings of non-complete graphs for $r \geq 4$. When the underlying graph is very dense, that is, close to the complete graph, we apply the same strategy as in Section 4.4 for the case $r = 4$, where we prove a proper stability result for containers. The case $r \geq 5$ is even simpler, in which we actually prove that the number of Gallai colorings in each container is small enough. When the underlying graph has edge density close to the $\frac{1}{4}$, i.e. the edge density of the extremal graph, some new ideas are needed, and we also adopt a result of Bollobás and Nikiforov [22] on book graphs. For the rest of the graphs whose edge densities are between $\frac{1}{4} + o(1)$ and $\frac{1}{2} - o(1)$, we use a supersaturation result of triangle-free graphs given by Balogh, Bushaw, Collares, Liu, Morris, and Sharifzadeh [7], and the above results on Gallai r-colorings for both high density graphs and low density graphs.

4.2 Preliminaries

4.2.1 The hypergraph container theorem

We use the following version of the hypergraph container theorem (Theorem 3.1 in [14]). Let $\mathcal{H}$ be a k-uniform hypergraph with average degree d. The *co-degree* of a set of vertices $S \subseteq V(\mathcal{H})$ is the number of edges containing S; that is,

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

For every integer $2 \leq j \leq k$, the j-th maximum co-degree of $\mathcal{H}$ is

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

When the underlying hypergraph is clear, we simply write it as Δ_j. For $0 < \tau < 1$, the *co-degree function* $\Delta(\mathcal{H}, \tau)$ is defined as

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

In particular, when $k = 3$,

$$\Delta(\mathcal{H}, \tau) = \frac{4\Delta_2}{d\tau} + \frac{2\Delta_3}{d\tau^2}.$$

Theorem 4.13. *[14] Let $\mathcal{H}$ be a k-uniform hypergraph on vertex set $[N]$. Let $0 < \varepsilon, \tau < 1/2$. Suppose that $\tau < 1/(200k!^2k)$ and $\Delta(\mathcal{H}, \tau) \leq \varepsilon/(12k!)$. Then there exists $c = c(k) \leq 1000k!^3k$ 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 $A \in \mathcal{C}$, $e(\mathcal{H}[A]) \leq \varepsilon \cdot e(\mathcal{H})$*;*

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

4.2.2 Definitions and multi-color container theorem

A key tool in applying container theory to multi-colored structures will be the notion of a *template*. This notion of 'template', which was first introduced in [42], goes back to [94] under the name of '2-colored multigraphs' and later to [15], where it is simply called 'containers'. For more studies about the multi-color container theory, we refer the interested reader to [11, 12, 15, 42, 94].

Definition 4.14 (Template and palette)**.** *An* r*-template of order* n *is a function* $P : E(K_n) \to 2^{[r]}$*, associating to each edge* e *of* K_n *a list of colors* $P(e) \subseteq [r]$*; we refer to this set* $P(e)$ *as the palette available at* e*.*

Definition 4.15 (Subtemplate)**.** *Let* P_1*,* P_2 *be two* r*-templates of order* n*. We say that* P_1 *is a subtemplate of* P_2 *(written as* $P_1 \subseteq P_2$*) if* $P_1(e) \subseteq P_2(e)$ *for every edge* $e \in E(K_n)$*.*

We observe that for $G \subseteq K_n$, an r-coloring of G can be considered as an r-template of order n, with only one color allowed at each edge of G and no color allowed at each non-edge. For an r-template P, write $\mathrm{RT}(P)$ for the number of subtemplates of P that are rainbow triangles. We say that P is *rainbow triangle-free* if $\mathrm{RT}(P) = 0$. Using the container method, Theorem 4.13, we obtain the following.

Theorem 4.16. *For every* $r \geq 3$*, there exists a constant* $c = c(r)$ *and a collection* $\mathcal{C}$ *of* r*-templates of order* n *such that*

(i) every rainbow triangle-free r*-template of order* n *is a subtemplate of some* $P \in \mathcal{C}$*;*

(ii) for every $P \in \mathcal{C}$*,* $\mathrm{RT}(P) \leq n^{-1/3}\binom{n}{3}$*;*

(iii) $|\mathcal{C}| \leq 2^{cn^{-1/3}\log^2 n\binom{n}{2}}$.

Proof. Let $\mathcal{H}$ be a 3-uniform hypergraph with vertex set $E(K_n) \times \{1, 2, \ldots, r\}$, whose edges are all triples $\{(e_1, d_1), (e_2, d_2), (e_3, d_3)\}$ such that e_1, e_2, e_3 form a triangle in K_n and d_1, d_2, d_3 are all different. In other words, every hyperedge in $\mathcal{H}$ corresponds to a rainbow triangle of K_n. Note that there are exactly $r(r-1)(r-2)$ ways to rainbow color a triangle with r colors. Hence, the average degree d of $\mathcal{H}$ is equal to

$$d = \frac{3e(\mathcal{H})}{v(\mathcal{H})} = \frac{3r(r-1)(r-2)\binom{n}{3}}{r\binom{n}{2}} = (r-1)(r-2)(n-2).$$

For the application of Theorem 4.13, let $\varepsilon = n^{-1/3}/r(r-1)(r-2)$ and $\tau = \sqrt{72 \cdot 3! \cdot r} n^{-1/3}$. Observe that $\Delta_2(\mathcal{H}) = r-2$, and $\Delta_3(\mathcal{H}) = 1$. For n sufficiently large, we have $\tau \leq 1/(200 \cdot 3!^2 \cdot 3)$ and

$$\Delta(\mathcal{H}, \tau) = \frac{4(r-2)}{d\tau} + \frac{2}{d\tau^2} \leq \frac{3}{d\tau^2} \leq \frac{\varepsilon}{12 \cdot 3!}.$$

Hence, there is a collection $\mathcal{C}$ of vertex subsets satisfying properties (i)-(iii) of Theorem 4.13. Observe that every vertex subset of $\mathcal{H}$ corresponds to an r-template of order n; every rainbow triangle-free r-template of order n corresponds to an independent set in $\mathcal{H}$. Therefore, $\mathcal{C}$ is a desired collection of r-templates. $\square$

Definition 4.17 (Gallai r-template)**.** *For a graph G of order n, an r-template P of order n is a Gallai r-template of G if it satisfies the following properties:*

(i) for every $e \in E(G)$, $|P(e)| \geq 1$;

(ii) $\mathrm{RT}(P) \leq n^{-1/3}\binom{n}{3}$.

For a graph G of order n and a collection $\mathcal{P}$ of r-templates of order n, denote by $\mathrm{Ga}(\mathcal{P}, G)$ the set of Gallai r-colorings of G which is a subtemplate of some $P \in \mathcal{P}$. If $\mathcal{P}$ consists of a single template P, then we simply write it as $\mathrm{Ga}(P, G)$.

4.2.3 A technical lemma

In this section, we provide a lemma that will be useful to us in what follows. We use a special case of the weak Kruskal-Katona theorem due to Lovàsz's [85].

Theorem 4.18 (Lovàsz [85])**.** *Suppose G is a graph with $\binom{x}{2}$ edges, for some real number $x \geq 2$. Then the number of triangles of G is at most $\binom{x}{3}$, with equality if and only if x is an integer and $G = K_x$.*

Lemma 4.19. *Let $n, r \in \mathbb{N}$ with $r \geq 3$ and $\frac{4}{n} - \frac{4}{n^2} \leq \varepsilon < \frac{1}{2}$. If G is an r-colored graph of order n, which contains at least $(1-\varepsilon)\binom{n}{3}$ monochromatic triangles, then there exists a color c such that the number of edges colored by c is at least $e(G) - 4r^2\varepsilon\binom{n}{2}$.*

Proof. We shall prove this lemma by contradiction. Let $\delta = 4r^2\varepsilon$. Assume that none of the colors is used on at least $e(G) - \delta\binom{n}{2}$ edges.

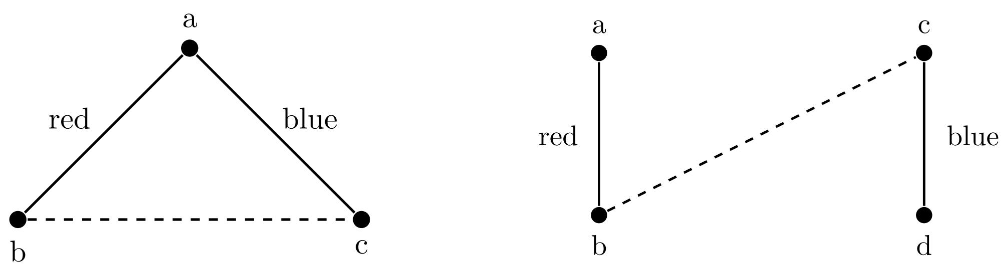
First, we conclude that $e(G) \geq (1-\varepsilon)\binom{n}{2}$. If not, then by Theorem 4.18, the number of triangles of G is less than

$$\frac{\sqrt{2}}{3}(1-\varepsilon)^{3/2}\binom{n}{2}^{3/2} \leq (1-\varepsilon)\binom{n}{3},$$

which contradicts the assumption.

By the pigeonhole principle, we can assume without loss of generality that the set of red edges in G, denoted by $\mathrm{R}(G)$, satisfies $|\mathrm{R}(G)| \geq (1-\varepsilon)\binom{n}{2}/r$. By the contradiction assumption, we have $|\mathrm{R}(G)| < e(G) - \delta\binom{n}{2}$. Therefore, the number of non-red edges is greater than $\delta\binom{n}{2}$. Again, without loss of generality, we can assume that the set of blue edges in G, denoted by $\mathrm{B}(G)$, satisfies $|\mathrm{B}(G)| \geq \delta\binom{n}{2}/r$.

For an edge in $\mathrm{R}(G)$ and an edge in $\mathrm{B}(G)$, these two edges either share one endpoint or are vertex disjoint, see Figure 4.1. In the first case, see Figure 4.1a, the triple abc could not form a monochromatic triangle of G. In the latter case, see Figure 4.1b, at least one of abc and bcd is not a monochromatic triangle of G.



(a) (b)

Figure 4.1: Two cases of a red-blue pair of edges.

Let $\mathrm{NT}(G)$ be the family of triples $\{a,b,c\}$ which does not form a monochromatic triangle of G. The above discussion shows that each pair of red and blue edges generates at least one triple in $\mathrm{NT}(G)$. Observe that each triple in $\mathrm{NT}(G)$ can be counted in at most $2+3(n-3)$ pairs of red and blue edges. Hence, we obtain that

$$|\mathrm{NT}(G)| \geq \frac{(1-\varepsilon)\binom{n}{2}/r \cdot \delta\binom{n}{2}/r}{2+3(n-3)} > \frac{\delta}{4r^2}\binom{n}{3} = \varepsilon\binom{n}{3},$$

which contradicts the assumption of the lemma. $\square$

4.3 Counting Gallai colorings in r-templates

In this section, we aim to prove the following technical theorem, which will be used repeatedly in the rest of the chapter.

Theorem 4.20. *Let $n, r \in \mathbb{N}$ with $r \geq 3$, and G be a graph of order n. Suppose that $\delta = \log^{-11} n$ and k is a positive constant, which does not depend on n. For two colors $i, j \in [r]$, denote by $\mathcal{F} = \mathcal{F}(i,j)$ the set of r-templates of order n, which contain at least $(1-k\delta)\binom{n}{2}$ edges with palette $\{i,j\}$. Then, for n sufficiently large,*

$$|\mathrm{Ga}(\mathcal{F}, G)| \leq 2^{e(G)} + 2^{-\frac{n}{3\log^2 n}} 2^{\binom{n}{2}}.$$

Fix two colors $1 \le i < j \le r$, and let $S = [r] - \{i, j\}$. For an r-coloring F of G, let $S(F)$ be the set of edges in G, which are colored by colors in S. From the definition of $\mathcal{F}$, we immediately obtain the following proposition.

Proposition 4.21. *For every $F \in \mathrm{Ga}(\mathcal{F}, G)$, the number of edges in $S(F)$ is at most $k\delta\binom{n}{2}$.*

Lemma 4.22. *Let $\mathcal{F}_1$ be the set of $F \in \mathrm{Ga}(\mathcal{F}, G)$ such that $S(F)$ contains a matching of size $\delta n \log^2 n$. Then, for n sufficiently large,*

$$|\mathcal{F}_1| \le 2^{-\frac{n^2}{5\log^9 n}} 2^{\binom{n}{2}}.$$

Proof. Let us consider the ways to color G so that the resulting colorings are in $\mathcal{F}_1$. We first choose the set of edges E^S which will be colored by the colors in S. Note that E^S must contain a matching of size $\delta n \log^2 n$ by the definition of $\mathcal{F}_1$. By Proposition 4.21, there are at most $\sum_{i \le k\delta\binom{n}{2}} \binom{\binom{n}{2}}{i}$ choices for such E^S, and the number of ways to color them is at most $r^{k\delta\binom{n}{2}}$. In the next step, take a matching M of size $\delta n \log^2 n$ in E^S; the number of ways to choose such matching is at most $\binom{\binom{n}{2}}{\delta n \log^2 n}$.

Let $A = V(M)$ and $B = [n] \setminus A$. Denote by $\mathcal{T}$ the set of triangles of K_n with a vertex in B and an edge from M, which contain no edge in $E^S \cap G[A, B]$. We claim that $|\mathcal{T}| \ge \frac{1}{4}\delta n^2 \log^2 n$ as otherwise we would obtain that

$$|E^S| \ge |B| \cdot \delta n \log^2 n - |\mathcal{T}| + |M| \ge \frac{1}{2}\delta n^2 \log^2 n - \frac{1}{4}\delta n^2 \log^2 n = \frac{1}{4}\delta n^2 \log^2 n > k\delta\binom{n}{2},$$

which, by Proposition 4.21, contradicts the fact that $F \in \mathrm{Ga}(\mathcal{F}, G)$. Note that if a triangle T in $\mathcal{T}$ contains more than one uncolored edge, then they must have the same color in order to avoid the rainbow triangle. Hence, the number of ways to color the uncolored edges in $\mathcal{T}$ is at most $2^{|\mathcal{T}|}$.

There remain at most $\binom{n}{2} - 2|\mathcal{T}|$ uncolored edges and they can only be colored by i or j, as edges in E_S are already colored. Hence, the number of ways to color the rest of edges is at most $2^{\binom{n}{2} - 2|\mathcal{T}|}$. In conclusion, we obtain that

$$\begin{aligned} |\mathcal{F}_1| &\le \textstyle\sum_{i \le k\delta\binom{n}{2}} \binom{\binom{n}{2}}{i} r^{k\delta\binom{n}{2}} \binom{\binom{n}{2}}{\delta n \log^2 n} \cdot 2^{|\mathcal{T}|} \cdot 2^{\binom{n}{2} - 2|\mathcal{T}|} \\ &\le 2^{O(\delta n^2 \log n)} \cdot 2^{O(\delta n \log^3 n)} \cdot 2^{\binom{n}{2} - \frac{1}{4}\delta n^2 \log^2 n} \le 2^{\binom{n}{2} - \frac{n^2}{5\log^9 n}}. \end{aligned}$$

∎

Lemma 4.23. *For every integer $1 \le t < \delta n \log^2 n$, let $\mathcal{F}(t)$ be the set of $F \in \mathrm{Ga}(\mathcal{F}, G)$, in which the maximum matching of $S(F)$ is of size t. Then, for n sufficiently large,*

$$|\mathcal{F}(t)| \le 2^{-\frac{n}{2\log^2 n}} 2^{\binom{n}{2}}.$$

Proof. For a fixed t, let us count the ways to color G so that the resulting colorings are in $\mathcal{F}(t)$. By the definition of $\mathcal{F}(t)$, among all edges which will be colored by the colors in S, there exists a maximum matching M of size t. We first choose such matching; the number of ways is at most $\binom{\binom{n}{2}}{t}$. Once we fix the matching M, let $A = V(M)$ and $B = [n] \setminus A$. By the maximality of M, we immediately obtain the following claim.

Claim 4.24. *None of the edges in $G[B]$ can be colored by the colors in S.* □

Denote by $\mathrm{Cr}(S)$ the set of edges in $G[A, B]$ which will be colored by the colors in S. For a vertex $u \in A$, denote by $\mathrm{Cr}(S, u)$ the set of edges in $\mathrm{Cr}(S)$ with one endpoint u. Similarly, define $\mathrm{Cr}(\{i, j\}, u)$ to be the set of edges in $G[u, B]$ which will be colored by the colors i or j. We shall divide the proof into three cases.

Case 1: $|\mathrm{Cr}(S)| \leq \frac{nt}{\log^2 n}$.

We first color the edges in $G[A]$ and the number of options is at most $r^{\binom{2t}{2}}$. In the next step, we select and color the edges in $\mathrm{Cr}(S)$; by the above inequality, the number of ways is at most $\sum_{i \leq \frac{nt}{\log^2 n}} \binom{2nt}{i} r^{\frac{nt}{\log^2 n}}$. By Claim 4.24, the remaining edges can only use the colors i or j. Let $\mathcal{T}$ be the set of triangles of K_n formed by a vertex in B and an edge from M, which contain no edge in $\mathrm{Cr}(S)$. We claim that $|\mathcal{T}| \geq \frac{1}{4}nt$ as otherwise we would obtain

$$|\mathrm{Cr}(S)| \geq |B|t - |\mathcal{T}| \geq \frac{1}{2}nt - \frac{1}{4}nt > \frac{nt}{\log^2 n},$$

which contradicts the assumption. If a triangle T in $\mathcal{T}$ contains more than one uncolored edge, then they must have the same color in order to avoid the rainbow triangle. Hence, the number of ways to color the uncolored edges in $\mathcal{T}$ is at most $2^{|\mathcal{T}|}$.

There remain at most $\binom{n}{2} - 2|\mathcal{T}| - \binom{2t}{2}$ uncolored edges, and they can be colored by i or j. Therefore the number of ways to color the rest of the edges is at most $2^{\binom{n}{2} - 2|\mathcal{T}| - \binom{2t}{2}}$. In conclusion, we obtain that the number of r-coloring $F \in \mathcal{F}(t)$ with $|\mathrm{Cr}(S)| \leq \frac{nt}{\log^2 n}$ is at most

$$\begin{aligned}
&\binom{\binom{n}{2}}{t} \cdot r^{\binom{2t}{2}} \cdot \textstyle\sum_{i \leq \frac{nt}{\log^2 n}} \binom{2nt}{i} r^{\frac{nt}{\log^2 n}} \cdot 2^{|\mathcal{T}|} \cdot 2^{\binom{n}{2} - 2|\mathcal{T}| - \binom{2t}{2}} \\
\leq &2^{O(t \log n)} \cdot 2^{O(t^2)} \cdot 2^{O\left(\frac{nt}{\log n}\right)} \cdot 2^{\binom{n}{2} - \frac{1}{4}nt} \leq 2^{\binom{n}{2} - \frac{1}{5}nt} \leq 2^{\binom{n}{2} - \frac{1}{5}n},
\end{aligned}$$

where the third inequality is given by $t^2 \leq t \cdot \delta n \log^2 n = nt / \log^9 n$.

Case 2: There exists a vertex $u \in A$ such that

$$|\mathrm{Cr}(S, u)| \geq \frac{n}{\log^4 n} \qquad \text{and} \qquad |\mathrm{Cr}(\{i, j\}, u)| \geq \frac{n}{\log^4 n}. \tag{4.3}$$

We first choose the vertex u, and the number of options is at most $2t$. Moreover, the number of ways to select and color edges in $\mathrm{Cr}(S,u)$ is at most $r^n 2^n$. In the next step, we color all the uncolored edges in $G[A,B]$ and $G[A]$, and the number of ways is at most $r^{2nt+\binom{2t}{2}}$. Let $\mathcal{T}$ be the set of triangles $T=\{uvw\}$ of K_n, in which $v,w\in B$, $uv\in \mathrm{Cr}(S,u)$, and $uw\in \mathrm{Cr}(\{i,j\},u)$. By the relation (4.3), we have $|\mathcal{T}|\geq \frac{n^2}{\log^8 n}$. For every triangle $T=\{uvw\}\in\mathcal{T}$, if vw is an edge of G, then by Claim 4.24 it can only be colored by i or j, and must have the same color with uw in order to avoid the rainbow triangle. Therefore, the number of ways to color the uncolored edges in $\mathcal{T}$ is 1.

