Extremal Problems in Graphs and Hypergraphs

by

Beka Ergemlidze

Supervisor: **Ervin Győri**

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Mathematics and its Applications
Central European University
Budapest, Hungary

Contents

Abstract

The main theme of the thesis is the investigation of Turán-type problems in graphs and hypergraphs. A big part of it is focused on studying Turán numbers of Berge and linear cycles in hypergraphs. In addition, we investigate behaviour of linear cycles in 3-uniform hypergraphs and observe some Turán-type problems for graphs.

The thesis is divided into 6 chapters. The first chapter contains the background on Turán type problems in graphs and hypergraphs, as well as describes characteristics of linear cycles in 3-uniform hypergraphs.

In the second chapter we study behaviour of linear cycles in 3-uniform hypergraphs. Gyárfás, Győri and Simonovits proved that if a 3-uniform hypergraph H has no linear cycles, then $\alpha(H) \geq \frac{2|V(H)|}{5}$. The hypergraph consisting of vertex disjoint copies of complete hypergraphs K_5^3 shows that equality can hold. They asked whether $\alpha(H)$ can be improved if we exclude K_5^3 as a subhypergraph and whether such a hypergraph is 2-colorable. We answer these questions affirmatively by showing that if a 3-uniform linear cycle-free hypergraph H, contains no subhypergraph K_5^3, then it is 2-colorable. Therefore, $\alpha(H) \geq \lceil \frac{V(H)}{2} \rceil$. Furthermore, we show that this bound is sharp. We also determine the exact upper-bound on minimum degree in linear cycle-free hypergraphs. These results are based on the paper "3-uniform hypergraphs and linear cycles" co-authored with Győri and Methuku.

Gyárfás and Sárközy conjectured that the following extension of the well-known theorem of Pósa holds: One can partition every k-uniform hypergraph H into at most $\alpha(H)$ linear cycles (here, as in Pósa's theorem, vertices and subsets of hyperedges are accepted as linear cycles). We show that their conjecture would be true for $k = 3$, if we allowed the linear cycles to be just edge-disjoint, instead of being vertex-disjoint, thus proving a weaker version of the conjecture. The proof is based on the paper "A note on the Linear Cycle Cover Conjecture of Gyárfás and Sárközy" co-authored with Győri and Methuku.

In Chapter 3 we investigate hypergraph Turán problems of Berge cycles and linear cycles. Given a family of 3-uniform hypergraphs $\mathcal{F}$, the *linear Turán number* of $\mathcal{F}$, denoted $\mathrm{ex}_3^{lin}(n, \mathcal{F})$, is the maximum number of hyperedges in an $\mathcal{F}$-free 3-uniform linear hypergraph on n vertices. $\mathrm{ex}_3(n, \mathcal{F})$ denotes the Turán number of $\mathcal{F}$ for 3-uniform hypergraphs. We give an upper bound of $\mathrm{ex}_3(n, C_5)$, which significantly improves the previous bound determined by Győri and Bollobás. In the linear case, we determine asymptotically sharp bounds and show that $\mathrm{ex}_3^{lin}(n, C_5) = \frac{1}{3\sqrt{3}} n^{3/2}$ asymptotically, by giving a new construction and proving the corresponding upper bound. We also show that asymptotics of $\mathrm{ex}_3^{lin}(n, C_4)$ is same as $\mathrm{ex}_3^{lin}(n, \{C_3, C_4\})$, strengthening the theorem of Lazebnik and Verstraëte. In the same chapter we provide constructions of 3-uniform hypergraphs without linear cycle of given odd length, which in special cases, gives us a lower bound with the matching order of magnitude of the upper bound provided by Collier-Cartaino, Graber and Jiang [14]. This chapter is

based on the papers “Asymptotics for Turán numbers of cycles in 3-uniform linear hypergraphs” and “3-uniform hypergraphs without a cycle of length five” co-authored with Győri and Methuku.

In the fourth chapter we study generalized Turán problems. The main question is to determine $ex(n, K_3, C_5)$, where $ex(n, K_3, C_5)$ denotes the maximum possible number of copies of K_3 in a C_5-free graph on n vertices. Bollobás and Győri initiated the study and showed that $\frac{1}{3\sqrt{3}}(1+o(1))n^{3/2} \leq \mathrm{ex}(n, K_3, C_5) \leq \frac{5}{4}(1+o(1))n^{3/2}$. Alon and Shikhelman improved this result by reducing the constant in the upper-bound to $\frac{\sqrt{3}}{2}$. In this chapter we introduce a new approach and further improve this bound showing that $\mathrm{ex}(n, K_3, C_5) < (1 + o(1))\frac{1}{3\sqrt{2}}n^{3/2}$. We also give a short proof for slightly weaker bound, based on the paper "A note on the maximum number of triangles in a C_5-free graph" co-authored by Győri, Methuku, and Salia. In the last part of the chapter, we give a new upper bound for maximum number of edges in a graph without C_5 and induced C_4 as a subgraph, which slightly improves the previous bound given by Ergemlidze, Győri and Methuku. This chapter is mainly based on the paper “Triangles in C_5-free graphs and Hypergraphs of Girth Six” co-authored with Methuku.