There remain at most $\binom{n}{2}-|\mathcal{T}|$ uncolored edges in B, as other edges are already colored. By Claim 4.24, none of the remaining edges in B could use the colors from S. Therefore, the number of ways to color the rest of edges is at most $2^{\binom{n}{2}-|\mathcal{T}|}$. In conclusion, we obtain that the number of $F\in\mathcal{F}(t)$ which is included in Case 2 is at most

$$\binom{\binom{n}{2}}{t}\cdot 2t\cdot r^n 2^n\cdot r^{2nt+\binom{2t}{2}}\cdot 2^{\binom{n}{2}-|\mathcal{T}|}\leq 2^{O(t\log n)}\cdot 2^{O(n)}\cdot 2^{O(nt)}\cdot 2^{\binom{n}{2}-\frac{n^2}{\log^8 n}}\leq 2^{\binom{n}{2}-\frac{n^2}{2\log^8 n}},$$

where the last inequality is given by the condition that $nt\leq n\cdot\delta n\log^2 n=n^2/\log^9 n$.

Case 3: $|\mathrm{Cr}(S)|>\frac{nt}{\log^2 n}$, and for every vertex $u\in A$,

$$|\mathrm{Cr}(S,u)|<\frac{n}{\log^4 n}\qquad\text{or}\qquad |\mathrm{Cr}(\{i,j\},u)|<\frac{n}{\log^4 n}. \tag{4.4}$$

We first color the edges in $G[A]$ and the number of ways is at most $r^{\binom{2t}{2}}$. By (4.4), for every vertex $u\in A$, the number of ways to select $\mathrm{Cr}(S,u)$ is at most $2\sum_{i\leq n/\log^4 n}\binom{n}{i}\leq 2^{n/\log^3 n}$. Therefore, the number of ways to select $\mathrm{Cr}(S)$ is at most $2^{2nt/\log^3 n}$.

Subcase 3.1: $e(G)\leq\binom{n}{2}-\frac{n^2}{4\log^6 n}$.

The number of ways to color $\mathrm{Cr}(S)$ is at most r^{2nt}. By Claim 4.24, the rest of the edges can only be colored by i or j, and the number of them is at most $e(G)-|\mathrm{Cr}(S)|$. Hence, the number of $F\in\mathcal{F}(t)$ covered in Case 3.1 is at most

$$\binom{\binom{n}{2}}{t}\cdot r^{\binom{2t}{2}}\cdot 2^{\frac{2nt}{\log^3 n}}\cdot r^{2nt}\cdot 2^{e(G)-|\mathrm{Cr}(S)|}\leq 2^{O(t\log n)}\cdot 2^{O(nt)}\cdot 2^{\binom{n}{2}-\frac{n^2}{4\log^6 n}-\frac{nt}{\log^2 n}}\leq 2^{\binom{n}{2}-\frac{n^2}{5\log^6 n}},$$

where the last inequality holds by the condition that $nt\leq n\cdot\delta n\log^2 n=n^2/\log^9 n$.

Subcase 3.2: $e(G) > \binom{n}{2} - \frac{n^2}{4\log^6 n}$.

For $u \in A$, define $N_S(u) = \{v \in B \mid uv \in \mathrm{Cr}(S,u)\}$. Let G_u be the induced subgraph of G on $N_S(u)$, and denote by $c(G_u)$ the number of components of G_u.

Claim 4.25. *For every $u \in A$, we have $c(G_u) \leq \frac{n}{\log^3 n}$.*

Proof. Suppose that there exists a vertex u in A with $c(G_u) > \frac{n}{\log^3 n}$. Then the number of non-edges in G_u is at least $\binom{\frac{n}{\log^3 n}}{2} \geq \frac{n^2}{4\log^6 n}$, which contradicts with the assumption of Case 3.2. $\square$

Claim 4.26. *For every $u \in A$, the number of ways to color $\mathrm{Cr}(S,u)$ is at most $r^{c(G_u)}$.*

Proof. Let C be an arbitrary component of G_u. It is sufficient to prove that for every $v, w \in V(C)$, uv and uw must have the same color. Assume that there exist $v, w \in V(C)$ such that uv and uw receive different colors. Since C is a connected component of G_u, there is a path $P = \{v = v_0, v_1, v_2, \ldots, v_k = w\}$ in G_u, in which uv_i is painted by a color in S for every $0 \leq i \leq k$. Moreover, since uv and uw receive different colors, there exists an integer $0 \leq j \leq k-1$ such that uv_j and uv_{j+1} receive different colors. On the other hand, by Claim 4.24, $v_j v_{j+1}$ can only be colored by i or j. Therefore, u, v_j, v_{j+1} form a rainbow triangle, which is not allowed in a Gallai r-coloring. $\square$

By Claims 4.25 and 4.26, the number of ways to color $\mathrm{Cr}(S,u)$ is at most $r^{\frac{n}{\log^3 n}}$, and therefore the total number of ways to color $\mathrm{Cr}(S)$ is at most $r^{\frac{2nt}{\log^3 n}}$. By Claim 4.24, the rest of the edges can only be colored by i or j, and the number of them is at most $e(G) - |\mathrm{Cr}(S)|$. Hence, the number of $F \in \mathcal{F}(t)$ included in Case 3.2 is at most

$$\binom{\binom{n}{2}}{t} \cdot r^{\binom{2t}{2}} \cdot 2^{\frac{2nt}{\log^3 n}} \cdot r^{\frac{2nt}{\log^3 n}} \cdot 2^{e(G)-|\mathrm{Cr}(S)|} \leq 2^{O(t\log n)} \cdot 2^{O\left(\frac{nt}{\log^3 n}\right)} \cdot 2^{\binom{n}{2} - \frac{nt}{\log^2 n}} \leq 2^{\binom{n}{2} - \frac{n}{2\log^2 n} - 1}.$$

Eventually, we conclude that

$$|\mathcal{F}(t)| \leq 2^{\binom{n}{2} - \frac{1}{5}n} + 2^{\binom{n}{2} - \frac{n^2}{2\log^8 n}} + 2^{\binom{n}{2} - \frac{n}{2\log^2 n} - 1} \leq 2^{-\frac{n}{2\log^2 n}} 2^{\binom{n}{2}}$$

for every $1 \leq t < \delta n \log^2 n$. $\square$

Observe that every r-coloring of G using at most 2 colors is a Gallai r-coloring. Then we immediately obtain the following lemma.

Lemma 4.27. *Let $\mathcal{F}_0$ be the set of $F \in \mathrm{Ga}(\mathcal{F}, G)$ such that $S(F) = \emptyset$. Then $|\mathcal{F}_0| = 2^{e(G)}$.*

Now, we have all the ingredients to prove Theorem 4.20.

Proof of Theorem 4.20. Applying Lemmas 4.22, 4.23 and 4.27, we obtain that

$$|\mathrm{Ga}(\mathcal{F}, G)| = |\mathcal{F}_1| + \sum_{t=1}^{\delta n/\log^2 n} |\mathcal{F}(t)| + |\mathcal{F}_0| \leq 2^{e(G)} + 2^{-\frac{n}{3\log^2 n}} 2^{\binom{n}{2}},$$

for n sufficiently large. □

4.4 Gallai r-colorings of complete graphs

4.4.1 Stability of the Gallai r-template of complete graphs

Proposition 4.28. *Let n, $r \in \mathbb{N}$ with $r \geq 3$. Suppose P is a Gallai r-template of K_n. Then the number of edges with at least 3 colors in its palette is at most $n^{-1/6}n^2$.*

Proof. Let $E = \{e \in E(K_n) : \; |P(e)| \geq 3\}$ and assume that $|E| > n^{-1/6}n^2$. Let F be a spanning subgraph of K_n with edge set E. For every $i \in [n]$, denote by d_i the degree of vertex i of F. Then the number of 3-paths in F is equal to

$$\sum_{i\in[n]} \binom{d_i}{2} \geq n\binom{\frac{\sum_{i\in[n]} d_i}{n}}{2} \geq n\binom{2|E|/n}{2} \geq \frac{|E|^2}{n} > 3n^{-1/3}\binom{n}{3}.$$

Observe that if i, j, k is a 3-path in F, then there is at least one rainbow triangle in P with vertex set $\{i, j, k\}$ since edges ij, jk have at least 3 colors in its palette and edge ik has at least one color in its palette. Therefore, there would be more than $n^{-1/3}\binom{n}{3}$ rainbow triangles in P, which contradicts the fact that P is a Gallai r-template. □

Lemma 4.29. *Let n, $r \in \mathbb{N}$ with $r \geq 3$ and $n^{-1/6} \ll \delta \ll 1$. Assume that P is a Gallai r-template of K_n with $|\mathrm{Ga}(P, K_n)| > 2^{(1-\delta)\binom{n}{2}}$. Then the number of triangles T of K_n with $\sum_{e\in T} |P(e)| = 6$ and $P(e) = P(e')$ for every e, $e' \in T$ is at least $(1 - 4\delta)\binom{n}{3}$.*

Proof. Let $\mathcal{T}$ be the collection of triangles of K_n. We define

$$\mathcal{T}_1 = \left\{T \in \mathcal{T} \mid \textstyle\sum_{e\in T} |P(e)| = 6 \text{ and } P(e) = P(e') \text{ for every } e, \; e' \in T\right\},$$

$$\mathcal{T}_2 = \{T \in \mathcal{T} \mid \exists \; e \in T, \; |P(e)| \geq 3\},$$

$$\mathcal{T}_3 = \left\{T \in \mathcal{T} \setminus (\mathcal{T}_1 \cup \mathcal{T}_2) \mid \textstyle\sum_{e\in T} |P(e)| = 6\right\},$$

$$\mathcal{T}_4 = \left\{T \in \mathcal{T} \setminus \mathcal{T}_2 \mid \textstyle\sum_{e\in T} |P(e)| \leq 5\right\}.$$

Let $|\mathcal{T}_1| = \alpha\binom{n}{3}$, $|\mathcal{T}_2| = \beta\binom{n}{3}$, $|\mathcal{T}_3| = \gamma\binom{n}{3}$. Then $|\mathcal{T}_4| \leq (1-\alpha)\binom{n}{3}$. By Proposition 4.28, we have $|\mathcal{T}_2| \leq n^{-1/6}n^3$ and therefore $\beta \leq 12n^{-1/6}$. Observe that for every $T \in \mathcal{T}_3$, the template P contains a rainbow triangle with edge set T; therefore, we obtain that $|\mathcal{T}_3| \leq \mathrm{RT}(P) \leq n^{-1/3}\binom{n}{3}$, which gives $\gamma \leq n^{-1/3} \leq n^{-1/6}$.

Assume that $\alpha < 1 - 4\delta$. Then the number of Gallai r-colorings of K_n, which are subtemplates of P, satisfies

$$\begin{aligned}
\log|\mathrm{Ga}(P, K_n)| &\leq \log\left(\prod_{e \in E(K_n)} |P(e)|\right) = \log\left(\prod_{T \in \mathcal{T}} \prod_{e \in T} |P(e)|\right)^{\frac{1}{n-2}} \\
&\leq \log\left(\prod_{T \in \mathcal{T}_1} 2^3 \prod_{T \in \mathcal{T}_2} r^3 \prod_{T \in \mathcal{T}_3} 2^3 \prod_{T \in \mathcal{T}_4} 2^2\right) \cdot \frac{1}{n-2} \\
&\leq (3\alpha + 3\beta \log r + 3\gamma + 2(1-\alpha))\, \frac{1}{3}\binom{n}{2} \\
&\leq \left(2 + \alpha + (36\log r + 3)n^{-1/6}\right) \frac{1}{3}\binom{n}{2} < (2 + (1-4\delta) + \delta)\, \frac{1}{3}\binom{n}{2} = (1-\delta)\binom{n}{2}.
\end{aligned}$$

This contradicts the assumption that $|\mathrm{Ga}(P, K_n)| > 2^{(1-\delta)\binom{n}{2}}$. □

We now prove a stability result for Gallai r-templates of K_n.

Theorem 4.30. *Let n, $r \in \mathbb{N}$ with $r \geq 3$ and $n^{-1/6} \ll \delta \ll 1$. Assume that P is a Gallai r-template of K_n with $|\mathrm{Ga}(P, K_n)| > 2^{(1-\delta)\binom{n}{2}}$. Then there exist two colors i, $j \in [r]$ such that the number of edges of K_n with palette $\{i, j\}$ is at least $(1 - 4r^4\delta)\binom{n}{2}$.*

Proof. Let G be an $\binom{r}{2}$-colored graph with edge set $E(G) = \{e \in E(K_n) \mid |P(e)| = 2\}$ and color set $\{(i, j) \mid 1 \leq i < j \leq r\}$, where each edge e is colored by color $P(e)$. By Lemma 4.29, the number of monochromatic triangles in G is at least $(1-4\delta)\binom{n}{3}$. Applying Lemma 4.19 on G, we obtain that there exist two colors i, j such that the number of edges with palette $\{i, j\}$ is at least $e(G) - 4\binom{r}{2}^2 \cdot 4\delta\binom{n}{2} \geq (1-4\delta)\binom{n}{2} - 4\binom{r}{2}^2 \cdot 4\delta\binom{n}{2} \geq (1 - 4r^4\delta)\binom{n}{2}$. □

4.4.2 Proof of Theorem 4.1

Proof of Theorem 4.1. Let $\mathcal{C}$ be the collection of containers given by Theorem 4.16. We observe that a Gallai r-coloring of K_n can be regarded as a rainbow triangle-free r-coloring template of order n, with only one color allowed at each edge. Therefore, by Property (i) of Theorem 4.16, every Gallai r-coloring of K_n is a subtemplate of some $P \in \mathcal{C}$.

Let $\delta = \log^{-11} n$. We define

$$\mathcal{C}_1 = \left\{P \in \mathcal{C} :\ |\mathrm{Ga}(P, K_n)| \leq 2^{(1-\delta)\binom{n}{2}}\right\}, \quad \mathcal{C}_2 = \left\{P \in \mathcal{C} :\ |\mathrm{Ga}(P, K_n)| > 2^{(1-\delta)\binom{n}{2}}\right\}.$$

By Property (iii) of Theorem 4.16, we have

$$|\mathrm{Ga}(\mathcal{C}_1, K_n)| \le |\mathcal{C}_1| \cdot 2^{(1-\delta)\binom{n}{2}} \le 2^{cn^{-1/3}\log^2 n\binom{n}{2}} \cdot 2^{\binom{n}{2}-\log^{-11} n\binom{n}{2}} \le 2^{-\frac{n^2}{4\log^{11} n}} 2^{\binom{n}{2}}.$$

We claim that every template P in $\mathcal{C}_2$ is a Gallai r-template of K_n. First, by Property (ii) of Theorem 4.16, we have $\mathrm{RT}(P) \le n^{-1/3}\binom{n}{3}$. Suppose that there exists an edge $e \in E(K_n)$ with $|P(e)| = 0$. Then we would obtain $\mathrm{Ga}(P, K_n) = \emptyset$ as a Gallai r-coloring of K_n requires at least one color on each edge, which contradicts the definition of $\mathcal{C}_2$. Now by Theorem 4.30, we can divide $\mathcal{C}_2$ into classes $\{\mathcal{F}_{i,j}, 1 \le i < j \le r\}$, where $\mathcal{F}_{i,j}$ consists of all the r-templates in $\mathcal{C}_2$ which contain at least $(1-4r^4\delta)\binom{n}{2}$ edges with palette $\{i, j\}$. Applying Theorem 4.20 on $\mathcal{F}_{i,j}$, we obtain that $|\mathrm{Ga}(\mathcal{F}_{i,j}, K_n)| \le \left(1 + 2^{-\frac{n}{3\log^2 n}}\right) 2^{\binom{n}{2}}$, and therefore

$$|\mathrm{Ga}(\mathcal{C}_2, K_n)| \le \sum_{1\le i<j\le r} |\mathrm{Ga}(\mathcal{F}_{i,j}, K_n)| \le \binom{r}{2}\left(1 + 2^{-\frac{n}{3\log^2 n}}\right) 2^{\binom{n}{2}}.$$

Finally, we conclude that

$$|\mathrm{Ga}(\mathcal{C}, K_n)| = |\mathrm{Ga}(\mathcal{C}_1, K_n)| + |\mathrm{Ga}(\mathcal{C}_2, K_n)| \le \left(\binom{r}{2} + 2^{-\frac{n}{4\log^2 n}}\right) 2^{\binom{n}{2}},$$

which gives the desired upper bound for the number of Gallai r-colorings of K_n. □

4.5 Gallai 3-colorings of non-complete graphs

In this section, we count Gallai 3-colorings of dense non-complete graphs. We shall explore the stability property first, and then follow a somewhat similar strategy as in the proof of Theorem 4.1. The main obstacle is that in a Gallai r-template of a non-complete graph, a palette of an edge could be an empty set, which leads to a more sophisticated discussion of templates.

4.5.1 Triangles in r-templates of dense graphs

Let $\mathcal{T}$ be the collection of triangles of K_n. For a given r-template P of order n, we partition the triangles into 5 classes. We set an extra class, as a $T \in \mathcal{T}$ may not be a triangle in G.

$$\begin{aligned}
&\mathcal{T}_1(P) = \left\{T \in \mathcal{T} \mid \textstyle\sum_{e\in T} |P(e)| = 6 \text{ and } P(e) = P(e') \text{ for every } e,\ e' \in T\right\},\\
&\mathcal{T}_2(P) = \{T \in \mathcal{T} \mid T = \{e_1, e_2, e_3\},\ |P(e_1)| \geq 3,\ |P(e_2)| \geq 3,\ \text{and } |P(e_3)| = 0\},\\
&\mathcal{T}_3(P) = \{T \in \mathcal{T} \mid T = \{e_1, e_2, e_3\},\ |P(e_1)| \geq 3,\ |P(e_2)| + |P(e_3)| \leq 2\},\\
&\mathcal{T}_4(P) = \left\{T \in \mathcal{T} \setminus (\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3) \mid \textstyle\sum_{e\in T} |P(e)| \geq 6\right\},\\
&\mathcal{T}_5(P) = \left\{T \in \mathcal{T} \setminus \mathcal{T}_3 \mid \textstyle\sum_{e\in T} |P(e)| \leq 5\right\}.
\end{aligned} \tag{4.5}$$

Lemma 4.31. *Let $n, r \in \mathbb{N}$ with $r \geq 4$ and $0 < k \leq 1$. For $0 < \xi \leq \left(\frac{k}{2+6k}\right)^2$, let G be a graph of order n, and $e(G) \geq (1-\xi)\binom{n}{2}$. Assume that P is a Gallai r-template of G. Then, for sufficiently large n,*

$$|\mathcal{T}_2(P)| \leq \max\left\{k|\mathcal{T}_3(P)|,\ \frac{3+9k}{k} n^{-\frac{1}{3}} \binom{n}{3}\right\}.$$

Proof. Let $E = \{e \in E(K_n) :\ |P(e)| \geq 3\}$ and F be a spanning subgraph of K_n with edge set E. For every $i \in [n]$, denote by d_i the degree of vertex i of F. Since $\sum_{i=1}^n d_i = 2|E|$, the number of vertices with $d_i > \sqrt{\xi} n$ is less than $\frac{2|E|}{\sqrt{\xi} n}$. Therefore, we obtain

$$\begin{aligned}
|\mathcal{T}_2(P)| &\leq \sum_{i=1}^n \min\left\{\binom{d_i}{2}, \xi\binom{n}{2}\right\} < \frac{2|E|}{\sqrt{\xi} n} \cdot \frac{\xi n^2}{2} + \sum_{d_i \leq \sqrt{\xi} n} \frac{d_i^2}{2}\\
&\leq \frac{2|E|}{\sqrt{\xi} n} \cdot \frac{\xi n^2}{2} + \frac{2|E|}{\sqrt{\xi} n} \cdot \frac{\xi n^2}{2} = 2|E|\sqrt{\xi} n \leq \frac{k}{1+3k} n|E|,
\end{aligned} \tag{4.6}$$

where the third inequality follows from the concavity of the function x^2. The rest of the proof is divided into two cases.

Case 1: $|E| \geq \frac{2+6k}{k} n^{-\frac{1}{3}} \binom{n}{2}$.

Consider all triangles of K_n with at least one edge in E. Note that if a triangle has at least one edge in E and belongs to neither $\mathcal{T}_3(P)$ nor $\mathcal{T}_2(P)$, then it induces a rainbow triangle in P. Together with (4.6), we have

$$\begin{aligned}
k|\mathcal{T}_3(P)| &\geq k\left(|E|(n-2) - 2|\mathcal{T}_2(P)| - 3n^{-\frac{1}{3}}\tbinom{n}{3}\right) \geq k\left(\tfrac{1+k}{1+3k} n|E| - 2|E| - 3n^{-\frac{1}{3}}\tbinom{n}{3}\right)\\
&= \tfrac{k}{1+3k} n|E| + k\left(\tfrac{k}{1+3k} n|E| - 2|E| - 3n^{-\frac{1}{3}}\tbinom{n}{3}\right) \geq \tfrac{k}{1+3k} n|E| \geq |\mathcal{T}_2(P)|,
\end{aligned}$$

where the fourth inequality is given by $|E| \ge \frac{2+6k}{k} n^{-\frac{1}{3}} \binom{n}{2}$ for sufficiently large n.

Case 2: $|E| < \frac{2+6k}{k} n^{-\frac{1}{3}} \binom{n}{2}$.

In this case, we have

$$|\mathcal{T}_2(P)| < \frac{1}{2}|E|(n-2) < \frac{3+9k}{k} n^{-1/3} \binom{n}{3}.$$

□

4.5.2 Stability of Gallai 3-templates of dense non-complete graphs

Lemma 4.32. *Let $0 < \xi \le \frac{1}{64}$ and $n^{-1/3} \ll \delta \ll 1$. Let G be a graph of order n, and $e(G) \ge (1-\xi)\binom{n}{2}$. Assume that P is a Gallai 3-template of G with $|\mathrm{Ga}(P,G)| > 2^{(1-\delta)\binom{n}{2}}$. Then $|\mathcal{T}_1(P)| \ge (1-40\delta)\binom{n}{3}$.*

Proof. Let $|\mathcal{T}_1(P)| = \alpha\binom{n}{3}$, $|\mathcal{T}_2(P)| = \beta\binom{n}{3}$, $|\mathcal{T}_3(P)| = \eta\binom{n}{3}$ and $|\mathcal{T}_4(P)| = \gamma\binom{n}{3}$. Then $|\mathcal{T}_5(P)| \le (1-\alpha-\beta-\eta)\binom{n}{3}$. Observe that for every $T \in \mathcal{T}_4(P)$, the template P contains a rainbow triangle with edge set T; therefore, we obtain that $|\mathcal{T}_4(P)| \le RT(P) \le n^{-1/3}\binom{n}{3}$, which gives $\gamma \le n^{-1/3}$.

Define for $e \in E(K_n)$ the weight function

$$w(e) = \begin{cases} 1 & \text{if } P(e) = \emptyset, \\ |P(e)| & \text{otherwise.} \end{cases}$$

Similarly to the proof of Lemma 4.29, the number of Gallai 3-colorings of G which are subtemplates of P satisfies

$$\begin{aligned} \log|\mathrm{Ga}(P,G)| &\le \log\left(\prod_{e\in K_n} |w(e)|\right) = \log\left(\prod_{T\in\mathcal{T}}\prod_{e\in T} |w(e)|\right)^{\frac{1}{n-2}} \\ &\le \log\left(\prod_{T\in\mathcal{T}_1} 2^3 \prod_{T\in\mathcal{T}_2} 3^2 \prod_{T\in\mathcal{T}_3} 6 \prod_{T\in\mathcal{T}_4} 3^3 \prod_{T\in\mathcal{T}_5} 2^2\right)\frac{1}{n-2} \\ &\le (3\alpha + 2\beta\log 3 + \eta\log 6 + 3\gamma\log 3 + 2(1-\alpha-\beta-\eta))\,\frac{1}{3}\binom{n}{2} \\ &= \left(2 + \alpha + (2\log 3 - 2)\beta + (\log 6 - 2)\eta + 3n^{-1/3}\log 3\right)\frac{1}{3}\binom{n}{2}. \end{aligned} \tag{4.7}$$

Let $k = 1$. By Lemma 4.31, we have $\beta \le \max\{\eta,\ 12n^{-1/3}\}$. Assume that $\alpha < 1 - 40\delta$. The rest of the proof shall be divided into two cases.

Case 1: $\beta \le \eta$.

If $\eta < 20\delta$, continuing (4.7) we have

$$\begin{aligned}\log|\mathrm{Ga}(P,G)| &\le \left(2+\alpha+(2\log 3+\log 6-4)\,\eta+3n^{-1/3}\log 3\right)\tfrac{1}{3}\tbinom{n}{2}\\ &\le (2+(1-40\delta)+1.8\cdot 20\delta+\delta)\,\tfrac{1}{3}\tbinom{n}{2}=(1-\delta)\tbinom{n}{2}.\end{aligned}$$

Otherwise, together with $\alpha \le 1-\beta-\eta$, continuing (4.7) we obtain that

$$\begin{aligned}\log|\mathrm{Ga}(P,G)| &\le \left(3+(2\log 3-3)\beta+(\log 6-3)\eta+3n^{-1/3}\log 3\right)\tfrac{1}{3}\tbinom{n}{2}\\ &\le \left(3+(2\log 3+\log 6-6)\,\eta+3n^{-1/3}\log 3\right)\tfrac{1}{3}\tbinom{n}{2}\\ &\le (3-0.2\cdot 20\delta+\delta)\,\tfrac{1}{3}\tbinom{n}{2}=(1-\delta)\tbinom{n}{2}.\end{aligned}$$

Case 2: $\beta \le 12n^{-1/3}$.

Together with $\eta \le 1-\alpha$ and $\alpha < 1-40\delta$, continuing (4.7) we have

$$\begin{aligned}\log|\mathrm{Ga}(P,G)| &\le \left(2+\alpha+2\log 3\cdot 12n^{-1/3}+(\log 6-2)(1-\alpha)+3n^{-1/3}\log 3\right)\tfrac{1}{3}\tbinom{n}{2}\\ &\le \left(\log 6+(3-\log 6)\alpha+27n^{-1/3}\log 3\right)\tfrac{1}{3}\tbinom{n}{2}\\ &\le (\log 6+(3-\log 6)(1-40\delta)+\delta)\,\tfrac{1}{3}\tbinom{n}{2}<(1-\delta)\tbinom{n}{2}.\end{aligned}$$

Both cases contradict our assumption that $|\mathrm{Ga}(P,G)| > 2^{(1-\delta)\binom{n}{2}}$. □

Similarly as in the proof of Theorem 4.30, using Lemmas 4.19 and 4.32, we obtain the following theorem.

Theorem 4.33. *Let $0<\xi\le\frac{1}{64}$ and $n^{-1/3}\ll\delta\ll 1$. Let G be a graph of order n and $e(G)\ge(1-\xi)\binom{n}{2}$. Assume that P is a Gallai 3-template of G with $|\mathrm{Ga}(P,G)|>2^{(1-\delta)\binom{n}{2}}$. Then there exist two colors $i,\,j\in[3]$ such that the number of edges of K_n with palette $\{i,j\}$ is at least $(1-37\cdot 40\delta)\binom{n}{2}$.*

4.5.3 Proof of Theorem 4.9

Proof of Theorem 4.9. Let $\mathcal{C}$ be the collection of containers given by Theorem 4.16 for $r=3$. Note that every Gallai 3-coloring of G is a subtemplate of some $P\in\mathcal{C}$. Let $\delta=\log^{-11} n$. We define

$$\mathcal{C}_1=\left\{P\in\mathcal{C}:\ |\mathrm{Ga}(P,K_n)|\le 2^{(1-\delta)\binom{n}{2}}\right\},\quad \mathcal{C}_2=\left\{P\in\mathcal{C}:\ |\mathrm{Ga}(P,K_n)|> 2^{(1-\delta)\binom{n}{2}}\right\}.$$

Similarly to the proof of Theorem 4.1, applying Theorems 4.16, 4.20, and 4.33, we obtain that

$$
\begin{aligned}
|\mathrm{Ga}(\mathcal{C}, G)| = |\mathrm{Ga}(\mathcal{C}_1, G)| + |\mathrm{Ga}(\mathcal{C}_2, G)| &\leq 2^{-\frac{n^2}{4\log^{11} n}} 2^{\binom{n}{2}} + 3 \cdot \left(2^{e(G)} + 2^{-\frac{n}{3\log^2 n}} 2^{\binom{n}{2}}\right) \\
&\leq 3 \cdot 2^{e(G)} + 2^{-\frac{n}{4\log^2 n}} 2^{\binom{n}{2}}.
\end{aligned}
$$

□

4.6 Gallai r-colorings of non-complete graphs

Theorem 4.11 is a direct consequence of the following three theorems.

Theorem 4.34. *For $n, r \in \mathbb{N}$ with $r \geq 4$, there exists n_0 such that for all $n > n_0$ the following holds. For a graph G of order n with $e(G) \geq (1 - \log^{-11} n)\binom{n}{2}$, the number of Gallai r-colorings of G is strictly less than $r^{\lfloor n^2/4 \rfloor}$.*

Theorem 4.35. *Let $n, r \in \mathbb{N}$ with $r \geq 4$, and $0 < \xi \ll 1$. For a graph G of order n with $\lfloor n^2/4 \rfloor < e(G) \leq \lfloor n^2/4 \rfloor + \xi n^2$, the number of Gallai r-colorings of G is strictly less than $r^{\lfloor n^2/4 \rfloor}$.*

Theorem 4.36. *For $n, r \in \mathbb{N}$ with $r \geq 4$, there exists n_0 such that for all $n > n_0$ the following holds. Let $n^{-1/36} \ll \xi \leq \frac{1}{2}\log^{-11} n \ll 1$. For a graph G of order n with $(\frac{1}{4} + 3\xi)n^2 \leq e(G) \leq (\frac{1}{2} - 3\xi)n^2$, the number of Gallai r-colorings of G is strictly less than $r^{\lfloor n^2/4 \rfloor}$.*

4.6.1 Proof of Theorem 4.34 for $r \geq 5$

Lemma 4.37. *Let $n, r \in \mathbb{N}$ with $r \geq 5$ and $0 < \xi \leq \frac{1}{900}$. Assume that G is a graph of order n with $e(G) \geq (1 - \xi)\binom{n}{2}$, and P is a Gallai r-template of G. Then, for sufficiently large n,*

$$
|\mathrm{Ga}(P, G)| \leq r^{\frac{1}{2}\binom{n}{2}} \cdot 2^{-0.007\binom{n}{2}}.
$$

Proof. Let $\mathcal{T}$ be the collection of triangles of K_n. For a given r-template P of order n, we again use the partition (4.5). Let $|\mathcal{T}_1(P)| = \alpha\binom{n}{3}$, $|\mathcal{T}_2(P)| = \beta\binom{n}{3}$, $|\mathcal{T}_3(P)| = \eta\binom{n}{3}$ and $|\mathcal{T}_4(P)| = \gamma\binom{n}{3}$. Then $|\mathcal{T}_5(P)| \leq (1 - \alpha - \beta - \eta)\binom{n}{3}$. Note that for every $T \in \mathcal{T}_4(P)$, the template P contains a rainbow triangle with edge set T; therefore, we obtain that $|\mathcal{T}_4(P)| \leq \mathrm{RT}(P) \leq n^{-1/3}\binom{n}{3}$, which gives $\gamma \leq n^{-1/3}$.

Define for $e \in E(K_n)$ the weight function

$$w(e) = \begin{cases} 1 & \text{if } P(e) = \emptyset \\ |P(e)| & \text{otherwise.} \end{cases}$$

Similarly, as in Lemma 4.32, the number of Gallai r-colorings of G, which is a subtemplate of P, satisfies

$$\begin{aligned} \log|\mathrm{Ga}(P,G)| &\le \log\left(\prod_{T\in\mathcal{T}_1} 2^3 \prod_{T\in\mathcal{T}_2} r^2 \prod_{T\in\mathcal{T}_3} 2r \prod_{T\in\mathcal{T}_4} r^3 \prod_{T\in\mathcal{T}_5} 2^2\right) \cdot \tfrac{1}{n-2} \\ &\le (3\alpha + 2\beta\log r + \eta\log 2r + 3\gamma\log r + 2(1-\alpha-\beta-\eta))\, \tfrac{1}{3}\tbinom{n}{2} \\ &\le \left(2 + \alpha + (2\log r - 2)\beta + (\log r - 1)\eta + 3n^{-1/3}\log r\right) \tfrac{1}{3}\tbinom{n}{2}. \end{aligned} \tag{4.8}$$

Let $k = 1/12$. By Lemma 4.31, we have $\beta \le \max\{k\eta,\ \frac{3+9k}{k} n^{-1/3}\}$. The rest of the proof shall be divided into two cases.

Case 1: $\beta \le k\eta$.

Together with $\alpha \le (1-\beta-\eta)$, continuing (4.8) we have

$$\begin{aligned} \log|\mathrm{Ga}(P,G)| &\le \left(3 + (2\log r - 3)\beta + (\log r - 2)\eta + 3n^{-1/3}\log r\right) \tfrac{1}{3}\tbinom{n}{2} \\ &\le \left(3 + ((2k+1)\log r - (3k+2))\,\eta + 3n^{-1/3}\log r\right) \tfrac{1}{3}\tbinom{n}{2}. \end{aligned}$$

Note that $(2k+1)\log r - (3k+2)$ is positive as $r \ge 4$. Therefore, together with $\eta \le 1$ and $k = \frac{1}{12}$, we obtain that

$$\begin{aligned} \log|\mathrm{Ga}(P,G)| &\le \left(\tfrac{7}{6}\log r + \tfrac{3}{4} + 3n^{-1/3}\log r\right)\tfrac{1}{3}\tbinom{n}{2} \le \left(\tfrac{3}{2}\log r - 0.023 + 3n^{-1/3}\log r\right)\tfrac{1}{3}\tbinom{n}{2} \\ &\le \tfrac{1}{2}\tbinom{n}{2}\log r - 0.007\tbinom{n}{2}, \end{aligned}$$

where the second inequality follows from $(\frac{1}{3}\log r - \frac{3}{4}) \ge 0.023$ as $r \ge 5$.

Case 2: $\beta \le \frac{3+9k}{k} n^{-1/3}$.

Together with $\alpha \le (1-\eta)$, continuing (4.8) we have

$$\begin{aligned}\log|\mathrm{Ga}(P,G)| &\le \left(3 + (\log r - 2)\eta + 2\log r \cdot \tfrac{3+9k}{k} n^{-1/3} + 3n^{-1/3}\log r\right) \tfrac{1}{3}\tbinom{n}{2} \\ &\le \left(\tfrac{3}{2}\log r - \left(\tfrac{1}{2}\log r - 1\right) + \left(\tfrac{2+6k}{k} + 1\right) 3n^{-1/3}\log r\right) \tfrac{1}{3}\tbinom{n}{2} \\ &\le \left(\tfrac{3}{2}\log r - 0.16 + 0.01\right) \tfrac{1}{3}\tbinom{n}{2} = \tfrac{1}{2}\tbinom{n}{2}\log r - 0.05\tbinom{n}{2},\end{aligned}$$

where the third inequality holds for $r \ge 5$ and sufficiently large n. □

Using Lemma 4.37, we prove a stronger theorem for the case $r \ge 5$.

Theorem 4.38. *For $n, r \in \mathbb{N}$ with $r \ge 5$ and $0 < \xi \le \frac{1}{900}$, there exists n_0 such that for all $n > n_0$ the following holds. If G is a graph of order n, and $e(G) \ge (1-\xi)\binom{n}{2}$, then the number of Gallai r-colorings of G is less than $r^{\frac{1}{2}\binom{n}{2}}$.*

Proof. Let $\mathcal{C}$ be the collection of containers given by Theorem 4.16. Theorem 4.16 indicates that every Gallai r-coloring of G is a subtemplate of some $P \in \mathcal{C}$ and $|\mathcal{C}| \le 2^{cn^{-1/3}\log^2 n\binom{n}{2}}$ for some constant c, which only depends on r. We may assume that all templates P in $\mathcal{C}$ are Gallai r-templates of G. By Property (ii) of Theorem 4.16, we always have $\mathrm{RT}(P) \le n^{-1/3}\binom{n}{3}$. Suppose that for a template P there exists an edge $e \in E(G)$ with $|P(e)| = 0$. Then we would obtain $|\mathrm{Ga}(P,G)| = 0$ as a Gallai r-coloring of G requires at least one color on each edge. Now applying Lemma 4.37 on every container $P \in \mathcal{C}$, we obtain that the number of Gallai r-colorings of G is at most

$$\sum_{P\in\mathcal{C}} |\mathrm{Ga}(P,G)| \le |\mathcal{C}| \cdot r^{\frac{1}{2}\binom{n}{2}} \cdot 2^{-0.007\binom{n}{2}} < r^{\frac{1}{2}\binom{n}{2}}$$

for n sufficiently large. □

4.6.2 Proof of Theorem 4.34 for $r = 4$

Given two colors R and B, consider a 4-template P of order n in which every edge of K_n has palette $\{R, B\}$. For a constant $0 < \varepsilon \ll 1$ and a graph G with $e(G) > \binom{n}{2} - 2\varepsilon n$, we can easily check that P is a Gallai 4-template of G and $|\mathrm{Ga}(P,G)| = 2^{e(G)} > 4^{\frac{1}{2}\binom{n}{2} - \varepsilon n}$. This indicates that Lemma 4.37 fails to hold when $r = 4$. Instead, we shall apply the same technique as for 3-colorings: prove a stability result to determine the approximate structure of r-templates, which would contain too many Gallai r-colorings, and then apply this together with Theorem 4.20 to obtain the desired bound.

Lemma 4.39. *Let $n^{-1/3} \ll \delta \ll 1$. Let G be a graph of order n with $e(G) \geq (1-\delta)\binom{n}{2}$. Assume that P is a Gallai 4-template of G with $|\mathrm{Ga}(P,G)| > 2^{(1-\delta)\binom{n}{2}}$. Then the number of triangles T of K_n with $\sum_{e\in T}|P(e)| = 6$ and $P(e) = P(e')$ for every $e,\ e' \in T$ is at least $(1-16\delta)\binom{n}{3}$.*

Proof. Let $\mathcal{T}$ be the collection of triangles of K_n. We define

$$\begin{aligned}
\mathcal{T}_1 &= \left\{T \in \mathcal{T} \mid \textstyle\sum_{e\in T}|P(e)| = 6 \text{ and } P(e) = P(e') \text{ for every } e,\ e' \in T\right\},\\
\mathcal{T}_2 &= \{T \in \mathcal{T} \mid \exists\, e \in T,\ |P(e)| = 0\},\\
\mathcal{T}_3 &= \{T \in \mathcal{T} \mid T = \{e_1,e_2,e_3\},\ |P(e_1)| = 4,\ |P(e_2)| = |P(e_3)| = 1\},\\
\mathcal{T}_4 &= \left\{T \in \mathcal{T} \setminus (\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3) \mid \textstyle\sum_{e\in T}|P(e)| \geq 6\right\},\\
\mathcal{T}_5 &= \left\{T \in \mathcal{T} \setminus \mathcal{T}_2 \mid \textstyle\sum_{e\in T}|P(e)| \leq 5\right\}.
\end{aligned}$$

Let $|\mathcal{T}_1| = \alpha\binom{n}{3}$, $|\mathcal{T}_2| = \beta\binom{n}{3}$, $|\mathcal{T}_3| = \eta\binom{n}{3}$ and $|\mathcal{T}_4| = \gamma\binom{n}{3}$. Then $|\mathcal{T}_5| = (1-\alpha-\beta-\eta-\gamma)\binom{n}{3}$. Since G satisfies $e(G) \geq (1-\delta)\binom{n}{2}$ and P is a Gallai template, we have $|\mathcal{T}_2| \leq \delta\binom{n}{2}\cdot n \leq 6\delta\binom{n}{3}$, and therefore $\beta \leq 6\delta$. Observe that for every $T \in \mathcal{T}_4$, the template P contains a rainbow triangle with edge set T; therefore, we obtain that $|\mathcal{T}_4| \leq RT(P) \leq n^{-1/3}\binom{n}{3}$, which gives $\gamma \leq n^{-1/3}$.