In Chapter 5 we investigate 15-year old question asked by Mubayi and Verstraëte. Let t, n be integers with $n \geq 3t, t \geq 3$. Let $K_{2,t}^{(3)}$ denote the triple system consisting of $2t$ triples $\{a\} \cup E_1, \{b\} \cup E_1, \{a\} \cup E_2, \{b\} \cup E_2, \ldots, \{a\} \cup E_t, \{b\} \cup E_t$, where a, b are distinct elements and $E_1, \ldots, E_t$ are pairwise disjoint 2-element sets that are disjoint from $\{a, b\}$. About 15 years ago Mubayi and Verstraëte proved that $\mathrm{ex}(n, K_{2,t}^{(3)}) < t^4\binom{n}{2}$, they showed that $g(t) := \lim_{n\to\infty} \mathrm{ex}(n, K_{2,t}^{(3)})/\binom{n}{2}$ and that $\frac{2t-1}{3} \leq g(t) \leq t^4$. and asked if one could determine the growth rate of $g(t)$. we prove that, $g(t) = \Theta(t^{1+o(1)})$, as $t \to \infty$. This shows that their lower bound is close to the truth. More precisely, we prove that $\mathrm{ex}(n, K_{2,t}^{(3)}) \leq (15t \log t + 40t)\, n^2$ for any $t \geq 2$. The chapter is based on the paper “New bounds for a hypergraph Bipartite Turán problem” co-authored with Jiang and Methuku.

In Chapter 6 we study another Turán-type problem. For a fixed graph F the *rainbow Turán number* of F, $\mathrm{ex}^*(n, F)$, is the maximum number of edges in a graph on n vertices that has a proper edge-coloring with no rainbow copy of F. Johnston, Palmer and Sarkar proved in [53] that for any positive integer k $\frac{k}{2}n \leq ex^*(n, P_{k+1}) \leq \left\lceil\frac{3k+1}{2}\right\rceil n$. In this chapter we show that the rainbow Turán number of a path with $k + 1$ edges is less than $\left(\frac{9k}{7} + 2\right) n$, improving an earlier estimate of Johnston, Palmer and Sarkar. The proof is based on the paper “On the Rainbow Turán number of paths” co-authored with Győri and Methuku.

Acknowledgements

I would like to express my gratitude towards my supervisor Ervin Győri for his continuous support during my time as a student at CEU. Special thanks goes to my collaborator and a friend Abhishek Methuku who, together with Ervin Győri, played a major role in my success as a researcher.

I also want to thank my other collaborators Nika Salia, Casey Tomkins and Oscar Zamora. Also I want to thank Antonio Alfieri and Ferenc Bencs for their massive help during my first year of the PhD program. I am very greatful to professor Gyula O.H. Katona who has been extremely kind and supportive in various situations.

Finally, I would like to thank my fiancée Dea Gigauri and my family members Dali Ramishvili, Paata Ergemlidze, Tengiz Ergemlidze and Nani Kurtanidze for their irreplacable love and support.

Chapter 1

Introduction

The *Turán number* $\mathrm{ex}(n, F)$ is the maximum number of edges in an F-free graph on n vertices. Extremal graph theory studies Turán numbers of various graphs. Investigation of this type of problem dates back to 1907, when Mantel [63] proved that the maximum possible number of edges in triangle-free graphs on n vertices is at most $\left\lfloor \frac{n^2}{4} \right\rfloor$. The complete bipartite graph with parts of sizes $\left\lfloor \frac{n}{2} \right\rfloor$ and $\left\lceil \frac{n}{2} \right\rceil$ is a construction for lower bound. It was only years after, when Turán initated systematic studying of similar problems, he proved:

Theorem 1.1 (Turán [71])**.** *The maximum number of edges in a graph on n vertices with no K_{t+1} is at most $\left(1 - \frac{1}{t}\right) \frac{n^2}{2}$.*

The matching lower bound comes from the construction, which is a complete t-partite graph with part sizes being as close as possible. Clearly, each part will be of size either $\left\lfloor \frac{n}{t} \right\rfloor$ or $\left\lceil \frac{n}{t} \right\rceil$.

For a graph G, the chromatic number $\chi(G)$ is the smallest number of colors needed to color the vertex set of G so that no two adjacent vertices share the same color.

For any non-bipartite forbidden graph F, $\mathrm{ex}(n, F)$ has order of magnitude n^2, moreover Erdős, Stone and Simonovits [21, 23] showed that asymptotics of the Turán number of a graph is determined by its chromatic number only.

Theorem 1.2 (Erdős, Stone, Simonovits [21, 23])**.** *For a graph F with $\chi(F) \geq 3$ we have $\mathrm{ex}(n, F) = \left(1 - \frac{1}{\chi(F)-1}\right) \frac{n^2}{2} + o(n^2)$.*

It is fascinating that this one theorem takes care of the huge class of Turán problems. Since then, the study has been mainly directed to the so-called 'degenerate' case, i.e., when the forbidden graph is bipartite. Kővári, Sós and Turán considered one of the most natural degenerate cases and estimated the Turán number of a complete bipartite graph $K_{s,t}$ with parts of sizes s and t.

Theorem 1.3 (Kővári, T.Sós, Turán [58])**.** *Let $K_{s,t}$ denote the complete bipartite graph with s and t vertices in its color-classes. Then*

$$\mathrm{ex}(n, K_{s,t}) \leq \frac{1}{2}\sqrt[s]{t-1}\,n^{2-\frac{1}{t}} + O(n)$$

Clearly, it makes sense to use this theorem for $s < t$ as it is a better estimation this way. Kollár, Rónyai and Szabó [56] provided a lower bound, which matches the order of magnitude of the upper bound, whenever $t > s!$, and later Alon, Rónyai, and Szabó [5] provided a matching lower bound if t is sufficiently big compared to s, more specifically when $t > (s-1)!$.