Define for $e \in E(K_n)$ the weight function

$$w(e) = \begin{cases} 1 & \text{if } P(e) = \emptyset \\ |P(e)| & \text{otherwise.} \end{cases}$$

Assume that $\alpha < 1-16\delta$. Similarly, as in Lemma 4.32, the number of Gallai 4-colorings of G which is a subtemplate of P satisfies

$$\begin{aligned}
\log|\mathrm{Ga}(P,G)| &\leq \log\left(\textstyle\prod_{T\in\mathcal{T}_1} 2^3 \prod_{T\in\mathcal{T}_2} 4^2 \prod_{T\in\mathcal{T}_3} 4 \prod_{T\in\mathcal{T}_4} 4^3 \prod_{T\in\mathcal{T}_4} 4\right)\cdot \tfrac{1}{n-2}\\
&\leq (3\alpha + 4\beta + 2\eta + 6\gamma + 2(1-\alpha-\beta-\eta-\gamma))\,\tfrac{1}{3}\tbinom{n}{2}\\
&= (2+\alpha+2\beta+4\gamma)\,\tfrac{1}{3}\tbinom{n}{2} < (2+(1-16\delta)+13\delta)\,\tfrac{1}{3}\tbinom{n}{2} = (1-\delta)\tbinom{n}{2}.
\end{aligned}$$

This contradicts the assumption that $|\mathrm{Ga}(P,G)| > 2^{(1-\delta)\binom{n}{2}}$. □

Similarly, as in Theorem 4.30, applying Lemmas 4.19 and 4.39, we obtain the following.

Theorem 4.40. *Let $n^{-1/3} \ll \delta \ll 1$. Let G be a graph of order n with $e(G) \geq (1-\delta)\binom{n}{2}$. Assume that P is a Gallai 4-template of G with $|\mathrm{Ga}(P,G)| > 2^{(1-\delta)\binom{n}{2}}$. Then there exist two colors $i,\ j \in [4]$ such that the* *number of edges of K_n with palette $\{i,j\}$ is at least* $(1-145\cdot 16\delta)\binom{n}{2}$.

Proof of Theorem 4.34 for $r=4$. Let $\mathcal{C}$ be the collection of containers given by Theorem 4.16 for $r=4$. Note that every Gallai 4-coloring of G is a subtemplate of some $P\in\mathcal{C}$. Let $\delta=\log^{-11} n$. We define

$$\mathcal{C}_1=\left\{P\in\mathcal{C}:\ |\mathrm{Ga}(P,G)|\le 2^{(1-\delta)\binom{n}{2}}\right\},\quad \mathcal{C}_2=\left\{P\in\mathcal{C}:\ |\mathrm{Ga}(P,G)|> 2^{(1-\delta)\binom{n}{2}}\right\}.$$

Similarly, as in the proof of Theorem 4.1, applying Theorems 4.16, 4.20, and 4.40, we obtain that

$$\begin{aligned}|\mathrm{Ga}(\mathcal{C},G)|=|\mathrm{Ga}(\mathcal{C}_1,G)|+|\mathrm{Ga}(\mathcal{C}_2,G)|&\le 2^{-\frac{n^2}{4\log^{11} n}}2^{\binom{n}{2}}+6\left(2^{e(G)}+2^{-\frac{n}{3\log^2 n}}2^{\binom{n}{2}}\right)\\ &\le 6\cdot 2^{e(G)}+2^{-\frac{n}{4\log^2 n}}2^{\binom{n}{2}}<4^{\lfloor n^2/4\rfloor}.\end{aligned}$$

4.6.3 Proof of Theorem 4.35

A *book* of size q consists of q triangles sharing a common edge, which is known as the *base* of the book. We write $\mathrm{bk}(G)$ for the size of the largest book in a graph G and call it the *booksize* of G.

Lemma 4.41. *Let $n,r\in\mathbb{Z}^+$ with $r\ge 4$, $0<\alpha,\beta\ll 1$, and G be a graph of order n. Assume that there exists a partition $V(G)=A\cup B$ satisfying the following conditions:*

(i) $\delta(G[A,B])\ge(\frac{1}{2}-\alpha)n$;

(ii) $\Delta(G[A]),\ \Delta(G[B])\le\beta n$.

Then the number of Gallai r-colorings of G is at most $r^{\lfloor n^2/4\rfloor}$. Furthermore, if $e(G)\neq\lfloor n^2/4\rfloor$, then the number of Gallai r-colorings of G is strictly less than $r^{\lfloor n^2/4\rfloor}$.

Proof. By Condition (i), we have $(\frac{1}{2}-\alpha)n\le|A|,|B|\le(\frac{1}{2}+\alpha)n$. Let $e(G)=\lfloor n^2/4\rfloor+m$. Without loss of generality, we can assume that $m>0$ and $e(G[A])\ge\frac{m}{2}$. Then there exists a matching M in $G[A]$ of size at least $\frac{e(G[A])}{2\Delta(G[A])-1}\ge\frac{m}{4\beta n}$.

For two vertices $u,v\in A$, the number of their common neighbors in B is at least

$$|B|-2\left(|B|-\delta(G[A,B])\right)=2\delta(G[A,B])-|B|\ge 2\left(\frac{1}{2}-\alpha\right)n-\left(\frac{1}{2}+\alpha\right)n\ge\frac{n}{3}.$$

Then, for every $e\in G[A]$, there exists a book graph B_e of size $n/3$ with the base e. Let $\mathcal{B}=\{B_e\mid e\in M\}$. Note that M is a matching, and therefore book graphs in $\mathcal{B}$ are edge-disjoint. Another crucial fact is that for every $B\in\mathcal{B}$, the number of r-colorings of B without rainbow triangles is at most $r\left(r+2(r-1)\right)^{n/3}<r(3r)^{n/3}$, since once we color the base edge, each triangle must be colored in the way that two of its edges

share the same color. Hence, the number of Gallai r-colorings of G is at most

$$\left(r(3r)^{\frac{n}{3}}\right)^{|M|} r^{e(G)-|M|\left(1+2\cdot\frac{n}{3}\right)} = r^{e(G)-(1-\log_r 3)|M|\cdot\frac{n}{3}} \le r^{\lfloor n^2/4\rfloor+m-(1-\log_r 3)\frac{m}{4\beta n}\cdot\frac{n}{3}} < r^{\lfloor n^2/4\rfloor},$$

where the last inequality is given by $\beta \ll 1$. $\square$

Lemma 4.42. *Let $n, r \in \mathbb{Z}^+$ with $r \ge 4$, $0 < \alpha', \beta \ll 1$, $0 < \alpha, \gamma, \xi \ll \varepsilon \ll 1$, and G be a graph of order n with $e(G) \le \lfloor n^2/4\rfloor + \xi n^2$. Assume that there exists a partition $V(G) = A \cup B \cup C$ satisfying the following conditions:*

(i) $d_{G[A,B]}(v) \ge \left(\frac{1}{2}-\alpha\right)n$ *for all but at most γn vertices in $A \cup B$;*

(ii) $\delta(G[A,B]) \ge \left(\frac{1}{2}-\alpha'\right)n$*;*

(iii) $\Delta(G[A])$, $\Delta(G[B]) \le \beta n$*;*

(iv) $0 < |C| \le \gamma n$*;*

(v) for every $v \in C$, both $d(v,A)$, $d(v,B) \ge r\varepsilon n$.

Then the number of Gallai r-colorings of G is strictly less than $r^{\lfloor n^2/4\rfloor}$.

Proof. By Condition (i), we have

$$\left(\frac{1}{2}-\alpha\right)n \le |A|, |B| \le \left(\frac{1}{2}+\alpha\right)n. \tag{4.9}$$

For a vertex v, a set S, a set of colors $\mathcal{R}$ and a coloring of G, let $N(v,S;\mathcal{R})$ be the set of vertices $u \in N(v,S)$, such that uv is colored by some color in $\mathcal{R}$. Let $d(v,S;\mathcal{R}) = |N(v,S;\mathcal{R})|$. Denote by $\mathcal{C}_1$ the set of Gallai r-colorings of G, in which there exist a vertex $v \in C$, and two disjoint sets of colors $\mathcal{R}_1$ and $\mathcal{R}_2$, such that both $d(v,A;\mathcal{R}_1)$, $d(v,B;\mathcal{R}_2) \ge \varepsilon n$. Let $\mathcal{C}_2$ be the set of Gallai r-colorings of G, which are not in $\mathcal{C}_1$.

We first show that $\mathcal{C}_1 = o(r^{\lfloor n^2/4\rfloor})$. We shall count the ways to color G so that the resulting colorings are in $\mathcal{C}_1$. First, we color the edges in $G[C, A\cup B]$; the number of ways is at most $r^{e(G[C,A\cup B])}$. Once we fix the colors of edges in $G[C, A\cup B]$, by the definition of $\mathcal{C}_1$, there exist a vertex $v \in C$, and two disjoint sets of colors $\mathcal{R}_1$ and $\mathcal{R}_2$, such that $d(v,A;\mathcal{R}_1)$, $d(v,B;\mathcal{R}_2) \ge \varepsilon n$. We observe that for every edge $e = uw$ between $N_1 = N(v,A;\mathcal{R}_1)$ and $N_2 = N(v,B;\mathcal{R}_2)$, e either shares the same color with uv, or with vw, as otherwise we would obtain a rainbow triangle uvw. Then the number of ways to color edges in $G[N_1,N_2]$ is at most $2^{e(G[N_1,N_2])} \le r^{\frac{1}{2}e(G[N_1,N_2])}$. Note that by Condition (i), inequality (4.9) and $\alpha, \gamma \ll \varepsilon$, we have

$$e(G[N_1,N_2]) \ge (|N_1| - \gamma n)(|N_2| - 2\alpha n) \ge \frac{1}{2}\varepsilon^2 n^2.$$

Hence, we obtain

$$\begin{aligned}\log_r |\mathcal{C}_1| &\le e(G[C, A \cup B]) + \frac{1}{2}e(G[N_1, N_2]) + (e(G) - e(G[C, A \cup B]) - e(G[N_1, N_2])) \\ &= e(G) - \frac{1}{2}e(G[N_1, N_2]) \le \lfloor n^2/4 \rfloor + \xi n^2 - \frac{1}{4}\varepsilon^2 n^2,\end{aligned}$$

which indicates $|\mathcal{C}_1| = o(r^{\lfloor n^2/4 \rfloor})$ as $\xi \ll \varepsilon$.

It remains to estimate the size of $\mathcal{C}_2$. Recall that for a coloring in $\mathcal{C}_2$, for every vertex $v \in C$, there are no two disjoint sets of colors $\mathcal{R}_1$ and $\mathcal{R}_2$ such that $d(v, A; \mathcal{R}_1)$, $d(v, B; \mathcal{R}_2) \ge \varepsilon n$.

Claim 4.43. *Let $\mathcal{S}$ be a set of r colors. For every coloring in $\mathcal{C}_2$, and every vertex $v \in C$, there exists a color $R \in \mathcal{S}$, such that both $d(v, A; \mathcal{S} \setminus \{R\}) < \varepsilon n$ and $d(v, B; \mathcal{S} \setminus \{R\}) < \varepsilon n$.*

Proof. We arbitrarily fix a coloring in $\mathcal{C}_2$ and a vertex $v \in C$. By Condition (v), there exists a color R such that $d(v, A; R) \ge \varepsilon n$. By the definition of $\mathcal{C}_2$, we obtain that $d(v, B; \mathcal{S} \setminus \{R\}) < \varepsilon n$. Then we also have $d(v, B; R) \ge d(v, B) - d(v, B; \mathcal{S} \setminus \{R\}) \ge r\varepsilon n - \varepsilon n > \varepsilon n$. For the same reason, we obtain that $d(v, A; \mathcal{S} \setminus \{R\}) < \varepsilon n$. □

By Claim 4.43, the number of ways to color edges in $G[C, A \cup B]$ is at most

$$\left(r \sum_{i \le \varepsilon n} \binom{n}{i} \sum_{i \le \varepsilon n} \binom{n}{i} r^{2\varepsilon n} \right)^{|C|} \le \left(4r \left(\frac{ne}{\epsilon n} \right)^{2\varepsilon n} r^{2\varepsilon n} \right)^{|C|} \le r^{((\log_r e - \log_r \varepsilon + 1)2\varepsilon n + 2)|C|} < r^{\frac{|C|n}{3}},$$

where the last inequality is given by $(\log_r e - \log_r \varepsilon + 1)\, 2\varepsilon \ll \frac{1}{3}$ as $\varepsilon \ll 1$. Note that by Conditions (ii)–(iv), we have

- $\delta(G[A, B]) \ge \left(\frac{1}{2} - \alpha'\right) n \ge \left(\frac{1}{2} - \alpha'\right) (|A| + |B|)$;
- $\Delta(G[A]), \Delta(G[B]) \le \beta n \le \frac{\beta}{1-\gamma}(|A| + |B|)$.

Applying Lemma 4.41 on $G[A \cup B]$, we obtain that the number of ways to color edges in $G[A \cup B]$ is at most $r^{\frac{(n-|C|)^2}{4}}$. A trivial upper bound for the ways to color the rest of the edges, that is, the edges in $G[C]$ is $r^{\binom{|C|}{2}}$. Hence, we have

$$\log_r |\mathcal{C}_2| \le \frac{|C|n}{3} + \frac{(n - |C|)^2}{4} + \binom{|C|}{2} = \frac{n^2}{4} - \left(\frac{n}{6} - \frac{3}{4}|C| + \frac{1}{2} \right) |C| \le \lfloor n^2/4 \rfloor - \frac{1}{4},$$

where the last inequality is given by $0 < |C| \le \gamma n$ and $\gamma \ll 1$. Finally, we obtain that the number of Gallai r-colorings of G is

$$|\mathcal{C}_1| + |\mathcal{C}_2| \le o(r^{\lfloor n^2/4 \rfloor}) + r^{\lfloor n^2/4 \rfloor - \frac{1}{4}} < r^{\lfloor n^2/4 \rfloor}.$$

□

Lemma 4.44. *Let $n, r \in \mathbb{Z}^+$ with $r \geq 4$, $\alpha, \beta, \gamma, \xi \ll 1$, and G be a graph of order n with $\lfloor n^2/4 \rfloor < e(G) \leq \lfloor n^2/4 \rfloor + \xi n^2$. Assume, that there exists a partition $V(G) = A \cup B \cup C$ satisfying the following conditions:*

(i) $\delta(G[A, B]) \geq (\frac{1}{2} - \alpha)n$;

(ii) $\Delta(G[A]), \Delta(G[B]) \leq \beta n$;

(iii) $0 < |C| \leq \gamma n$;

(iv) *for every* $v \in C$, $d(v) \geq n/2$.

Then the number of Gallai r-colorings of G is strictly less than $r^{\lfloor n^2/4 \rfloor}$.

Proof. Let $\alpha, \gamma, \xi \ll \varepsilon \ll 1$. Let $C_1 = \{v \in C \mid d(v, A) < r\varepsilon n\}$, and $C_2 = \{v \in C \mid d(v, B) < r\varepsilon n\}$. By Conditions (iii) and (iv), for every $v \in C_1$, we have $d(v, B) \geq \left(\frac{1}{2} - \gamma - r\varepsilon\right) n$. Similarly, for every $v \in C_2$, we have $d(v, A) \geq \left(\frac{1}{2} - \gamma - r\varepsilon\right) n$. Define

$$A' = A \cup C_1, \quad B' = B \cup C_2, \quad C' = C \setminus (C_1 \cup C_2).$$

If $C' = \emptyset$, then we obtain a new partition $V(G) = A' \cup B'$ satisfying the following properties:

- $\delta(G[A', B']) \geq \min\{\left(\frac{1}{2} - \alpha\right) n, \left(\frac{1}{2} - \gamma - r\varepsilon\right) n\} = \left(\frac{1}{2} - \gamma - r\varepsilon\right) n$;
- $\Delta(G[A']), \Delta(G[B']) \leq \min\{(\beta + \gamma)n, (r\varepsilon + \gamma)n\}$.

Together with $e(G) > \lfloor n^2/4 \rfloor$, by Lemma 4.41, we obtain that the number of Gallai r-colorings of G is strictly less than $r^{\lfloor n^2/4 \rfloor}$. Otherwise, we obtain a new partition $V(G) = A' \cup B' \cup C'$ satisfying the following properties:

- $d_{G[A', B']}(v) \geq \left(\frac{1}{2} - \alpha\right) n$ for all but at most γn vertices in $A' \cup B'$;
- $\delta(G[A', B']) \geq \left(\frac{1}{2} - \gamma - r\varepsilon\right) n$;
- $\Delta(G[A']), \Delta(G[B']) \leq \min\{(\beta + \gamma)n, (r\varepsilon + \gamma)n\}$;
- $0 < |C'| \leq |C| \leq \gamma n$;
- for every $v \in C'$, both $d(v, A'), d(v, B') \geq r\varepsilon n$.

Together with $e(G) \leq \lfloor n^2/4 \rfloor + \xi n^2$, by Lemma 4.42, the number of Gallai r-colorings of G is strictly less than $r^{\lfloor n^2/4 \rfloor}$. □

Now, we prove a lemma which is crucial to the proof of Theorem 4.35.

Lemma 4.45. *Let $n, r \in \mathbb{Z}^+$ with $r \geq 4$, $\alpha, \beta, \gamma, \xi \ll 1$, and G be a graph of order n with $\lfloor n^2/4 \rfloor < e(G) \leq \lfloor n^2/4 \rfloor + \xi n^2$. Assume that there exists a partition $V(G) = A \cup B \cup C$ satisfying the following conditions:*

(i) $\delta(G[A, B]) \geq (\frac{1}{2} - \alpha)n$;

(ii) $\Delta(G[A])$, $\Delta(G[B]) \leq \beta n$;

(iii) $|C| \leq \gamma n$.

Then the number of Gallai r-colorings of G is strictly less than $r^{\lfloor n^2/4 \rfloor}$.

Proof. By Lemma 4.41, we can assume that $|C| > 0$ without loss of generality. We begin with the graph G, greedily remove a vertex in C with degree strictly less than $|G|/2$ in G to obtain a smaller subgraph. Let G' be the resulting graph when the algorithm terminates, and $n' = |V(G')|$. We remark that G' is not unique and it depends on the order of removing vertices. Without loss of generality, we can assume that $n' < n$, as otherwise we are done by applying Lemma 4.44 on G.

Let $A' = A$, $B' = B$, and $C' = V(G') \cap C$. Clearly, we have $G' = G[A' \cup B' \cup C']$. Furthermore, by the mechanics of the algorithm, we have

$$e(G) \leq e(G') + \frac{1}{2}\left(\binom{n}{2} - \binom{n'}{2}\right). \tag{4.10}$$

We first claim that $e(G') > \lfloor (n')^2/4 \rfloor$, as otherwise we would have

$$e(G) \leq \lfloor (n')^2/4 \rfloor + \frac{1}{2}\left(\binom{n}{2} - \binom{n'}{2}\right) \leq \lfloor n^2/4 \rfloor,$$

which contradicts the assumption of the lemma. On the other hand, since $n' \geq (1-\gamma)n$, we obtain that

$$e(G) \leq \lfloor n^2/4 \rfloor + \xi n^2 \leq \lfloor (n')^2/4 \rfloor + \frac{\gamma + 2\xi}{2(1-\gamma)^2}(n')^2.$$

Let $\xi' = \frac{\gamma+2\xi}{2(1-\gamma)^2}$. Then we have

$$\lfloor (n')^2/4 \rfloor < e(G') \leq \lfloor (n')^2/4 \rfloor + \xi'(n')^2. \tag{4.11}$$

If $C' = \emptyset$, we obtain a vertex partition $V(G') = A' \cup B'$ satisfying:

- $\delta(G'[A', B']) \geq (\frac{1}{2} - \alpha)n \geq (\frac{1}{2} - \alpha)n'$;
- $\Delta(G'[A])$, $\Delta(G'[B]) \leq \beta n \leq \frac{\beta}{1-\gamma}n'$.

Together with (4.11), by Lemma 4.41, we obtain that the number of Gallai r-colorings of G', denoted by $|\mathcal{C}(G')|$, is strictly less than $r^{\lfloor n^2/4 \rfloor}$. Otherwise, we find the partition $V(G') = A' \cup B' \cup C'$ satisfying:

- $\delta(G'[A', B']) \geq (\frac{1}{2} - \alpha)n \geq (\frac{1}{2} - \alpha)n'$;
- $\Delta(G'[A]), \Delta(G'[B]) \leq \frac{\beta}{1-\gamma}n'$;
- $0 < |C'| \leq \gamma n \leq \frac{\gamma}{1-\gamma}n'$;
- for every $v \in C'$, $d(v) \geq \frac{n'}{2}$.

Together with (4.11), by Lemma 4.44, we obtain that $|\mathcal{C}(G')| < r^{\lfloor (n')^2/4 \rfloor}$. Combining with (4.10), we conclude that the number of Gallai r-colorings of G, denoted by $|\mathcal{C}(G)|$, satisfies

$$\log_r |\mathcal{C}(G)| \leq \log_r |\mathcal{C}(G')| + (e(G) - e(G')) < \lfloor (n')^2/4 \rfloor + \frac{1}{2}\left(\binom{n}{2} - \binom{n'}{2}\right) \leq \lfloor n^2/4 \rfloor,$$

which completes the proof. $\square$

Another important tool we need is the stability property of book graphs proved by Bollobás and Nikiforov [22].

Theorem 4.46. *[22] For every $0 < \alpha < 10^{-5}$ and every graph G of order n with $e(G) \geq (\frac{1}{4} - \alpha)n^2$, either*

$$\mathrm{bk}(G) > \left(\frac{1}{6} - 2\alpha^{1/3}\right)n$$

or G contains an induced bipartite graph G_1 of order at least $(1 - \alpha^{1/3})n$ and with minimum degree

$$\delta(G_1) \geq \left(\frac{1}{2} - 4\alpha^{1/3}\right)n.$$

Proof of Theorem 4.35: Let $e(G) = \lfloor n^2/4 \rfloor + m$, where $0 < m \leq \xi n^2$. We construct a family $\mathcal{B}$ of book graphs by the following algorithm. We start the algorithm with $\mathcal{B} = \emptyset$ and $G_0 = G$. In the i-th iteration step, if there exists a book graph B of size $\frac{n}{7}$ in G_i, we let $\mathcal{B} = \mathcal{B} \cup \{B\}$, and $G_i = G_{i-1} - e$, where e is the base edge of B. The algorithm terminates when there is no book graph of size $n/7$. Let E_0 be the set of base edges of $\mathcal{B}$, and $\tau = 7/(1 - \log_r 3)$.

Suppose that $|\mathcal{B}| \geq 2\tau m$. Since $|E_0| = |\mathcal{B}| \geq 2\tau m$, the edge set E_0 contains a matching M of size $\frac{|E_0|}{2(n-1)-1} > \tau m/n$. Let $\mathcal{B}'$ be the set of book graphs in $\mathcal{B}$ whose base edges are in M. Since M is a matching, book graphs in $\mathcal{B}'$ are edge-disjoint. Note that for every $B \in \mathcal{B}$, the number of r-colorings of B without rainbow triangles is at most $r(r + 2(r-1))^{n/7} < r(3r)^{n/7}$. Then the number of Gallai colorings of G is at most

$$\left(r(3r)^{\frac{n}{7}}\right)^{|M|} r^{e(G)-|M|(1+2\cdot\frac{n}{7})} = r^{\lfloor n^2/4\rfloor+m-(1-\log_r 3)|M|\frac{n}{7}} < r^{\lfloor n^2/4\rfloor+m-m} = r^{\lfloor n^2/4\rfloor}.$$

It remains to consider the case for $|\mathcal{B}| < 2\tau m$. Without loss of generality, we can assume that there is no matching of size greater than $\tau m/n$ in E_0. Let $G' = G - E_0$. Then we have

$$e(G') > \lfloor n^2/4\rfloor - (2\tau - 1)\, m.$$

Furthermore, by the construction of G', we obtain that $\mathrm{bk}(G') < n/7$. Let $\alpha = (2\tau - 1)\,\xi$. By applying Theorem 4.46 on G', we obtain that there is a vertex partition $V(G') = A' \cup B' \cup C'$ with $|C'| \leq \alpha^{1/3} n$, such that A', B' are independent sets, and

$$\delta(G'[A', B']) \geq \left(\frac{1}{2} - 4\alpha^{1/3}\right) n.$$

Let G_0 be the spanning subgraph of G with edge set E_0. For a small constant β with $\xi \ll \beta \ll 1$, let V_0 be the set of vertices in G_0 with degree more than βn. Since $|E_0| < 2\tau m \leq 2\tau\xi n^2$, we have $|V_0| \leq (4\tau\xi/\beta) n \leq \beta n$. Let $A = A' \setminus V_0$, $B = B' \setminus V_0$, and $C = C' \cup V_0$. Then we obtain a vertex partition $V(G) = A \cup B \cup C$ satisfying the following conditions:

- $\delta(G[A, B]) \geq (\frac{1}{2} - 4\alpha^{1/3} - \beta)n$;
- $\Delta(G[A]), \Delta(G[B]) \leq \beta n$;
- $|C| \leq (\alpha^{1/3} + \beta)n$.

By Lemma 4.45, we obtain that the number of Gallai r-colorings of G is strictly less than $r^{\lfloor n^2/4\rfloor}$. □

4.6.4 Proof of Theorem 4.36

We say that a graph G is *t-far* from being k-partite if $\chi(G') > k$ for every subgraph $G' \subset G$ with $e(G') > e(G) - t$. We will use the following theorem of Balogh, Bushaw, Collares, Liu, Morris, and Sharifzadeh [7].

Theorem 4.47. *[7] For every $n, k, t \in \mathbb{N}$, the following holds. Every graph G of order n which is t-far from being k-partite contains at least*

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

copies of K_{k+1}.

Proposition 4.48. *Let $n \in \mathbb{N}$ and $0 < \varepsilon \leq 1$. Every graph F on at least εn vertices, which contains at most $n^{-1/3}\binom{n}{3}$ triangles, satisfies*

$$e(F) \leq \frac{|F|^2}{4} + \frac{e^4}{6n^{1/3}\varepsilon^3}|F|^2.$$

Proof. Let $t = \frac{e^4}{6n^{1/3}\varepsilon^3}|F|^2$. Assume that $e(F) > \frac{|F|^2}{4} + t$. Then F is t-far from being bipartite. By Theorem 4.47, the number of triangles in F is at least

$$\frac{|F|}{2e^4}\left(e(F) + t - \frac{|F|^2}{4}\right) > \frac{|F|}{2e^4} \cdot 2t = \frac{1}{6n^{1/3}\varepsilon^3}|F|^3 > n^{-1/3}\binom{n}{3},$$

which gives a contradiction. □

For an r-template P of order n, we say that an edge e of K_n is an *r-edge* of P if $|P(e)| \geq 3$. An r-edge e is *typical* if the number of rainbow triangles containing e is at most $n^{11/12}$. We then immediately obtain the following proposition.

Proposition 4.49. *For an r-template of order n containing at most $n^{-1/3}\binom{n}{3}$ rainbow triangles, the number of r-edges of P, which is not typical, is at most $n^{11/6}$.*

We now prove the following lemma.

Lemma 4.50. *Let $n, r \in \mathbb{N}$ with $r \geq 4$, and $n^{-1/33} \ll \xi \leq \frac{1}{2}\log^{-11} n \ll 1$. Assume that G is a graph of order n with $(\frac{1}{4} + 3\xi)n^2 \leq e(G) \leq (\frac{1}{2} - 3\xi)n^2$, and P is a Gallai r-template of G. Then, for sufficiently large n,*

$$\log_r |\mathrm{Ga}(P, G)| \leq \frac{n^2}{4} - \xi^3\frac{n^2}{2} + 4n^{23/12}.$$

Proof. We first construct a subset I of $[n]$ and a sequence of graphs $\{G_0, G_1, \ldots, G_\ell\}$ by the following algorithm. We start the algorithm with $I = \emptyset$ and $G_0 = G$. In the i-th iteration step, we either add a vertex v to I, whose degree is at most $(\frac{1}{2} - \xi^2)(|G_i| - 1)$ in the graph G_i, or add a pair of vertices $\{u, v\}$ to I, where uv is a typical r-edge satisfying $|N_{G_i}(u) \cap N_{G_i}(v)| \geq 2\xi^2(|G_i| - 2)$. In both cases, we define $G_{i+1} = G - I$. The algorithm terminates when neither of the above types of vertices exists.

Assume that the algorithm terminates after ℓ steps. Let $G' = G_\ell$ and $k = |G'|$. We now make the following claim.

Claim 4.51.

$$\log_r |\mathrm{Ga}(P, G)| \leq \left(\frac{1}{2} - \xi^2\right)\left(\frac{n^2}{2} - \frac{k^2}{2}\right) + 3n^{23/12} + \log_r |\mathrm{Ga}(P, G')|.$$

Proof. In the i-th iteration step of the above algorithm, if we add to I a single vertex v, then the number of ways to color the incident edges of v in G_i satisfies

$$\log_r \prod_{\substack{e \text{ is incident to} \\ v \text{ in } G_i}} |P(e)| \le d_{G_i}(v) \le (\tfrac{1}{2} - \xi^2)(|G_i| - 1).$$

Now we assume that what we add is a pair of vertices $\{u, v\}$. For every $w \in N_{G_i}(u) \cap N_{G_i}(v)$, vertices uvw either span a rainbow triangle in P, or satisfy $|P(uw)| = |P(vw)| = 1$. Together with the fact that uv is a typical r-edge, we obtain that the number of ways to color the edges, which are incident to v or u in G_i, satisfies

$$\begin{aligned} \log_r \prod_{\substack{e \text{ is incident to} \\ u \text{ or } v \text{ in } G_i}} |P(e)| &\le |G_i| - 2 - |N_{G_i}(u) \cap N_{G_i}(v)| + 2n^{11/12} + 1 \\ &\le (1 - 2\xi^2)(|G_i| - 2) + 2n^{11/12} + 1. \end{aligned}$$

From the above discussion, we conclude that the number of ways to color edges in $E(G) - E(G')$ satisfies

$$\log_r \prod_{e \in E(G) - E(G')} |P(e)| \le \left(\tfrac{1}{2} - \xi^2\right) \left(\tfrac{n^2}{2} - \tfrac{k^2}{2}\right) + n(1 + 2n^{11/12}),$$

which implies the claim. $\square$

We now split the proof into several cases.

Case 1: $k \le \xi^2 n$.

Then $|\mathrm{Ga}(P, G')| \le r^{k^2/2} \le r^{\xi^4 n^2/2}$, and therefore by Claim 4.51 and $\xi \ll 1$, we obtain that

$$\log_r |\mathrm{Ga}(P, G)| \le \left(\frac{1}{2} - \xi^2\right) \frac{n^2}{2} + 3n^{23/12} + \xi^4 n^2/2 \le \frac{n^2}{4} - \xi^2 \frac{n^2}{4} + 3n^{23/12}.$$

Case 2: $e(G') > \left(\frac{1}{2} - 2\xi\right) k^2$ and $k > \xi^2 n$.

Since $2\xi \le \log^{-11} n \le \log^{-11} k$, for sufficiently large n, Theorem 4.34 indicates that $|\mathrm{Ga}(P, G')| \le r^{k^2/4}$. We claim that $k \le (1 - \xi)n$, as otherwise we would have

$$e(G) \ge e(G') > \left(\frac{1}{2} - 2\xi\right) k^2 > \left(\frac{1}{2} - 2\xi\right) (1 - \xi)^2 n^2 \ge \left(\frac{1}{2} - 3\xi\right) n^2,$$

which is contradiction with the assumption of the lemma. Therefore, by Claim 4.51, we obtain that

$$\log_r |\mathrm{Ga}(P,G)| \le \left(\tfrac{1}{2}-\xi^2\right)\left(\tfrac{n^2}{2}-\tfrac{k^2}{2}\right)+3n^{23/12}+\tfrac{k^2}{4} \le \tfrac{n^2}{4}-\xi^2\tfrac{n^2}{2}+\xi^2\tfrac{k^2}{2}+3n^{23/12}$$
$$\le \tfrac{n^2}{4}-\xi^2\tfrac{n^2}{2}+\xi^2(1-\xi)^2\tfrac{n^2}{2}+3n^{23/12} \le \tfrac{n^2}{4}-\xi^3\tfrac{n^2}{2}+3n^{23/12}.$$

Case 3: $e(G') < \left(\frac{1}{4}+2\xi\right)k^2$ and $k > \xi^2 n$.

Since $2\xi \ll 1$, Theorem 4.35 indicates that $|\mathrm{Ga}(P,G')| \le r^{k^2/4}$. We claim that $k \le (1-\xi)n$, as otherwise we would have

$$e(G) < \left(\tfrac{n^2}{2}-\tfrac{k^2}{2}\right)+\left(\tfrac{1}{4}+2\xi\right)k^2 < \tfrac{n^2}{2}-\left(\tfrac{1}{4}-2\xi\right)k^2$$
$$< \tfrac{n^2}{2}-\left(\tfrac{1}{4}-2\xi\right)(1-\xi)^2n^2 \le \tfrac{n^2}{2}-\left(\tfrac{1}{4}-3\xi\right)n^2=\left(\tfrac{1}{4}+3\xi\right)n^2,$$

which is contradiction with the assumption of the lemma. Similarly, as in Case 2, we obtain that

$$\log_r |\mathrm{Ga}(P,G)| \le \left(\frac{1}{2}-\xi^2\right)\left(\frac{n^2}{2}-\frac{k^2}{2}\right)+3n^{23/12}+\frac{k^2}{4} \le \frac{n^2}{4}-\xi^3\frac{n^2}{2}+3n^{23/12}.$$

Case 4: $(\frac{1}{4}+2\xi)k^2 \le e(G') \le (\frac{1}{2}-2\xi)k^2$ and $k > \xi^2 n$.

Denote by $e_r(G')$ the number of r-edges of P in G'. Let $A=\{v\in V(G') \mid d_{G'}(v) \le \left(\frac{1}{2}+\xi\right)k\}$.

Claim 4.52. *All the typical r-edges of G' have both endpoints in A.*

Proof. First, by the construction of G', we have the following two properties: for every $v \in V(G')$,

$$d_{G'}(v) > \left(\frac{1}{2}-\xi^2\right)(k-1), \tag{4.12}$$

and for every typical r-edge uv in G',

$$d_{G'}(u)+d_{G'}(v) \le 2+(k-2)+|N_{G_i}(u)\cap N_{G_i}(v)| < (1+2\xi^2)k. \tag{4.13}$$

Suppose that there exists a typical r-edge uv such that u is not in A, i.e. $d_{G'}(u) > \left(\frac{1}{2}+\xi\right)k$. Then by (4.12) and $\xi \ll 1$, we have

$$d_{G'}(u)+d_{G'}(v) > \left(\frac{1}{2}+\xi\right)k+\left(\frac{1}{2}-\xi^2\right)(k-1) > (1+2\xi^2)k,$$

which contradicts (4.13). □

Subcase 4.1: $|A| \le \xi k$.

By Proposition 4.49 and Claim 4.52, we have

$$e_r(G') \le \binom{|A|}{2} + n^{11/6} \le \xi^2 \frac{k^2}{2} + n^{11/6}.$$

Therefore, together with the assumption of Case 4, we obtain that

$$\begin{aligned}\log_r |\mathrm{Ga}(P, G')| &\le \log_r \left(r^{e_r(G')} 2^{e(G') - e_r(G')}\right) \le \tfrac{1}{2}\left(e(G') + e_r(G')\right) \\ &\le \tfrac{1}{2}\left(\left(\tfrac{1}{2} - 2\xi\right) k^2 + \xi^2 \tfrac{k^2}{2} + n^{11/6}\right) = \tfrac{k^2}{4} - \left(\xi - \tfrac{1}{4}\xi^2\right) k^2 + \tfrac{1}{2} n^{11/6}.\end{aligned}$$

Then by Claim 4.51,

$$\begin{aligned}\log_r |\mathrm{Ga}(P, G)| &\le \left(\tfrac{1}{2} - \xi^2\right)\left(\tfrac{n^2}{2} - \tfrac{k^2}{2}\right) + 3n^{23/12} + \tfrac{k^2}{4} - \left(\xi - \tfrac{1}{4}\xi^2\right) k^2 + \tfrac{1}{2} n^{11/6} \\ &\le \tfrac{n^2}{4} - \xi^2 \tfrac{n^2}{2} - \left(\xi - \tfrac{3}{4}\xi^2\right) k^2 + 4n^{23/12} \le \tfrac{n^2}{4} - \xi^2 \tfrac{n^2}{2} + 4n^{23/12},\end{aligned}$$

where the last inequality is given by $\xi \ll 1$.

Subcase 4.2: $|A| > \xi k$.