A very interesting and widely used special case of the theorem 1.3 is when $s = t = 2$, i.e. estimating the Turán number of $K_{2,2}$, which also happens to be a 4-cycle. The theorem clearly implies that

$$\mathrm{ex}(n, C_4) \leq (1 + o(1))\frac{1}{2}n^{3/2}.$$

One would wonder if there exists a matching lower bound and the answer is it does. The construction for the lower bound is following:

Construction. *Let p be a prime number and $n = p^2 - 1$. Let the vertex set be non-zero pairs (x, y) of the residues modulo p. As the edge set we take distinct pairs of vertices (x, y) and (a, b) such that $ax + by = 1$ (modulo p).*

If there is a cycle of length 4 in the constructed graph, then there are vertices (a, b), (u, v), (a', b') and $(u'v')$ such that $au + bv = au' + bv' = a'u + b'v = a'u' + b'v' = 1$. So the system of equations $ax + by = 1$ and $a'x + b'y = 1$ have two distinct pairs of solutions, which is impossible. So the constructed graph is C_4-free.

For each (a, b) the equation $ax + by = 1$ has p solutions and at least $p - 1$ of them is different from (a, b). This implies that the number of edges of the constructed graph is at least $\frac{1}{2}(p^2-1)(p-1)$, so $\mathrm{ex}(p^2-1, C_4) \geq \frac{1}{2}(p^2-1)(p-1)$. By the fact, that prime numbers are 'densely' distributed in integers, we can extend the lower bound for an arbitrary n, therefore, we get $\mathrm{ex}(n, C_4) \geq (1 + o(1))\frac{1}{2}n^{3/2}$.

The next natural step in understanding Turán numbers of bipartite graphs is to determine extremal number of cycles of even length.

Theorem 1.4 (Bondy, Simonovits [10])**.** *For any $k \geq 2$, we have*

$$\mathrm{ex}(n, C_{2k}) = O(n^{1+\frac{1}{k}}).$$

For $k = 2, 3$ and 5, it is proven that the upper bound (the order of magnitude) can not be improved, but generally, whether the upper bound is sharp or not, remains as one of the most intriguing open questions in extremal graph theory.

1.1 Generalized Turán numbers

Since forming of Turán theory, people generalized classical Turán problems in many ways, one class of these generalizations officially carries the name *generalized Turán numbers*. In this section we overview this topic.

Unsurprisingly, we start this topic by a problem provided by Erdős [24]. Erdős has made several conjectures concerning triangles and pentagons, one of them is following:

Conjecture 1.5 (Erdős)**.** The number of cycles of length 5 in a triangle-free graph on n vertices is at most $(n/5)^5$ and equality holds for the blown-up pentagon if $5 \mid n$.

It was only recently, that the conjecture was proven by Hatami, Hladký, Král, Norine, and Razborov [51] and independently by Grzesik [43].

For graphs F and H, let $ex(n, H, F)$ denote the maximum possible number of copies of H in an F-free graph on n vertices. These types of problems are called *generalized Turán problems* and the study of them began after Győri and Bollobás [8] considered the similar problem to Conjecture 1.5, they estimated the number of triangles in graphs without pentagons.

Theorem 1.6 (Győri, Bollobás [8])**.**

$$(1+o(1))\frac{1}{3\sqrt{3}}n^{3/2} \le ex(n, K_3, C_5) \le (1+o(1))\frac{5}{4}n^{3/2}.$$

Their lower bound comes from the following construction: Take a C_4-free bipartite graph G_0 on $n/3 + n/3$ vertices with about $(n/3)^{3/2}$ edges and double each vertex in one of the color classes (each corresponding edge will also be doubled) and add an edge joining the old and the new copy of each vertex to produce a graph G. It is easy to see that G is C_5-free and it contains $(n/3)^{3/2}$ triangles.

More systematic study of the function $\mathrm{ex}(n, H, F)$ was initiated by Alon and Shikhelman in [4], where they improved the result of Bollobás and Győri by showing that $\mathrm{ex}(n, C_3, C_5) \le (1+o(1))\frac{\sqrt{3}}{2}n^{3/2}$. This bound was further improved in [33] by Ergemlidze, Győri, Methuku and Salia and then very recently in [26], by Ergemlidze and Methuku, who showed that $\mathrm{ex}(n, C_3, C_5) < (1+o(1))0.232n^{3/2}$. We provide the proof of this result in Chapter 4.

Clearly, unlike classical Turán problems, determining the order of magnitude of generalized Turán numbers is not trivial for non-bipartite forbidden graphs. Győri and Li [45] provided bounds on number of triangles in C_{2k}-free graphs. They proved

Theorem 1.7 (Győri, Li [45])**.**

$$\mathrm{ex}(n, C_3, C_{2k+1}) < \frac{(2k-1)(16k-2)}{3}\,\mathrm{ex}(n, C_{2k}).$$

The lower bound contains more than $\binom{k}{2}\mathrm{ex}_{bip}(\frac{2n}{k+1}, \{C_4, C_6, \dots, C_{2k}\})$ triangles, where $\mathrm{ex}_{bip}(n, \mathcal{F})$ denotes the maximum number of edges in an $\mathcal{F}$-free bipartite graph on n vertices. Below we consider the corresponding construction:

Take a maximum size bipartite graph $H(X_0, Y)$ where $|X_0| = |Y| = \frac{n}{k+1}$ such that $C_4, C_6, \dots, C_{2k} \notin H$. To get the desired graph G, "blow up" the vertices in X_0, i.e., for every vertex $x \in X_0$ replace x by k vertices $x_1, x_2, \dots, x_k$ joined to each other and to all neighbors of x (in the graph H). The set of these new vertices is denoted by X, and clearly $|X| = k\,|X_0|$, i.e., $|X \cup Y| = k\frac{n}{k+1} + \frac{n}{k+1}$, so the resulting graph G has n vertices. This graph G contains many cycles of length $3, 4, \dots, 2k$, but it can be easily checked that if there is a cycle of length $2k+1$ in G, after contracting back the vertices of the blown up set, we would find an even cycle of length at most $2k+1$ in H, which is a contradiction.

Now let us count the number of triangles in G. For simplicity we count the number of triangles with one vertex in Y and two vertices in X, as most of the triangles are of this type. It is easy to see that each edge of H is replaced by a clique of size $k+1$ in G, therefore there is at least $\binom{k}{2}e(H)$ triangles with one vertex in Y. So this construction gives us a lower bound $\binom{k}{2}\,\mathrm{ex}_{bip}(\frac{2n}{k+1},\{C_4,C_6,\dots,C_{2k+1}\})$. We know that the functions $\mathrm{ex}_{bip}(\frac{2n}{k+1},\{C_4,C_6,\dots,C_{2k+1}\})$, $\mathrm{ex}(\frac{2n}{k+1},\{C_4,C_6,\dots,C_{2k+1}\})$ and $\mathrm{ex}(\frac{2n}{k+1},C_{2k+1})$ are essentially the same, therefore, this construction proves that Theorem 1.7 is very close to being sharp.

1.2 Turán problems in hypergraphs

Counting triangles in graphs is closely related to counting hyperedges in 3-uniform hypergraphs. This leads us to another closely related topic, which is one of the main themes of the thesis, Turán numbers of hypergraphs.

A hypergraph $H=(V,E)$ is a family E of distinct subsets of a finite set V. The members of E are called *hyperedges* and the elements of V are called *vertices*. A hypergraph is called r-uniform if each member of E has size r. A hypergraph $H=(V,E)$ is called *linear* if every two hyperedges have at most one vertex in common.

For a family of forbidden r-uniform hyergraphs $\mathcal{F}$ the Turán number $\mathrm{ex}_r(n,\mathcal{F})$ denotes the maximum number of hyperedges in an r-uniform hypergraph on n vertices with no element of $\mathcal{F}$ as a subhypergraph. For convenience, whenever $\mathcal{F}=\{F\}$ consists of a single forbidden hypergraph, we write $\mathrm{ex}_r(n,F)$ instead of $\mathrm{ex}_r(n,\{F\})$.

The linear Turán number $\mathrm{ex}_r^{lin}(n,\mathcal{F})$ is the maximum number of hyperedges in an r-uniform linear hypergraph on n vertices with no element of $\mathcal{F}$ as a subhypergraph.

A very natural and widely studied topic is Turán numbers of cycles in hypergraphs. Unlike graphs, there are several types of cycles in hypergraphs, most common of them would be Berge cycles and linear cycles.

Definition 1.8. *For an integer $k\geq 2$, a* Berge cycle *of length k, denoted by C_k, is an alternating sequence $v_1h_1v_2h_2\dots v_kh_kv_1$ of distinct vertices and edges such that $\{v_i,v_{i+1}\}\subseteq h_i$ for $1\leq i\leq k-1$, and $\{v_k,v_1\}\subseteq h_k$.*

A linear cycle *(often also called a loose cycle) in a hypergraph is a Berge cycle where only the cyclically consecutive hyperedges intersect and they intersect in exactly one vertex.*

It is worth noting that even in linear hypergraphs, Berge and linear cycles differ from each other. Although, in the case of cycles of length 3, in linear hypergraphs these two classes coincide, so linear Turán number of Berge triangle is the same as linear Turán number of linear triangle. Determining $ex_3^{lin}(n,C_3)$ is basically equivalent to the famous $(6,3)$-problem, which is a special case of a general problem of Brown, Erdős, and Sós. The famous theorem of Ruzsa and Szemerédi states:

Theorem 1.9 (Ruzsa, Szemerédi [68]). *There exists a constant $c>0$ for which we have*

$$n^{2-\frac{c}{\sqrt{logn}}} < ex_3^{lin}(n,C_3)=o(n^2).$$

The systematic study of Turán numbers of Berge cycles started with the investigation of Berge triangles by Győri [47], who proved that the maximum number of hyperedges in a Berge triangle-free 3-uniform hypergraph on n vertices is at most $n^2/8$. The construction for lower bound is following: Take 3 disjoint sets, $A = \{a_1, a_2, \ldots, a_{n/4}\}$, $A' = \{a'_1, a'_2, \ldots, a'_{n/4}\}$ and $B = \{b_1, b_2, \ldots, b_{n/2}\}$. The hypergraph H, whose vertex set is $A \cup A' \cup B$ and the edge set is $\{a_i, a'_i, b_j \mid 1 \le i \le n/4, 1 \le j \le n/2\}$, is Berge triangle-free and has $n^2/8$ hyperedges. There is an informal, but convenient way to see this construction. First we take a complete bipartite graph and then we make a copy of each vertex on one side, this way we create a triple corresponding each edge of the original bipartite graph and we assign hyperedges to these triples. It is worth noting that this type of hypergraph extension of a graph is quite common and we will come across similarly obtained hypergraphs throughout this thesis.