By the definition of A, the number of non-edges of G' is at least

$$\frac{1}{2}\left(k - 1 - \left(\frac{1}{2} + \xi\right) k\right)|A| = \frac{1}{2}\left(\left(\frac{1}{2} - \xi\right) k - 1\right)|A|. \tag{4.14}$$

We first claim that

$$|A| \le \frac{1 - 8\xi}{1 - 2\xi} k, \tag{4.15}$$

as otherwise we would obtain that the number of non-edges of G' is more than

$$\frac{1}{2}\left(\frac{1}{2} - \xi\right) k \cdot \frac{1 - 8\xi}{1 - 2\xi} k - \frac{|A|}{2} \ge \left(\frac{1}{4} - 2\xi\right) k^2 - \frac{k}{2}$$

which contradicts the assumption of Case 4. Inequality (4.15) implies that

$$(1 - 2\xi)k - |A| \ge 4\xi k. \tag{4.16}$$

By Propositions 4.48 and 4.49, since $|A| > \xi k > \xi^3 n$, we have

$$e_r(G') \le e(G'[A]) + n^{11/6} \le \frac{|A|^2}{4} + \frac{e^4}{6n^{1/3}\xi^9}|A|^2 + n^{11/6},$$

as otherwise we would find more than $n^{-1/3}\binom{n}{3}$ rainbow triangles, which contradicts the assumption that P is a Gallai r-template of G. Since $\xi \gg n^{-1/33}$, we have

$$e_r(G') \leq \frac{|A|^2}{4} + \frac{\xi^2}{2}|A|^2 + n^{11/6}. \tag{4.17}$$

Combining (4.14), (4.16) and (4.17), we have

$$\begin{aligned}
\log_r |\mathrm{Ga}(P, G')| &\leq \tfrac{1}{2}\left(e(G') + e_r(G')\right) \\
&\leq \tfrac{1}{2}\left(\tbinom{k}{2} - \tfrac{1}{2}\left(\left(\tfrac{1}{2} - \xi\right)k - 1\right)|A| + \tfrac{|A|^2}{4} + \tfrac{\xi^2}{2}|A|^2 + n^{11/6}\right) \\
&\leq \tfrac{k^2}{4} - \tfrac{|A|}{8}\left((1-2\xi)k - |A|\right) + \tfrac{\xi^2}{4}|A|^2 + \tfrac{1}{2}n^{11/6} \\
&\leq \tfrac{k^2}{4} - \tfrac{\xi}{2}|A|k + \tfrac{\xi^2}{4}|A|^2 + \tfrac{1}{2}n^{11/6}.
\end{aligned}$$

Then by Claim 4.51 and the assumption of Subcase 4.2, we obtain that

$$\begin{aligned}
\log_r |\mathrm{Ga}(P, G)| &\leq \left(\tfrac{1}{2} - \xi^2\right)\left(\tfrac{n^2}{2} - \tfrac{k^2}{2}\right) + 3n^{23/12} + \tfrac{k^2}{4} - \tfrac{\xi}{2}|A|k + \tfrac{\xi^2}{4}|A|^2 + \tfrac{1}{2}n^{11/6} \\
&< \left(\tfrac{1}{2} - \xi^2\right)\left(\tfrac{n^2}{2} - \tfrac{k^2}{2}\right) + 3n^{23/12} + \tfrac{k^2}{4} - \tfrac{\xi^2}{2}k^2 + \tfrac{\xi^2}{4}n^2 + \tfrac{1}{2}n^{11/6} \\
&\leq \tfrac{n^2}{4} - \xi^2\tfrac{n^2}{4} + 4n^{23/12}.
\end{aligned}$$

□

Proof of Theorem 4.36. Let $\mathcal{C}$ be the collection of containers given by Theorem 4.16. Theorem 4.16 indicates that every Gallai r-coloring of G is a subtemplate of some $P \in \mathcal{C}$ and $|\mathcal{C}| \leq 2^{cn^{-1/3}\log^2 n\binom{n}{2}}$ for some constant c, which only depends on r. We may assume that all templates P in $\mathcal{C}$ are Gallai r-templates of G. By Property (ii) of Theorem 4.16, we always have $\mathrm{RT}(P) \leq n^{-1/3}\binom{n}{3}$. Suppose that for a template P there exists an edge $e \in E(G)$ with $|P(e)| = 0$. Then we would obtain $|\mathrm{Ga}(P, G)| = 0$ as a Gallai r-coloring of G requires at least one color on each edge. Now applying Lemma 4.50 on every container $P \in \mathcal{C}$, we obtain that the number of Gallai r-colorings of G is at most

$$\sum_{P\in\mathcal{C}} |\mathrm{Ga}(P, G)| \leq |\mathcal{C}| \cdot r^{\frac{n^2}{4} - \xi^3\frac{n^2}{2} + 4n^{23/12}} < r^{\lfloor n^2/4 \rfloor},$$

where the last inequality follows from $\xi \gg n^{-1/36}$ for n sufficiently large.

Chapter 5

An analogue of the Erdős–Gallai theorem for random graphs

5.1 Introduction

A celebrated theorem of Erdős and Gallai [41] from 1959 determines the maximum number of edges in an n-vertex graph with no k-vertex path P_k.

Theorem 5.1 (Erdős and Gallai [41]). *For $n, k \geq 2$, if G is an n-vertex graph with no copy of P_k, then the number of edges of G satisfies $e(G) \leq \frac{1}{2}(k-2)n$. If n is divisible by $k-1$, then the maximum is achieved by a union of disjoint copies of K_{k-1}.*

An important direction of combinatorics in recent years is the study of sparse random analogues of classical extremal results; that is, the extent to which of these results remain true in a random setting. For graphs G and F, we write $\mathrm{ex}(G, F)$ for the maximum number of edges in an F-free subgraph of G. For example, the Erdős–Gallai theorem asserts that $\mathrm{ex}(K_n, P_k) = \frac{1}{2}(k-2)n$ if n is divisible by $k-1$.

The study of the random variable $\mathrm{ex}(G, F)$, where G is the Erdős-Rényi random graph $G(n, p)$, was initiated by Babai, Simonovits and Spencer [3], and by Frankl and Rödl [45]. After efforts by several researchers [55, 56, 69, 71, 72, 98], Conlon and Gowers [28] and Schacht [95] finally proved a sparse random version of the Erdős-Stone theorem, showing a *transference principle* of Turán-type results, that is, when a random graph inherits its (relative) extremal properties from the classical deterministic case. Note that via the hypergraph container method the same results were proved ([11] and [94]), even when $|F|$ is a reasonable large function of n. A special case of this result, when F is the k-vertex path P_k, can be viewed as a weak analogue (as the Turán density is 0) of the Erdős-Gallai theorem on the random graph for paths with a fixed size. In this chapter, we investigate the random analogue of the Erdős-Gallai theorem for general paths, whose length might increase with the order of the random graph.

We say that events A_n in a probability space hold *asymptotically almost surely* (or a.a.s.), if the probability that A_n holds tends to 1 as n goes to infinity. The typical appearance of long paths and cycles is one of the most thoroughly studied direction in random graph theory. Over the past decades, there were many diverse and beautiful results in this subject. In a seminal paper, Ajtai, Komlós and Szemerédi [1], confirming a conjecture of Erdős, proved that for $p = \frac{c}{n}$ with $c > 1$, $G(n,p)$ contains a path of length $\alpha(c)n$ a.a.s. where $\lim_{c\to\infty} \alpha(c) = 1$. Frieze [46] later determined the asymptotics of the number of vertices not covered by a longest path in $G(n,p)$. For Hamiltonicity, Bollobás [21] and Komlós and Szemerédi [73] independently proved that for $p \geq \frac{\ln n + \ln\ln n + \omega(1)}{n}$, the random graph $G(n,p)$ is a.a.s. Hamiltonian. Turán-type results for long cycles in $G(n,p)$ was also studied under the name of *global resilience*, that is, the minimum number r such that one can destroy the graph property by deleting r edges. Dellamonica Jr, Kohayakawa, Marciniszyn and Steger [30] determined the global resilience of $G(n,p)$ with respect to the property of containing a cycle of length proportional to the number of vertices. Very recently, Krivelevich, Kronenberg and Mond [80] studied the transference principle in the context of long cycles and in particular showing the following.

Theorem 5.2 (Corollary 1.10 in [80])**.** *For every $0 < \beta < \frac{1}{5}$, there exists $C > 0$ such that if $G = G(N,p)$ where $p \geq \frac{C}{N}$, then for any $\frac{C_1}{\ln(1/\beta)} \cdot \ln N \leq n \leq (1 - C_2\beta)N$, with probability $1 - e^{\Omega(N)}$,*

$$\mathrm{ex}(G(N,p), C_n) \leq \left(\frac{\mathrm{ex}(K_N, C_n)}{\binom{N}{2}} + \beta \right) e(G(N,p)), \tag{5.1}$$

where $C_1, C_2 > 0$ are absolute constants.

We aim to explore the global resilience of general long paths. More formally, given integers $N > n$, we are interested in determining the asymptotic behavior of random variable $\mathrm{ex}(G(N,p), P_{n+1})$ as N and n go to infinity at the same time.

We start with an observation, which is proved in Section 5.3.

Proposition 5.3. *For every $\frac{1}{N^2} \ll p \leq \frac{1}{N}$ and $n \geq 2$, a.a.s. we have $\mathrm{ex}(G(N,p), P_{n+1}) = \Theta(pN^2)$. In particular, a.a.s. $\mathrm{ex}(G(N, 1/N), P_{n+1}) \geq N/15$.*

Therefore, throughout this chapter, we naturally restrict ourselves to the regime $p \geq 1/N$ and have the following trivial lower bound

$$a.a.s. \quad \mathrm{ex}(G(N,p), P_{n+1}) \geq \mathrm{ex}\left(G\left(N, 1/N\right), P_{n+1}\right) \geq N/15. \tag{5.2}$$

We prove the following results.

Theorem 5.4. *Let $3n \leq N \leq ne^{2n}$. The following hold a.a.s. as n approaches infinity.*

(i) For $p \geq \left(\ln \frac{N}{n}\right)/(6n)$, we have $\frac{1}{4}pnN \leq \mathrm{ex}(G(N,p), P_{n+1}) \leq 18pnN$.

(ii) Let $\omega = \left(\ln \frac{N}{n}\right)/(np)$*. For* $N^{-1} \le p \le \left(\ln \frac{N}{n}\right)/(6n)$*, we have*

$$\frac{1}{75}\frac{\omega}{\ln \omega} pnN \le \mathrm{ex}(G(N,p), P_{n+1}) \le 8\frac{\omega}{\ln \omega} pnN.$$

Theorem 5.5. *Let* $N \ge ne^{2n}$*. The following hold a.a.s. as* n *approaches infinity.*

(i) For $p \ge N^{-\frac{2}{5n}}$*, we have* $\frac{1}{16}nN \le \mathrm{ex}(G(N,p), P_{n+1}) \le \frac{1}{2}nN$*.*

(ii) Let $\omega = (\ln N)/(np)$*. For* $N^{-1} \le p \le N^{-\frac{2}{5n}}$*, we have*

$$\frac{1}{75}\frac{\omega}{\ln \omega} pnN \le \mathrm{ex}(G(N,p), P_{n+1}) \le 8\frac{\omega}{\ln \omega} pnN.$$

Remark 5.6. *Assume that* n *is even. Then* (5.1) *together with* $\mathrm{ex}(K_N, C_n) \le nN^{1+2/n}$ *(e.g., see [90]) imply that*

$$\begin{aligned}\mathrm{ex}(G(N,p), P_n) \le \mathrm{ex}(G(N,p), C_n) &\le \left(\frac{\mathrm{ex}(K_N, C_n)}{\binom{N}{2}} + \beta\right) e(G(N,p)) \\ &\le \left(\frac{nN^{1+2/n}}{\binom{N}{2}} + \beta\right)\frac{pN^2}{2} \sim pnN^{1+2/n} + \beta\frac{pN^2}{2},\end{aligned}$$

which is weaker than our bounds. (Recall that $p \ge \frac{C}{n}$*, where* $C = C(\beta)$*.) Of course, there are some better upper bounds for* $\mathrm{ex}(K_N, C_n)$*, which could be used to make an improvement. However, since, in general,* $\mathrm{ex}(K_N, C_n)$ *behaves differently with* $\mathrm{ex}(K_N, P_n)$ *and is indeed much greater, Krivelevich, Kronenberg, and Mond's result [80] and ours do not imply one another.*

Remark 5.7. *One can run the same proof and show that Theorem 5.5 holds when* n *is a constant greater than 1 and* N *approaches infinity. Note also that a result of Johansson, Kahn and Vu [66] on the threshold function of the property that* $G(N,p)$ *contains a* K_n*-factor (*n *is a constant) implies* $\mathrm{ex}(G(N,p), P_{n+1}) = \frac{1}{2}(n-1)N$ *for* $p = \Omega\left(N^{-2/n}(\ln n)^{1/\binom{n}{2}}\right)$*, whenever* N *is divisible by* n*. Indeed, they determined the threshold function for containing a* H*-factor (*H *is a fixed graph), which might be useful for further improving the above result.*

5.2 Tools

In this section, we list several results that we will use. The first lemma is a direct application of the depth first search algorithm (DFS), which has appeared in [32]. Using the DFS algorithm in finding long paths was first introduced by Ben-Eliezer, Krivelevich, and Sudakov [18], and then it became a particularly suitable tool in this topic.

Lemma 5.8 ([32])**.** *For every P_{n+1}-free graph H on N vertices, we can find a decomposition of edges into $\bigcup_{i=1}^{N/n} F_i$, where $F_i = E(S_i) \cup E(S_i, T_i)$ for two disjoint sets $S_i, T_i \subseteq [N]$ with $|S_i| = |T_i| = n$.*

We also need the following form of Chernoff's bound.

Lemma 5.9 (Chernoff's Bound)**.** *Let $X = \sum_{i=1}^{n} X_i$, where $X_i = 1$ with probability p_i and $X_i = 0$ with probability $1 - p_i$, and all X_i's are independent. Let $\mu = \mathbb{E}(X) = \sum_{i=1}^{n} p_i$. Then, for all $0 < \delta < 1$,*

$$\mathbb{P}(X \leq (1-\delta)\mu) \leq e^{-\mu\delta^2/2}.$$

The third lemma is a key ingredient of our proof, which is used to find dense subsets in random graphs. This may be of independent interest.

Lemma 5.10. *For $N > 2n$, $0 < p < 1$ and a constant $0 < \alpha \leq 1/2$, let $r = N/n$ and choose an arbitrary β satisfying*

$$\max\left\{2\ln(2e),\ \frac{2}{\alpha np}\ln\left(\frac{1}{\alpha np}\right)\right\} \leq 2\beta\ln\beta \leq \min\left\{2\left(\frac{1}{p}\right)\ln\left(\frac{1}{p}\right),\ \frac{1}{np}\left(\ln r - \ln\alpha 2^{\frac{1}{\alpha}}\right)\right\}. \tag{5.3}$$

Then there exists a positive constant $c = c(\alpha)$ such that with probability at least $1 - \exp(-cr^{\alpha}n)$ there exists an n-set in $G(N, p)$ with at least $\left(\frac{1-\alpha}{2}\right)\beta pn^2$ edges.

Remark 5.11. *Lemma 5.10 essentially states that given N, n, for some range of p, we can find an n-vertex subgraph, which is denser than the random graph by some factor β. For instance, as it will be explained in the proof of Theorem 5.4 (ii), when $135n \leq N \leq ne^{2n}$, we can choose $\frac{\ln r}{nr^{1/5}} \leq p \leq \frac{\ln r}{6n}$, so that $2\beta\ln\beta = \frac{1}{np}\ln\left(\frac{3}{8}r\right)$ satisfying (5.3). Note that if $p \ll \frac{\ln r}{n}$, we have $\beta = \omega(1)$, and therefore the graph we produce here is much denser than the random graph.*

Proof. One can check that the function $f(x) = x\ln x$ is non-negative and increasing for $x \geq 1$. Thus, $\ln(2e) \leq f(\beta) \leq f(1/p)$ implies that

$$\max\left\{2, \frac{1}{\alpha np}\right\} < \beta \leq 1/p. \tag{5.4}$$

Let $B_0 = [N]$. We will construct the desired set iteratively. In each step, take an arbitrary subset $A_i \subseteq B_{i-1}$ of size αn, and let

$$B_i = \{v \in B_{i-1} \setminus A_i :\ \deg(v, A_i) \geq \beta\alpha np\}.$$

We will show that a.a.s. we can continue this process $\lceil \frac{1}{\alpha} \rceil$ steps. For convenience, in the rest of the proof, we ignore all floor and ceiling signs.

Claim 5.12. $|B_i| \geq \frac{rn}{2^i} \exp\left(-2i\beta \ln \beta \cdot \alpha np\right)$, *for all* $0 \leq i \leq \frac{1}{\alpha} - 1$ *with probability at least* $1 - \exp(-\Omega(r^\alpha n))$.

We prove it by induction on $i \geq 0$. For $i = 0$, it is trivial. Suppose the statement holds for $i - 1$. That means

$$|B_{i-1}| \geq \frac{rn}{2^{i-1}} \exp\left(-2(i-1)\beta \ln \beta \cdot \alpha np\right) \tag{5.5}$$

with probability at least $1 - \exp(-\Omega(r^\alpha n))$. Furthermore, $0 \leq i \leq \frac{1}{\alpha} - 1$ yields that $(i-1)\alpha < i\alpha \leq 1 - \alpha < 1$ and hence,

$$|B_{i-1}| \geq \frac{rn}{2^{\frac{1}{\alpha}-2}} \exp\left(-2\beta \ln \beta \cdot np\right) \geq \frac{rn}{2^{\frac{1}{\alpha}-2}} \exp\left(-\left(\ln r - \ln \alpha 2^{\frac{1}{\alpha}}\right)\right) = 4\alpha n,$$

consequently

$$|B_{i-1}| - \alpha n \geq \frac{3}{4}|B_{i-1}| > \frac{|B_{i-1}|}{\sqrt{2}}.$$

Then, the expected size of B_i is

$$\mathbb{E}(|B_i|) = (|B_{i-1}| - \alpha n)\mathbb{P}(\deg(v, A_i) \geq \beta \alpha np) \geq \frac{1}{\sqrt{2}}|B_{i-1}|\binom{\alpha n}{\beta \alpha np} p^{\beta \alpha np}(1-p)^{\alpha n}.$$

Due to (5.4), we get that $p \leq 1/\beta \leq 1/2$ and $\beta \alpha np \geq 1$. Now we use $\binom{\alpha n}{\beta \alpha np} \geq \left(\frac{\alpha n}{\beta \alpha np}\right)^{\beta \alpha np} = \left(\frac{1}{\beta p}\right)^{\beta \alpha np}$ and the inequality $1 - p \geq (2e)^{-p}$, which is valid for $0 \leq p \leq 1/2$. Thus,

$$\begin{aligned}
\mathbb{E}(|B_i|) = (|B_{i-1}| - \alpha n)\mathbb{P}(\deg(v, A_i) \geq \beta \alpha np) &\geq \frac{1}{\sqrt{2}}|B_{i-1}| \exp(-(\beta \ln \beta + \ln 2e)\alpha np) \\
&\geq \frac{1}{\sqrt{2}}|B_{i-1}| \exp(-2\beta \ln \beta \cdot \alpha np).
\end{aligned}$$

Observe that conditioning on (5.5) gives

$$\begin{aligned}
\mathbb{E}(|B_i|) &\geq \frac{1}{\sqrt{2}}|B_{i-1}| \exp(-2\beta \ln \beta \cdot \alpha np) \\
&\geq \frac{1}{\sqrt{2}} \cdot \frac{rn}{2^{i-1}} \exp\left(-2(i-1)\beta \ln \beta \cdot \alpha np\right) \cdot \exp(-2\beta \ln \beta \cdot \alpha np) \\
&= \frac{1}{\sqrt{2}} \cdot \frac{rn}{2^{i-1}} \exp\left(-2i\beta \ln \beta \cdot \alpha np\right) \geq \frac{1}{\sqrt{2}} \cdot \frac{rn}{2^{i-1}} \exp\left(-\alpha i \left(\ln r - \ln \alpha 2^{\frac{1}{\alpha}}\right)\right) \\
&\geq \frac{1}{\sqrt{2}} \cdot \frac{rn}{2^{\frac{1}{\alpha}-1}} \exp\left(-(1-\alpha)\left(\ln r - \ln \alpha 2^{\frac{1}{\alpha}}\right)\right) = \Omega(r^\alpha n),
\end{aligned}$$

which goes to infinity together with n. Therefore, Chernoff's bound (applied with $\delta = 1 - 1/\sqrt{2}$) yields that with probability at least $1 - \exp(-\Omega(r^\alpha n))$ we have

$$|B_i| \geq \frac{1}{\sqrt{2}}\mathbb{E}(|B_i|) \geq \frac{1}{2}|B_{i-1}| \exp(-2\beta \ln \beta \cdot \alpha np) \geq \frac{rn}{2^i} \exp\left(-2i\beta \ln \beta \cdot \alpha np\right),$$

where the last inequality follows from (5.5).

Now we finish the proof of Lemma 5.10. Claim 5.12 gives that with probability at least $1-\exp(-\Omega(r^{\alpha}n))$ the set $B_{\frac{1}{\alpha}-1}$ exists and satisfies

$$\left|B_{\frac{1}{\alpha}-1}\right| \geq \frac{rn}{2^{\frac{1}{\alpha}-1}} \exp\left(-\left(\ln r - \ln \alpha 2^{\frac{1}{\alpha}}\right)\right) = 2\alpha n > \alpha n.$$

Therefore, we can find disjoint sets $A_1, \ldots, A_{1/\alpha}$ of size αn with $e(A_i, A_j) \geq \alpha n \cdot \beta \alpha n p$ for all $1 \leq i < j \leq 1/\alpha$. Let $A = \bigcup_{i=1}^{1/\alpha} A_i$. Then we have $|A| = n$ and

$$e(A) \geq \binom{1/\alpha}{2} \alpha n \cdot \beta \alpha n p = \left(\frac{1-\alpha}{2}\right) \beta p n^2.$$

□

We also present the following two probabilistic results which will be used later.