The study of Berge cycle-free hypergraphs were continued by Bollobás and Győri [8], who showed that $n^{3/2}/3\sqrt{3} \le ex_3(n, C_5) \le \sqrt{2}n^{3/2} + 4.5n$. Very recently, this estimate was considerably improved by Ergemlidze, Győri and Methuku [29]. They also considered [30] the analogous question for linear hypergraphs and proved that $\mathrm{ex}_3^{\mathrm{lin}}(n, C_5) = \frac{1}{3\sqrt{3}}n^{3/2} + O(n)$. Surprisingly, even though the lower bound here is the same as the lower bound in the Bollobás-Győri theorem, the hypergraph they construct in order to establish their lower bound is very different from the hypergraph used in the Bollobás-Győri theorem. The latter is far from being linear. We discuss more details about these problems and provide proofs for some of them in Chapter 3.

Győri and Lemons considered a more general question and estimated Turán number of Berge cycles of any given length.

Theorem 1.10 (Győri, Lemons, [46, 44])**.** *For $r \ge 2$, we have $\mathrm{ex}_r(n, C_{2l}) = O(n^{1+1/l})$.*
For $r \ge 3$, we have $\mathrm{ex}_r(n, C_{2l+1}) = O(n^{1+1/l})$.

Recently, Füredi, Kostochka and Luo [39] proved similar results for Berge cycles. Instead of forbidding Berge cycles of fixed length they forbid all Berge cycles of length at least k.

Theorem 1.1 (Füredi, Kostochka, Luo [39])**.** *Let $r \ge 3$ and $k \ge r+3$, and suppose $\mathcal{H}$ is an n-vertex r-graph with no Berge cycle of length k or longer. Then $e(\mathcal{H}) \le \frac{n-1}{k-2}\binom{k-1}{r}$.*

Moreover, Kostochka and Luo [57] found bounds for $k \le r-1$ and for $k = r$. For the remaining two cases $k = r+2$ and $k = r+1$, Füredi, Kostochka and Luo [39] conjectured that a similar statement as that of Theorem 1.1 holds. Recently, Ergemlidze, Győri, Methuku, Tompkins, Salia and Zamora [35] proved these conjectures.

Apart from cycles, we consider a hypergraph Turán problem of following extension of a complete bipartite graph.

Let $K_{2,t}^{(r)}$ denote the r-uniform hypergraph consisting of $2t$ hyperedges $\{a\} \cup E_1, \{b\} \cup E_1, \{a\} \cup E_2, \{b\} \cup E_2, \ldots, \{a\} \cup E_t, \{b\} \cup E_t$, where a, b are distinct vertices and $E_1, \ldots, E_t$ are pairwise disjoint $(r-1)$-uniform sets that are disjoint from $\{a, b\}$. For more clarity, in Figure 1.1 we see an example of $K_{2,3}^{(3)}$. (Note that in Figure 1.1 triangles correspond to hyperedges)

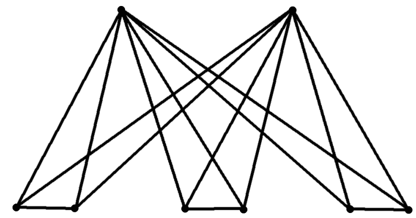

Figure 1.1: Example of $K_{2,3}^{(3)}$

Definition 1.11. *For all $n \geq r \geq 3$, let $f_r(n)$ denote the maximum number of edges in an n-vertex r-uniform hypergraph containing no four edges A, B, C, D with $A \cup B = C \cup D$ and $A \cap B = C \cap D = \emptyset$.*

Note that $f_3(n) = \mathrm{ex}(n, K_{2,2}^{(3)})$, and in general $f_r(n) \leq \mathrm{ex}(n, K_{2,2}^{(r)})$. Erdős [17] asked whether $f_r(n) = O(n^{r-1})$ when $r \geq 3$. Erdős and Frankl proved that $f_r(n) = O(n^{1-\frac{1}{2}})$ but they never published it. Later Füredi [36] answered Erdős' question affirmatively.

Theorem 1.12 (Füredi [36])**.** *For all integers n, r with $r \geq 3$ and $n \geq 2r$,*

$$\binom{n-1}{r-1} + \left\lfloor \frac{n-1}{r} \right\rfloor \leq f_r(n) < 3.5\binom{n}{r-1}.$$

The lower bound is obtained by taking the family of all r-element subsets of $[n] := \{1, 2, \ldots, n\}$ containing a fixed element, say 1, and adding to the family any collection of $\left\lfloor \frac{n-1}{r} \right\rfloor$ pairwise disjoint r-element subsets not containing 1. For $r = 3$, Füredi also gave an alternative lower bound construction using Steiner systems. An (n, r, t)-*Steiner system* $S(n, r, t)$ is an r-uniform hypergraph on $[n]$ in which every t-element subset of $[n]$ is contained in exactly one hyperedge. Füredi observed that if we replace every hyperedge in $S(n, 5, 2)$ by all its 3-element subsets then the resulting triple system has $\binom{n}{2}$ triples and contains no copy of $K_{2,2}^{(3)}$. This slightly improves the lower bound in Theorem 1.12 for $r = 3$ to $\binom{n}{2}$, for those n for which $S(n, 5, 2)$ exists. The upper bound in Theorem 1.12 was improved by Mubayi and Verstraëte [64] to $3\binom{n}{r-1} + O(n^{r-2})$. They obtain this bound by first showing $f_3(n) = \mathrm{ex}(n, K_{2,2}^{(3)}) < 3\binom{n}{2} + 6n$, and then combining it with a simple reduction lemma. This was later improved to $f_3(n) \leq \frac{13}{9}\binom{n}{2}$ by Pikhurko and Verstraëte [65].

Motivated by Füredi's work, Mubayi and Verstraëte [64] initiated the study of the general problem of determining $\mathrm{ex}(n, K_{2,t}^{(r)})$ for any $t \geq 2$.

Theorem 1.13. *$t \geq 2$ and $n \geq 2t$*

$$\mathrm{ex}(n, K_{2,t}^{(3)}) < t^4\binom{n}{2}.$$

Morover, for infinitely many n,

$$\mathrm{ex}(n, K_{2,t}^{(3)}) \geq \frac{2t-1}{3}\binom{n}{2}$$

.

Mubayi and Verstraëte noted that $g(t) := \lim_{n\to\infty} \mathrm{ex}(n, K_{2,t}^{(3)})/\binom{n}{2}$ exists and raised the question of determining the growth rate of $g(t)$. Ergemlidze, Jiang and Methuku [27] determined $g(t)$ within a margin of $\log t$ factor. The proof is provided in Chapter 5.

1.3 Linear cycle-free 3-uniform hypergraphs

In this section we again consider extremal hypergraph problems, but instead of estimating the number of hyperedges of hypergraphs without particular forbidden structures, we observe some properties.

Much like the graph case (as a graph without a cycle is a forest), a hypergraph without a Berge cycle is a disjoint union of *linear trees* (a linear forest), where a *linear tree* is a hypergraph obtained from a vertex by repeatedly adding hyperedges that intersect the previous hypergraph in exactly one vertex. The same is true for linear cycle-free linear hypergraphs, but things are not as straightforward when we forbid just linear cycles in hypergraphs. The second chapter of the thesis is dedicated to understanding more about linear cycle-free 3-uniform hypergraphs.

It comes as no surprise, that problems in hypergraphs often arise from natural extension of similar graph problems. Below we address several of them.

An independent set of a hypergraph H is a set of vertices that contain no hyperedges of H. Let $\alpha(H)$ denote the size of a largest independent set of H and we call it the independence number of H. A well-known theorem of Pósa [67] states that the vertex set of every graph G can be partitioned into at most $\alpha(G)$ cycles where $\alpha(G)$ denotes the independence number of G (where a vertex or an edge is accepted as a cycle). Gyárfás and Sárközy [49] conjectured that the following extension of Pósa's theorem holds.

Conjecture 1.14 (Gyárfás, Sárközy [49])**.** One can partition every k-uniform hypergraph H into at most $\alpha(H)$ linear cycles, hyperedges and subsets of hyperedges.

While the original conjecture stays open, in [49] Gyárfás and Sárközy proved a weaker form of the conjecture, where they used weak cycles instead of linear cycles. In weak cycles only consecutive hyperedges are allowed to intersect, but unlike linear cycles, the intersection can be more than a single vertex. Recently, Ergemlidze, Győri and Methuku [32] proved another weaker version of the conjecture, where they showed that every 3-uniform hypergraph can be covered with at most $\alpha(H)$ edge-disjoint linear cycles. Proof of this theorem is provided in Chapter 2.

Motivated by solving Conjecture 1.14, Gyárfás, Győri and Simonovits showed that the conjecture holds for linear cycle-free 3-uniform hypergraphs. Before stating the theorem we need a definition of a chromatic number for hypergraphs.

For a hypergraph H a chromatic number $\chi(H)$ is the smallest number of colors needed to color the vertex set of H so that there is no hyperedge of H with all of its vertices sharing the same color.

Theorem 1.2 (Gyárfás, Győri, Simonovits [48]). *If H is a 3-uniform hypergraph without linear cycles then it can be partitioned into $\alpha(H)$ linear cycles, hyperedges and subsets of hyperedges. Moreover $\chi(H) \leq 3$.*

Unlike the equivalent graph problem, Theorem 1.2 is far from being trivial. Gyárfás, Győri and Simonovits further investigated the relation between linear cycles and independent number of 3-uniform hypergraphs.

Theorem 1.3 (Gyárfás, Győri, Simonovits [48]). *If H is a 3-uniform hypergraph without linear cycles on n vertices then $\alpha(H) \geq \frac{2}{5}n$.*

The hypergraph consisting of vertex disjoint copies of K_5^3 (a complete 3-uniform hypergraph on 5 vertices) shows that equality can hold in Theorem 1.3. Gyárfás, Győri, Simonovits asked whether the lower bound of $\alpha(H)$ can be improved if we exclude K_5^3 as a subhypergraph and whether such hypergraph is 2-colorable. Ergemlidze, Győri and Methuku [28] answered these questions affirmatively. The proof is provided in Chapter 2.

1.4 Rainbow Turán numbers

In this section, we overview the study of rainbow Turán numbers, which effectively merges classical Turán problems with the extremal problems on edge-colorings of graphs.

A graph is properly edge-colored if every pair of incident edges have distinct colors. An edge-colored graph is called *rainbow* if all its edges have different colors.