Lemma 5.13. *Assume that $np \geq \left(\ln \frac{N}{n}\right)/6$ and $N \geq 3n$. Then a.a.s. for every two disjoint sets $S, T \subseteq [N]$, $|S| = |T| = n$, the number of edges in $G \in G(N,p)$ induced by $S \cup T$ with at least one endpoint in S is at most $18n^2p$.*

Proof. Let $X_{S,T}$ be the number of edges in $G(N,p)$ with one endpoint in S and one endpoint in T. Observe that $\mathbb{E}(X_{S,T}) = \left(\frac{3}{2} - \frac{1}{2n}\right) n^2 p$. Note that if $3n^2/2 \leq 18n^2p$, then the statement is trivial. Otherwise, the union bound implies that

$$\begin{aligned}\mathbb{P}(\exists S, T, X_{S,T} \geq 18n^2p) &\leq \binom{N}{n}^2 \binom{3n^2/2}{18n^2p} p^{18n^2p} \leq \left(\frac{Ne}{n}\right)^{2n} \left(\frac{e}{12}\right)^{18n^2p} \\ &= \exp\left(-n\left(18np\ln\left(\frac{12}{e}\right) - 2\ln\left(\frac{Ne}{n}\right)\right)\right).\end{aligned}$$

Since $np \geq \left(\ln \frac{N}{n}\right)/6$ and $N \geq 3n$, we obtain that

$$\begin{aligned}18np\ln\left(\frac{12}{e}\right) - 2\ln\left(\frac{Ne}{n}\right) &\geq 3\ln\left(\frac{12}{e}\right)\ln\left(\frac{N}{n}\right) - 2\ln\left(\frac{Ne}{n}\right) \geq \\ &4\ln\left(\frac{N}{n}\right) - 2\ln\left(\frac{N}{n}\right) - 2 = 2\ln\left(\frac{N}{n}\right) - 2 \geq 2\ln 3 - 2 \geq 0.19.\end{aligned}$$

Finally, we conclude that $\mathbb{P}(\exists S, T, X_{S,T} \geq 18n^2p) \leq \exp(-0.19n) = o(1)$, which completes the proof. □

Lemma 5.14. *Let $\beta = \frac{\frac{1}{np}\ln\frac{N}{n}}{\ln\left(\frac{1}{np}\ln\frac{N}{n}\right)} > 1$ and $m = 8\beta n^2 p$. Then a.a.s. for every two disjoint sets $S, T \subseteq [N]$, $|S| = |T| = n$, the number of edges induced by $S \cup T$ with at least one endpoint in S is at most m.*

Proof. We assume $m < 3n^2/2$ since otherwise Lemma 5.14 holds trivially. By a simple union bound, we obtain

$$\begin{aligned}\mathbb{P}(\exists S, T, X_{S,T} \geq m) &\leq \binom{N}{n}^2 \binom{3n^2/2}{m} p^m \leq \exp\left(2n \ln\left(\frac{Ne}{n}\right)\right) \exp(-\ln\beta \cdot m)\\ &= \exp\left(2n\ln\left(\frac{Ne}{n}\right) - 8\beta\ln\beta \cdot n^2 p\right).\end{aligned}$$

Now we bound from below $\beta \ln \beta$ by

$$\beta\ln\beta = \frac{\frac{1}{np}\ln\frac{N}{n}}{\ln\left(\frac{1}{np}\ln\frac{N}{n}\right)} \ln\left(\frac{\frac{1}{np}\ln\frac{N}{n}}{\ln\left(\frac{1}{np}\ln\frac{N}{n}\right)}\right) \geq \frac{\frac{1}{np}\ln\frac{N}{n}}{\ln\left(\frac{1}{np}\ln\frac{N}{n}\right)} \ln\sqrt{\frac{1}{np}\ln\frac{N}{n}} = \frac{1}{2np}\ln\left(\frac{N}{n}\right).$$

Thus,

$$\begin{aligned}\mathbb{P}(\exists S, T, X_{S,T} \geq m) &\leq \exp\left(2n\ln\left(\frac{Ne}{n}\right) - 8\beta\ln\beta\cdot n^2p\right)\\ &\leq \exp\left(2n\ln\left(\frac{Ne}{n}\right) - \frac{4}{np}\ln\left(\frac{N}{n}\right)\cdot n^2p\right)\\ &\leq \exp\left(-n\left(4\ln\left(\frac{N}{n}\right) - 2\ln\left(\frac{Ne}{n}\right)\right)\right) = o(1),\end{aligned}$$

where the last inequality follows from $N \geq 3n$ as $4\ln\left(\frac{N}{n}\right) - 2\ln\left(\frac{Ne}{n}\right) = 2\ln\left(\frac{N}{n}\right) - 2 \geq 2\ln(3) - 2 \geq 0.19$. $\square$

5.3 Proofs of the main results

5.3.1 Proof of Proposition 5.3

Let $G = (V, E) = G(N, p)$. We will count the number of isolated edges. For a given pair of vertices $e \in \binom{V}{2}$, let X_e be an indicator random variable that takes value 1 if e is an isolated edge in G. Set $X = \sum_e X_e$. Observe that $\Pr(X_e = 1) = p(1-p)^{2(N-2)}$ and so

$$\mathbb{E}(X) = \binom{N}{2} p(1-p)^{2(N-2)} \sim \binom{N}{2} pe^{-2pN} \geq \binom{N}{2} pe^{-2} \to \infty,$$

by assumption. Furthermore, since for vertex disjoint $e, f \in \binom{V}{2}$, $\Pr(X_e = X_f = 1) = p^2(1-p)^{4(n-4)+4}$, we obtain that

$$\mathbb{E}(X^2) = \mathbb{E}(X) + \sum_{e\cap f=\emptyset} \Pr(X_e = X_f = 1) = \mathbb{E}(X) + 6\binom{N}{4} p^2(1-p)^{4(N-4)+4}.$$

Thus,

$$\frac{\mathbb{E}(X^2)}{\mathbb{E}(X)^2} = \frac{1}{\mathbb{E}(X)} + \frac{(N-2)(N-3)}{N(N-1)(1-p)^4} \leq \frac{1}{\mathbb{E}(X)} + \frac{1}{(1-p)^4} \leq \frac{1}{\mathbb{E}(X)} + \frac{1}{1-4p}$$

and

$$\frac{\mathrm{Var}(X)}{\mathbb{E}(X^2)} \leq \frac{1}{\mathbb{E}(X)} + \frac{1}{1-4p} - 1 = \frac{1}{\mathbb{E}(X)} + \frac{4p}{1-4p} = o(1),$$

since $\mathbb{E}(X) \to \infty$ and also by assumption $p \to 0$. Now Chebyshev's inequality yields that X is concentrated around its mean and consequently a.a.s. we have

$$\mathrm{ex}(G(N,p), P_{n+1}) \geq (1+o(1))\mathbb{E}(X) = \Omega(pN^2).$$

The upper bound easily follows from the fact that $\mathrm{ex}(G(N,p), P_{n+1}) \leq e(G(N,p))$.

Finally observe that a.a.s.

$$\mathrm{ex}(G(N,1/N), P_{n+1}) \geq (1+o(1))\mathbb{E}(X) \geq (1+o(1))\binom{N}{2}\frac{1}{N}e^{-2} \geq N/15.$$

□

5.3.2 Proof of Theorem 5.4

Proof of Theorem 5.4 (i). This proof is by now quite standard which applies the DFS algorithm and the first moment method. Recall that $np \geq \left(\ln \frac{N}{n}\right)/6$ and $N \geq 3n$.

Observe that Lemma 5.13 together with Lemma 5.8 imply that for every P_{n+1}-free subgraph H of $G \in G(N,p)$ a.a.s.

$$e(H) \leq \frac{N}{n} \cdot 18n^2p = 18pnN,$$

which establishes the upper bound.

For the lower bound, take an arbitrary vertex partition $[N] = \bigcup_{i=1}^{N/n} S_i$, where $|S_i| = n$ for all i. Let H be the subgraph of $G \in G(N,p)$ whose edge set is $\bigcup E(G[S_i])$. Clearly, H is P_{n+1}-free. Note that $\mathbb{E}(e(H)) = \frac{N}{n}\left(\frac{1}{2} - \frac{1}{2n}\right)n^2p = \left(\frac{1}{2} - \frac{1}{2n}\right)pnN$. By Chernoff's bound,

$$\mathbb{P}\left(e(H) \leq \frac{1}{4}pnN\right) \leq \exp\left(-\Omega(pnN)\right) = o(1),$$

since $pnN \to \infty$. Therefore, a.a.s. we have $\mathrm{ex}(G(N,p), P_{n+1}) \geq e(H) \geq \frac{1}{4}pnN$. □

Proof of Theorem 5.4 (ii). We first show the upper bound. Let $\beta_1 = \frac{\frac{1}{np}\ln\frac{N}{n}}{\ln\left(\frac{1}{np}\ln\frac{N}{n}\right)}$ and $m = 8\beta_1 n^2 p$. Since $np \le \left(\ln\frac{N}{n}\right)/6$, we know that $\beta_1 > 1$.

For every P_{n+1}-free subgraph H of $G \in G(N,p)$, Lemma 5.8 and Lemma 5.14 imply that a.a.s

$$e(H) \le \frac{N}{n}\cdot m = 8\beta_1 pnN = 8\frac{\frac{1}{np}\ln\frac{N}{n}}{\ln\left(\frac{1}{np}\ln\frac{N}{n}\right)}pnN,$$

which establishes the upper bound.

For the lower bound, we shall divide the discussion into three cases. First, let us assume $N \le 135n$. Together with $\frac{1}{np}\ln\left(\frac{N}{n}\right) \ge 6 \ge e$, we have

$$\frac{\omega}{\ln\omega}pnN = \frac{\ln\left(\frac{N}{n}\right)}{\ln\left(\frac{1}{np}\ln\left(\frac{N}{n}\right)\right)}N \le \ln\left(\frac{N}{n}\right)N < 5N.$$

Therefore, by (5.2), we trivially have

$$\mathrm{ex}(G(N,p),P_{n+1}) \ge N/15 \ge \frac{1}{75}\frac{\omega}{\ln\omega}pnN.$$

Next, let us assume $p \le \ln\left(\frac{N}{n}\right)/\left(n\left(\frac{N}{n}\right)^{1/5}\right)$. Similarly, we complete the proof by observing that

$$\frac{\omega}{\ln\omega}pnN = \frac{\ln\left(\frac{N}{n}\right)}{\ln\left(\frac{1}{np}\ln\left(\frac{N}{n}\right)\right)}N \le \frac{\ln\left(\frac{N}{n}\right)}{\frac{1}{5}\ln\left(\frac{N}{n}\right)}N = 5N.$$

It remains to prove the lower bound for the case when $N \ge 135n$ and

$$\frac{\ln\left(\frac{N}{n}\right)}{n\left(\frac{N}{n}\right)^{1/5}} \le p \le \frac{\ln(\frac{N}{n})}{6n}. \tag{5.6}$$

Indeed, such range of p only exists for $N \ge 6^5 n$. In this case, we will apply Lemma 5.10 repeatedly to find a dense subgraph with no P_{n+1}. Let

$$2\beta_2\ln\beta_2 = \min\left\{2\left(\frac{1}{p}\right)\ln\left(\frac{1}{p}\right),\ \frac{1}{np}\ln\left(\frac{3N}{8n}\right)\right\}.$$

Since $N \le ne^{2n}$ and $p \le \ln\left(\frac{N}{n}\right)/(6n) \le \frac{1}{3}$, we have

$$2\left(\frac{1}{p}\right)\ln\left(\frac{1}{p}\right) \ge 2\left(\frac{1}{p}\right)\ln 3 > \frac{2}{p} \ge \frac{1}{np}\ln\left(\frac{3N}{8n}\right).$$

Furthermore, since $N \geq 6^5 n$, we obtain

$$\ln\left(\frac{3N}{8n}\right) \geq \ln\left(\frac{3}{8}\right) + \frac{1}{5}\ln 6^5 + \frac{4}{5}\ln\left(\frac{N}{n}\right) > \frac{4}{5}\ln\left(\frac{N}{n}\right),$$

and

$$2\beta_2 \ln \beta_2 = \frac{1}{np}\ln\left(\frac{3N}{8n}\right) \geq \frac{4}{5np}\ln\left(\frac{N}{n}\right) > 2\ln(2e).$$

Finally, observe that for $\alpha = 1/2$,

$$\frac{1}{np}\ln\left(\frac{3N}{8n}\right) \geq \frac{1}{np}\cdot 4\ln\left(\frac{2\left(\frac{N}{n}\right)^{1/5}}{\ln\left(\frac{N}{n}\right)}\right) \geq \frac{2}{\alpha np}\ln\left(\frac{1}{\alpha np}\right),$$

where the first inequality is given by $N \geq 135n$ and the last inequality follows from (5.6). Thus, we can iteratively apply Lemma 5.10 $N/4n$ times with $\alpha = \frac{1}{2}$ and $r = \frac{3N}{4n}$ and find $N/4n$ disjoint n-sets A_i, where a.a.s. for all i

$$e(A_i) \geq \left(\frac{1-\alpha}{2}\right)\beta_2 pn^2 \geq \frac{1-\alpha}{4}\frac{\frac{1}{np}\ln\left(\frac{3N}{8n}\right)}{\ln\left(\frac{1}{np}\ln\left(\frac{3N}{8n}\right)\right)}pn^2 \geq \frac{1}{10}\frac{\frac{1}{np}\ln\left(\frac{N}{n}\right)}{\ln\left(\frac{1}{np}\ln\left(\frac{N}{n}\right)\right)}pn^2.$$

Let H be the subgraph of G with vertex set $\bigcup_{i=1}^{N/4n} A_i$, and edge set $\bigcup_{i=1}^{N/4n} E(A_i)$. Note that H is P_{n+1}-free and therefore, a.a.s. we have

$$\mathrm{ex}(G(N,p), P_{n+1}) \geq e(H) \geq \frac{1}{10}\frac{\frac{1}{np}\ln\left(\frac{N}{n}\right)}{\ln\left(\frac{1}{np}\ln\left(\frac{N}{n}\right)\right)}pn^2 \cdot \frac{N}{4n} = \frac{1}{40}\frac{\frac{1}{np}\ln\left(\frac{N}{n}\right)}{\ln\left(\frac{1}{np}\ln\left(\frac{N}{n}\right)\right)}pnN.$$

□

5.3.3 Proof of Theorem 5.5

Proof of Theorem 5.5 (i). By the the Erdős-Gallai Theorem (Theorem 5.1), it is sufficient to prove the lower bound. Let

$$2\beta\ln\beta = \min\left\{2\left(\frac{1}{p}\right)\ln\left(\frac{1}{p}\right),\ \frac{4}{5np}\ln N\right\}.$$

Since $p \geq N^{-\frac{2}{5n}}$, we have $\beta = 1/p$. If $p > 1/3$, then the proof simply follows from the proof of Theorem 5.4 (i). Otherwise, we have $2\beta\ln\beta \geq 6\ln 3 > 2\ln(2e)$. Similarly as in the proof of Theorem 5.4 (ii), we can iteratively apply Lemma 5.10 $N/4n$ times with $\alpha = \frac{1}{2}$ and $r = \frac{3N}{4n}$, and a.a.s. find a P_{n+1}-free subgraph H of $G(N,p)$

with

$$e(H) \geq \left(\frac{1-\alpha}{2}\right)\beta pn^2 \cdot \frac{N}{4n} = \frac{1}{16}nN.$$

□

Proof of Theorem 5.5 (ii). The proof of the upper bound is the same as in Theorem 5.4 (ii) and we skip here the full details. For the lower bound, we first assume that $p < N^{-1/5}$. Observe that

$$\frac{\omega}{\ln \omega} pnN = \frac{\ln N}{\ln\left(\frac{1}{np}\ln N\right)} N \leq \frac{\ln N}{\ln N^{1/5}} N = 5N,$$

where the inequality holds for $N \geq ne^{2n}$. Therefore, by (5.2), we trivially have

$$\mathrm{ex}(G(N,p), P_{n+1}) \geq N/15 \geq \frac{1}{75}\frac{\omega}{\ln \omega} pnN.$$

It remains to show the lower bound for $p \geq N^{-1/5}$. Let

$$2\beta \ln \beta = \min\left\{2\left(\frac{1}{p}\right)\ln\left(\frac{1}{p}\right),\ \frac{4}{5np}\ln N\right\}.$$

Since $p \leq N^{-\frac{2}{5n}}$, we have $2\beta\ln\beta = \frac{4}{5np}\ln N$. Since $N \geq ne^{2n}$, we have

$$\frac{1}{np}\ln\left(\frac{3N}{8n}\right) \geq 2\beta\ln\beta \geq \frac{4}{5np}\ln\left(ne^{2n}\right) \geq \frac{8}{5p} \geq \frac{8}{5}N^{\frac{2}{5n}} \geq \frac{8e^{\frac{4}{5}}}{5} > 2\ln(2e).$$

Moreover, observe that for $\alpha = \frac{1}{2}$ and $p \geq N^{-1/5}$, we have $2\beta\ln\beta \geq \frac{2}{\alpha np}\ln\left(\frac{1}{\alpha np}\right)$. Similarly as in the proof of Theorem 5.4 (ii), the proof is completed by iteratively applying Lemma 5.10 $N/4n$ times with $\alpha = \frac{1}{2}$ and $r = \frac{3N}{4n}$. □

5.4 Long paths and multicolor size-Ramsey number

The size-Ramsey number $\hat{R}(F, r)$ of a graph F is the smallest integer m such that there exists a graph G on m edges with the property that any r-coloring of the edges of G yields a monochromatic copy of F. The study of size-Ramsey number was initiated by Erdős, Faudree, Rousseau and Schelp [38]. For paths, Beck [16], resolving a \$100 question of Erdős, proved that $\hat{R}(P_n, 2) < 900n$ for sufficiently large n. The strongest upper bound, $\hat{R}(P_n, 2) \leq 74n$, was given by Dudek and Prałat [31], and they also provide the lower bound, $\hat{R}(P_n, 2) \geq 5n/2 - O(1)$. Very recently, Bal and DeBiasio [4] further improved the lower bound to $(3.75 - o(1))n$.

For more colors, it was proved in [31] that $\frac{(r+3)r}{4}n - O(r^2) \le \hat{R}(P_n, r) \le 33r4^r n$. Subsequently, Krivelevich [79] (see also [78]) showed that $\hat{R}(P_n, r) = O((\ln r)r^2 n)$. An alternative proof of the above result was later given by Dudek and Prałat [32]. Both proofs indeed give a stronger *density-type* result, which shows that any dense subset of a large enough structure contain the desired substructure. In particular, the proof in [32] implies the following result.

Theorem 5.15 ([32]). *For $r \ge 2$ and $c \ge 7$, there exists a constant $\alpha = \alpha(c)$ such that the following statement holds a.a.s. for $p \ge \alpha(\ln r)/n$. Every subgraph H of $G \in G(crn, p)$ with $e(H) \ge e(G)/r$ contains a P_{n+1}.*

Note that any improvement of the order of magnitude of p in the above theorem would improve the upper bound for $\hat{R}(P_n, r)$. However, Theorem 5.4 (ii) implies that when $p \ll (\ln cr)/(6n)$, i.e. $(\ln cr)/np \gg 6$, a.a.s. there exists a P_{n+1}-free subgraph of $G \in G(crn, p)$ which contains more than

$$\frac{1}{40}\frac{(\ln cr)/np}{\ln((\ln cr)/np)} pn \cdot crn \ge cpn \cdot crn > e(G)/r$$

edges. Therefore, $(\ln r)/n$ is the threshold function for the density statement in Theorem 5.15. It would be interesting to know if $(\ln r)/n$ is still the threshold function for the corresponding Ramsey-type statement.

5.5 Concluding remarks

Our investigation raises some open problems. The most interesting question is to investigate the corresponding Ramsey properties on random graphs. The Ramsey-type questions on sparse random graphs has been studied by several researchers, for example, see [20, 97].