Given a graph F, the *rainbow Turán number* of F is defined as the maximum number of edges in a graph on n vertices that has a proper edge-coloring with no rainbow copy of F, and it is denoted by $\mathrm{ex}^*(n, F)$. Clearly, $\mathrm{ex}(n, F) \leq \mathrm{ex}^*(n, F)$.

The special case of the Canonical Ramsey Theorem of Erdős and Rado [22], says that any proper edge-coloring of K_n contains a rainbow K_m as a subgraph, provided that n is sufficiently large in relation to m. Motivated by this, Alon, Jiang, Miller and Pritikin [3] introduced a problem of finding a rainbow copy of a graph H in a coloring of K_n in which each color appears at most m times at each vertex. The rainbow Turán problem is a natural extension of this problem.

The systematic study of rainbow Turán numbers was initiated in [55] by Keevash, Mubayi, Sudakov and Verstraëte. Before stating their result, we need a definition:

We say that a graph G is color-critical if there exists an edge $e \in E(G)$ such that $\chi(G \setminus e) = \chi(G) - 1$ (note that the definition is non-standard).

Proposition 1.15 (Keevash, Mubayi, Sudakov, Verstraëte.). *For a non-bipartite graph F we have*

$$\mathrm{ex}^*(n, F) = \mathrm{ex}(n, F) + o(n^2).$$

Morover, if G is color critical then $\mathrm{ex}^*(n, F) = \mathrm{ex}(n, F)$ *for large enough n.*

For every bipartite graph F with a maximum degree of s in one of the parts, they proved $\mathrm{ex}^*(n,F) = O(n^{2-1/s})$. This matches the upper bound for the (usual) Turán numbers of such graphs.

Keevash, Mubayi, Sudakov and Verstraëte also studied the rainbow Turán problem for even cycles. More precisely, they showed that

$$\mathrm{ex}^*(n, C_{2k}) = \Omega(n^{1+1/k}).$$

For this they used the construction of large B_k^*-sets of Bose and Chowla [11]– it is conjectured that the same lower bound holds for $\mathrm{ex}^*(n, C_{2k})$ and is a well-known difficult open problem in extremal graph theory. They also proved the matching upper bound in the case of the six-cycle C_6, so it is known that $\mathrm{ex}^*(n, C_6) = \Theta(n^{4/3}) = \mathrm{ex}(n, C_6)$. However, interestingly, they showed that $\mathrm{ex}^*(n, C_6)$ is asymptotically larger than $\mathrm{ex}(n, C_6)$ by a multiplicative constant. Recently, Das, Lee and Sudakov [15] showed that

$$\mathrm{ex}^*(n, C_{2k}) = O(n^{1+\frac{(1+\epsilon_k)\ln k}{k}}),$$

where $\epsilon_k \to 0$ as $k \to \infty$.

Johnston, Palmer and Sarkar continued studying by investigating rainbow Turán numbers of matchings, paths and forests of stars. Let M_k denote a matching of size k, and let P_k denote a path of length k. In [53], Johnston, Palmer and Sarkar showed that for sufficiently large n, perhaps surprisingly, $\mathrm{ex}^*(n, M_k) = \mathrm{ex}(n, M_k)$. They also showed that $\mathrm{ex}^*(n, P_k) \leq \left\lceil \frac{3k-2}{2} \right\rceil n$. Recently, Ergemlidze, Methuku and Győri [34] improved the upper bound of $\mathrm{ex}^*(n, P_k)$ and we provide the proof in Chapter(6).

Chapter 2

3-uniform hypergraphs and linear cycles

2.1 Introduction

A hypergraph H is 2 colorable if there is a coloring of the vertices of H such that there is no monochromatic hyperedge in H. We denote the complete 3-uniform hypergraph on 5 vertices by K_5^3. Throughout the chapter, we mostly use the terminology introduced in [48].

Definition 2.1. *A* linear tree *is a hypergraph obtained from a vertex by repeatedly adding hyperedges that intersect the previous hypergraph in exactly one vertex. A* linear path *is a linear tree built so that the next hyperedge always intersects the previous hyperedge in a vertex of degree one.*

A linear cycle *is obtained from a linear path of at least two edges, by adding an edge that intersects the first and the last edges of the linear path in one of their degree one vertices.*

A skeleton *T in H is a linear subtree of H which cannot be extended to a larger linear subtree by adding a hyperedge e of H for which $|e \cap V(T)| = 1$.*

Recall that an independent set of a hypergraph H is a set of vertices that contain no hyperedges of H. $\alpha(H)$ denotes the size of a largest independent set of H and we call it the independence number of H.

Gyárfás, Győri and Simonovits [48] initiated the study of linear cycle-free hypergraphs by showing,

Theorem 2.1 (Gyárfás, Győri, Simonovits [48])**.** *If H is a 3-uniform hypergraph on n vertices without linear cycles, then it is 3-colorable. Moreover, $\alpha(H) \leq \frac{2n}{5}$.*

We proved,

Theorem 2.2 (E., Győri, Methuku [28])**.** *Let H be a 3-uniform hypergraph without linear cycles, and no K_5^3 as a sub-hypergraph. Then it is 2-colorable.*

Corollary 2.2. *Let H be a 3-uniform hypergraph without linear cycles, and no K_5^3 as a sub-hypergraph. Then $\alpha(H) \geq \lceil \frac{n}{2} \rceil$ and it is sharp.*