Problem 5.16. *Determine the threshold function $p(n)$ for the following statement. For some constant c and $r \ge 2$ (c is independent of r), every r-coloring of $G(crn, p)$ contains a monochromatic P_{n+1}.*

Theorem 5.15 implies that $p(n) = O((\ln r)/n)$, while the lower bound of $\hat{R}(P_n, r)$ shows that $p(n) = \Omega(1/n)$, where n goes to infinity. The exact behavior of $p(n)$ remains open and its determination would be very useful for studying the size-Ramsey number of paths.

Another direction is to consider the following graph parameter. Denote by $c(G, F)$ the minimum number of colors k such that there exists a k-coloring of G without monochromatic F. Clearly, we have

$$c(G(N, p), P_{n+1}) \ge \frac{\binom{N}{2}p}{\mathrm{ex}(G(N, p), P_{n+1})} \ge \frac{pN^2}{3\mathrm{ex}(G(N, p), P_{n+1})}. \tag{5.7}$$

Let $r = N/n$. We first present two general upper bounds on $c(G(N,p), P_{n+1})$.

Theorem 5.17. *Suppose r is a prime power, then $c(G(N,p), P_{n+1}) \leq r+1$.*

Proof. We use a construction from [50] (also appeared in [79]). Let A_r be an affine plane of order r, i.e. r^2 points with $r^2 + r$ lines, where every pair of points is contained in a unique line, and the lines can be split into $r+1$ disjoint families $F_1, \ldots, F_{r+1}$ so that the lines inside the families are parallel.

We arbitrarily partition $[N]$ into r^2 parts $V_1, V_2, \ldots, V_{r^2}$, where each part has size $N/r^2 = n/r$. We define an $r+1$-coloring as follows. If e is an edge crossing between V_x and V_y, where the unique line containing xy is in the family F_i, then we color e by i. Observe that every connected subgraph in color i has its vertex set V inside $\cup_{x \in L} V_x$ for some line $L \in A_r$. Therefore, we have $|V| \leq r \cdot n/r = n$, and there is no monochromatic P_{n+1}. $\square$

Theorem 5.18. *A.a.s. $c(G(N,p), P_{n+1}) \leq 2pN$.*

Proof. Let $k = 2pN$, and we can assume $k \leq r+1$. Consider a random k-coloring of $G(N,p)$. Then the subgraph G_i, whose edges are all edges in color i, is in $G(N,p')$, where $p' = p/k = 1/2N$. A fundamental result of Erdős and Rényi shows that a.a.s the largest component of G_i has size $O(\ln N) \leq n$. Therefore, a.a.s. there is no monochromatic P_{n+1}. $\square$

Corollary 5.19. *If $p = \frac{1}{\omega \cdot n}$, where $\omega = \omega(r) \geq 2$, then a.a.s. $c(G(N,p), P_{n+1}) \leq 2r/\omega$.*

For the lower bound, the proof of Theorem 1.2. in [32] implies the following.

Theorem 5.20. *For $p \geq 22(\ln(r/7))/n$, a.a.s. $c(G(N,p), P_{n+1}) > r/7$.*

This together with Theorem 5.17 shows that a.a.s. $c(G(N,p), P_{n+1}) = \Theta(r)$ for $p = \Omega((\ln r)/n)$. On the other hand, Theorem 5.4 and (5.7) give a lower bound for small p.

Theorem 5.21. *For $p \leq (\ln r)/34n$, a.a.s. $c(G(N,p), P_{n+1}) \geq \frac{\ln \omega}{24\omega} r$, where $\omega = (\ln r)/np$.*

This naturally raises the following question.

Problem 5.22. *What is the exact behavior of $c(G(N,p), P_{n+1})$ for $p = o((\ln r)/n)$, where n goes to infinity?*

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi. The longest path in a random graph. *Combinatorica*, 1(1):1–12, 1981.

[2] N. Alon, J. Balogh, P. Keevash, and B. Sudakov. The number of edge colorings with no monochromatic cliques. *Journal of the London Mathematical Society*, 70(2):273–288, 2004.

[3] L. Babai, M. Simonovits, and J. Spencer. Extremal subgraphs of random graphs. *Journal of Graph Theory*, 14(5):599–622, 1990.

[4] D. Bal and L. DeBiasio. New lower bounds on the size-Ramsey number of a path. *arXiv preprint arXiv:1909.06354*, 2019.

[5] J. Balogh. A remark on the number of edge colorings of graphs. *European Journal of Combinatorics*, 27(4):565–573, 2006.

[6] J. Balogh, B. Bollobás, and M. Simonovits. The number of graphs without forbidden subgraphs. *Journal of Combinatorial Theory, Series B*, 91(1):1–24, 2004.

[7] J. Balogh, N. Bushaw, M. Collares, H. Liu, R. Morris, and M. Sharifzadeh. The typical structure of graphs with no large cliques. *Combinatorica*, 37(4):617–632, 2017.

[8] J. Balogh, A. Dudek, and L. Li. An analogue of the Erdős-Gallai theorem for random graphs. *arXiv preprint arXiv:1909.00214*, 2019.

[9] J. Balogh and L. Li. On the number of generalized Sidon sets. *arXiv preprint arXiv:1803.00659*, 2018.

[10] J. Balogh and L. Li. On the number of linear hypergraphs of large girth. *Journal of Graph Theory*, 93(1):113–141, 2020.

[11] J. Balogh, R. Morris, and W. Samotij. Independent sets in hypergraphs. *Journal of the American Mathematical Society*, 28(3):669–709, 2015.

[12] J. Balogh, R. Morris, and W. Samotij. The method of hypergraph containers. *Proceedings of the International Congress of MathematiciansRio de Janeiro*, 3:3045–3078, 2018.

[13] J. Balogh, B. Narayanan, and J. Skokan. The number of hypergraphs without linear cycles. *Journal of Combinatorial Theory, Series B*, 2018.

[14] J. Balogh and J. Solymosi. On the number of points in general position in the plane. *Discrete Analysis*, Paper No. 16:20, 2018.

[15] J. Balogh and A. Z. Wagner. Further applications of the container method. In *Recent Trends in Combinatorics*, pages 191–213. Springer, 2016.

[16] J. Beck. On size-Ramsey number of paths, trees, and circuits. I. *Journal of Graph Theory*, 7(1):115–129, 1983.

[17] F. A. Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proceedings of the National Academy of Sciences of the United States of America*, 32(12):331, 1946.

[18] I. Ben-Eliezer, M. Krivelevich, and B. Sudakov. The size-Ramsey number of a directed path. *Journal of Combinatorial Theory, Series B*, 102(3):743–755, 2012.

[19] F. S. Benevides, C. Hoppen, and R. M. Sampaio. Edge-colorings of graphs avoiding complete graphs with a prescribed coloring. *Discrete Mathematics*, 340(9):2143–2160, 2017.

[20] T. Bohman, A. Frieze, M. Krivelevich, P.-S. Loh, and B. Sudakov. Ramsey games with giants. *Random Structures & Algorithms*, 38(1-2):1–32, 2011.

[21] B. Bollobás. The evolution of sparse graphs, in graph theory and combinatorics proceedings. In *Cambridge Combinatorial Conference in Honour of Paul Erdős*, pages 335–357, 1984.

[22] B. Bollobás and V. Nikiforov. Books in graphs. *European Journal of Combinatorics*, 26(2):259–270, 2005.

[23] J. A. Bondy and M. Simonovits. Cycles of even length in graphs. *Journal of Combinatorial Theory, Series B*, 16(2):97–105, 1974.

[24] K. Cameron, J. Edmonds, and L. Lovász. A note on perfect graphs. *Periodica Mathematica Hungarica*, 17(3):173–175, 1986.

[25] P. Cameron and P. Erdős. On the number of sets of integers with various properties. *Number Theory (RA Mollin, ed.)*, pages 61–79, 1990.

[26] S. Chowla. Solution of a problem of Erdős and Turán in additive-number theory. *Proc. Nat. Acad. Sci. India. Sect. A*, 14(1-2):5–4, 1944.

[27] C. Collier-Cartaino, N. Graber, and T. Jiang. Linear Turán numbers of r-uniform linear cycles and related Ramsey numbers. *Combinatorics, Probability, and Computing*, 27:358–386, 2018.

[28] D. Conlon and W. T. Gowers. Combinatorial theorems in sparse random sets. *Annals of Mathematics*, pages 367–454, 2016.

[29] J. de Oliveira Bastos, F. S. Benevides, G. O. Mota, and I. Sau. Counting Gallai 3-colorings of complete graphs. *Discrete Mathematics*, 342(9):2618–2631, 2019.

[30] D. Dellamonica Jr, Y. Kohayakawa, M. Marciniszyn, and A. Steger. On the resilience of long cycles in random graphs. *Electronic Journal of Combinatorics*, 15(1):32, 2008.

[31] A. Dudek and P. Prałat. On some multicolor Ramsey properties of random graphs. *SIAM Journal on Discrete Mathematics*, 31(3):2079–2092, 2017.

[32] A. Dudek and P. Prałat. Note on the multicolour size-Ramsey number for paths. *Electronic Journal of Combinatorics*, 25(3):3–35, 2018.

[33] P. Erdős. On a problem of Sidon in additive number theory and on some related problems addendum. *Journal of the London Mathematical Society*, 19(76_Part_4):208–208, 1944.

[34] P. Erdős. *Some new applications of probability methods to combinatorial analysis and graph theory.* University of Calgary, Department of Mathematics, Statistics and Computing , 1974.

[35] P. Erdős, D. J. Kleitman, and B. L. Rothschild. Asymptotic enumeration of k_n-free graphs, in: Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II, 19–27. Atti dei Convegni Lincei, No. 17. *Accad. Naz. Lincei*, pages 19–27, 1976.

[36] P. Erdős and M. Simonovits. Cube-supersaturated graphs and related problems. *Progress in graph theory (Waterloo, Ont., 1982)*, pages 203–218, 1984.

[37] P. Erdős and P. Turán. On a problem of Sidon in additive number theory, and on some related problems. *Journal of the London Mathematical Society*, 1(4):212–215, 1941.

[38] P. Erdős, R. J. Faudree, C. C. Rousseau, and R. H. Schelp. The size-Ramsey number. *Periodica Mathematica Hungarica*, 9(1-2):145–161, 1978.

[39] P. Erdös, P. Frankl, and V. Rödl. The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs and Combinatorics*, 2(1):113–121, 1986.

[40] P. Erdös, P. Frankl, and V. Rödl. The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs and Combinatorics*, 2(1):113–121, 1986.

[41] P. Erdős and T. Gallai. On maximal paths and circuits of graphs. *Acta Mathematica Hungarica*, 10(3-4):337–356, 1959.

[42] V. Falgas-Ravry, K. O'Connell, and A. Uzzell. Multicolour containers, extremal entropy and counting. *Random Structures & Algorithms*, 2018.

[43] R. Faudree and M. Simonovits. Cycle-supersaturated graphs. Unpublished manuscript.

[44] J. Fox, A. Grinshpun, and J. Pach. The Erdős–Hajnal conjecture for rainbow triangles. *Journal of Combinatorial Theory, Series B*, 111:75–125, 2015.

[45] P. Frankl and V. Rödl. Large triangle-free subgraphs in graphs without K_4. *Graphs and Combinatorics*, 2(1):135–144, 1986.

[46] A. M. Frieze. On large matchings and cycles in sparse random graphs. *Discrete Mathematics*, 59(3):243–256, 1986.

[47] Z. Füredi and T. Jiang. Hypergraph Turán numbers of linear cycles. *Journal of Combinatorial Theory, Series A*, 123(1):252–270, 2014.

[48] Z. Füredi and M. Simonovits. The history of degenerate (bipartite) extremal graph problems. In *Erdős Centennial*, pages 169–264. Springer, 2013.

[49] T. Gallai. Transitiv orientierbare graphen. *Acta Mathematica Hungarica*, 18(1-2):25–66, 1967.

[50] A. Gyárfás. Partition coverings and blocking sets in hypergraphs (in Hungarian). *Communications of the Computer and Automation Institute of the Hungarian Academy of Sciences*, 71:62, 1977.

[51] A. Gyárfás and G. N. Sárközy. Gallai colorings of non-complete graphs. *Discrete Mathematics*, 310(5):977–980, 2010.

[52] A. Gyárfás, G. N. Sárközy, A. Sebő, and S. Selkow. Ramsey-type results for Gallai colorings. *Journal of Graph Theory*, 64(3):233–243, 2010.

[53] A. Gyárfás and G. Simony. Edge colorings of complete graphs without tricolored triangles. *Journal of Graph Theory*, 46(3):211–216, 2004.

[54] J. Han and Y. Kohayakawa. On hypergraphs without loose cycles. *Discrete Mathematics*, 341(4):946–949, 2018.

[55] P. E. Haxell, Y. Kohayakawa, and T. Łuczak. Turán's extremal problem in random graphs: Forbidding even cycles. *Journal of Combinatorial Theory, Series B*, 64(2):273–287, 1995.

[56] P. E. Haxell, Y. Kohayakawa, and T. Łuczak. Turán's extremal problem in random graphs: Forbidding odd cycles. *Combinatorica*, 16(1):107–122, 1996.

[57] C. Hoppen, Y. Kohayakawa, and H. Lefmann. Edge colourings of graphs avoiding monochromatic matchings of a given size. *Combinatorics, Probability and Computing*, 21(1-2):203–218, 2012.

[58] C. Hoppen, Y. Kohayakawa, and H. Lefmann. Edge-colorings of graphs avoiding fixed monochromatic subgraphs with linear turán number. *European Journal of Combinatorics*, 35:354–373, 2014.

[59] C. Hoppen, Y. Kohayakawa, and H. Lefmann. Edge-colorings of uniform hypergraphs avoiding monochromatic matchings. *Discrete Mathematics*, 338(2):262–271, 2015.

[60] C. Hoppen and H. Lefmann. Edge-colorings avoiding a fixed matching with a prescribed color pattern. *European Journal of Combinatorics*, 47:75–94, 2015.

[61] C. Hoppen, H. Lefmann, and K. Odermann. A rainbow Erdős-Rothschild problem. *Electronic Notes in Discrete Mathematics*, 49:473–480, 2015.

[62] C. Hoppen, H. Lefmann, and K. Odermann. On graphs with a large number of edge-colorings avoiding a rainbow triangle. *European Journal of Combinatorics*, 66:168–190, 2017.

[63] C. Hoppen, H. Lefmann, and K. Odermann. A rainbow Erdős-Rothschild problem. *SIAM Journal on Discrete Mathematics*, 31(4):2647–2674, 2017.

[64] S. Janson. Large deviations for sums of partly dependent random variables. *Random Structures & Algorithms*, 24(3):234–248, 2004.

[65] T. Jiang and L. Yepremyan. Supersaturation of even linear cycles in linear hypergraphs. *arXiv:1707.03091*, 2017.

[66] A. Johansson, J. Kahn, and V. Vu. Factors in random graphs. *Random Structures & Algorithms*, 33(1):1–28, 2008.

[67] D. J. Kleitman and D. Wilson. On the number of graphs which lack small cycles. 1996. Unpublished manuscript.

[68] D. J. Kleitman and K. J. Winston. On the number of graphs without 4-cycles. *Discrete Mathematics*, 41(2):167–172, 1982.

[69] Y. Kohayakawa, B. Kreuter, and A. Steger. An extremal problem for random graphs and the number of graphs with large even-girth. *Combinatorica*, 18(1):101–120, 1998.

[70] Y. Kohayakawa, S. J. 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 & Algorithms*, 46(1):1–25, 2015.

[71] Y. Kohayakawa, T. Łuczak, and V. Rödl. On K_4-free subgraphs of random graphs. *Combinatorica*, 17(2):173–213, 1997.

[72] Y. Kohayakawa, V. Rödl, and M. Schacht. The Turán theorem for random graphs. *Combinatorics, Probability and Computing*, 13(1):61–91, 2004.

[73] J. Komlós and E. Szemerédi. Limit distribution for the existence of Hamiltonian cycles in a random graph. *Discrete Mathematics*, 43(1):55–63, 1983.

[74] J. Körner and G. Simonyi. Graph pairs and their entropies: modularity problems. *Combinatorica*, 20(2):227–240, 2000.

[75] A. Kostochka, D. Mubayi, and J. Verstraëte. Personal Communication.

[76] A. Kostochka, D. Mubayi, and J. Verstraëte. Turán problems and shadows I: paths and cycles. *Journal of Combinatorial Theory, Series A*, 129:57–79, 2015.

[77] B. Kreuter. *Extremale und Asymptotische Graphentheorie für verbotene bipartite Untergraphen*. PhD thesis, Diplomarbeit, Forschungsinstitut für Diskrete Mathematik, Universität Bonn, 1994.

[78] M. Krivelevich. Expanders—how to find them, and what to find in them. *Surveys in Combinatorics*, 2019.

[79] M. Krivelevich. Long cycles in locally expanding graphs, with applications. *Combinatorica*, 39(1):135–151, 2019.

[80] M. Krivelevich, G. Kronenberg, and A. Mond. Turán-type problems for long cycles in random and pseudo-random graphs. *arXiv preprint arXiv:1911.08539*, 2019.

[81] F. Lazebnik and J. Verstraëte. On hypergraphs of girth five. *Electronic journal of combinatorics*, 10(1):25, 2003.

[82] H. Lefmann and Y. Person. Exact results on the number of restricted edge colorings for some families of linear hypergraphs. *Journal of Graph Theory*, 73(1):1–31, 2013.

[83] H. Lefmann, Y. Person, V. Rödl, and M. Schacht. On colourings of hypergraphs without monochromatic fano planes. *Combinatorics, Probability and Computing*, 18(5):803–818, 2009.

[84] H. Lefmann, Y. Person, and M. Schacht. A structural result for hypergraphs with many restricted edge colorings. *Journal of Combinatorics*, 1(4):441–475, 2010.

[85] L. Lovász. *Combinatorial problems and exercises*, volume 361. American Mathematical Soc., 2007.

[86] R. Morris and D. Saxton. The number of $C_{2\ell}$-free graphs. *Advances in Mathematics*, 298:534–580, 2016.

[87] D. Mubayi and L. Wang. The number of triple systems without even cycles. *Combinatorica*, 39(3):679–704, 2019.

[88] B. Nagle, V. Rödl, and M. Schacht. Extremal hypergraph problems and the regularity method. In *Topics in discrete mathematics*, pages 247–278. Springer, 2006.

[89] C. Palmer, M. Tait, C. Timmons, and A. Z. Wagner. Turán numbers for Berge-hypergraphs and related extremal problems. *Discrete Mathematics*, 342(6):1553–1563, 2019.

[90] O. Pikhurko. A note on the Turán function of even cycles. *Proceedings of the American Mathematical Society*, 140(11):3687–3692, 2012.

[91] O. Pikhurko and Z. B. Yilma. The maximum number of k_3-free and k_4-free edge 4-colorings. *Journal of the London Mathematical Society*, 85(3):593–615, 2012.

[92] I. Z. Ruzsa and E. Szemerédi. Triple systems with no six points carrying three triangles. *Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai*, 18:939–945, 1978.

[93] W. Samotij. Counting independent sets in graphs. *European Journal of Combinatorics*, 48:5–18, 2015.

[94] D. Saxton and A. Thomason. Hypergraph containers. *Inventiones mathematicae*, 201(3):925–992, 2015.

[95] M. Schacht. Extremal results for random discrete structures. *Annals of Mathematics*, pages 333–365, 2016.

[96] J. Singer. A theorem in finite projective geometry and some applications to number theory. *Transactions of the American Mathematical Society*, 43(3):377–385, 1938.

[97] R. Spöhel, A. Steger, and H. Thomas. Coloring the edges of a random graph without a monochromatic giant component. *the electronic journal of combinatorics*, 17(1):133, 2010.

[98] T. Szabó and V. H. Vu. Turán's theorem in sparse random graphs. *Random Structures & Algorithms*, 23(3):225–234, 2003.

[99] A. Z. Wagner. Large subgraphs in rainbow-triangle free colorings. *Journal of Graph Theory*, 86(2):141–148, 2017.

[100] R. Yuster. The number of edge colorings with no monochromatic triangle. *Journal of Graph Theory*, 21(4):441–452, 1996.