Indeed, from Theorem 2.2, it trivially follows that $\alpha(H) \geq \lceil \frac{n}{2} \rceil$. The hypergraph H_n on n vertices obtained from the following construction shows that this inequality is sharp. Let H_3 be the hypergraph on 3 vertices v_1, v_2, v_3 such that $v_1v_2v_3 \in E(H_3)$ and let H_4 be the complete 3-uniform hypergraph K_4^3 on 4 vertices v_1, v_2, v_3, v_4. Now for each $3 \leq i \leq n-2$ let us define the hypergraph H_{i+2} such that $V(H_{i+2}) := V(H_i) \cup \{v_{i+1}, v_{i+2}\}$ and $E(H_{i+2}) := E(H_i) \cup \cup_{j=1}^{i} \{v_{i+1}v_{i+2}v_j\}$. If n is even, we start this iterative process with the hypergraph H_4 and if n is odd, we start with H_3. Notice that $\alpha(H_{i+2}) = \alpha(H_i) + 1$ for each i, which implies that $\alpha(H_n) = \lceil \frac{n}{2} \rceil$.

Given a 3-uniform hypergraph H, if $v \in V(H)$ the link of v in H is defined as the graph with vertex set $V(H)$ and edge set $\{(x, y) : (v, x, y) \in E(H)\}$. The *strong degree* $d^+(v)$ for $v \in V$ is the maximum number of independent edges in the link of v. The *degree* of $v \in V$ is simply the number of hyperedges of H containing v.

Theorem 2.3 (Gyárfás, Győri, Simonovits [48]). *Suppose that H is a 3-uniform hypergraph with $d^+(v) \geq 3$ for all $v \in V$. Then H contains a linear cycle.*

We showed,

Theorem 2.4 (E., Győri, Methuku [28]). *Let H be a 3-uniform hypergraph on $n \geq 10$ vertices without linear cycles. Then, there is a vertex whose degree is at most $n-2$.*

We remark that on 9 vertices there is a 3-uniform hypergraph without linear cycles where the degree of every vertex is 8. This hypergraph H is defined by taking a copy of K_4^3 on vertices $\{u_1, u_2, v_1, v_2\}$ and a vertex disjoint copy of K_5^3 such that $u_1u_2x, v_1v_2x \in E(H)$ for each $x \in V(K_5^3)$ and there are no other hyperedges in H.

Also notice that Theorem 2.4 cannot be improved because there is a 3-uniform hypergraph H', with $E(H') := \{xab \mid \{a, b\} \in V(H') \setminus \{x\}\}$, in which every vertex has degree at least $n-2$.

In this chapter we investigate one more problem which describes a connection between the independence number and linear cycles. Recall from Chapter 1, that a theorem of Pósa [67] states that the vertex set of every graph G can be partitioned into at most $\alpha(G)$ cycles (where a vertex or an edge is accepted as a cycle). Gyárfás and Sárközy [49] conjectured that the following extension of Pósa's theorem holds: One can partition every k-uniform hypergraph H into at most $\alpha(H)$ linear cycles (here, as in Pósa's theorem, vertices and subsets of hyperedges are accepted as linear cycles).

We show their conjecture is true for $k = 3$ provided we allow the linear cycles to be edge-disjoint, instead of being vertex-disjoint.

Theorem 2.5 (E., Győri, Methuku [32]). *If H is a 3-uniform hypergraph, then its vertex set can be covered by at most $\alpha(H)$ edge-disjoint linear cycles (where we accept a single vertex or a hyperedge as a linear cycle).*

Our proof uses induction on $\alpha(H)$. However, perhaps surprisingly, in order to make induction work, our main idea is to allow the hypergraph H to contain hyperedges of size 2 (in addition to hyperedges of size 3). First we will delete some vertices, and add certain hyperedges of size 2 into the remaining hypergraph so as to ensure the independence number of the remaining hypergraph is smaller than that of H. Then applying induction we will find edge-disjoint linear cycles (which may contain these added hyperedges) covering the remaining hypergraph. It will turn out that the added hyperedges behave nicely, allowing us to construct edge-disjoint linear cycles in H covering all of its vertices.

This Chapter is organized as follows: In Section 2.2 we prove Theorem 2.2 using our main lemma - Lemma 2.6 (which is proved in Section 2.2.1). In Section 2.3 we prove Theorem 2.4. Finally in Section 2.4, we present the proof of Theorem 2.5.

2.2 Proof of Theorem 2.2

Let H be our 3-uniform hypergraph without linear cycles. From now on, we write the hyperedge $\{a, b, c\} \in E(H)$ as abc for convenience.

Definition 2.3. *Given a vertex $v \in V(H)$ and a hyperedge $abc \in E(H)$, we say that v is* "strongly associated" *to abc if at least two of the three edges vab, vbc, vca are in $E(H)$. We say that v is* "weakly associated" *to abc if exactly one of the three edges vab, vbc, vca is in $E(H)$. We say that v is associated to abc if it is either strongly or weakly associated.*

The set of pairs $\{\{x, y\} \subset \{a, b, c\} \mid vxy \in E(H)\}$ is called the "support" of v in abc, denoted $s_{abc}(v)$.

Definition 2.4 (thick pair)**.** *For any two vertices, $a, b \in V(H)$, we call the pair $\{a, b\}$ "thick" if there are at least two different hyperedges each containing $\{a, b\}$. We call a hyperedge abc "thick" if all the pairs $\{a, b\}$, $\{b, c\}$ and $\{c, a\}$ are thick.*

Lemma 2.5. *If $abc \in E(H)$ is a thick hyperedge, then the set of vertices associated to it consists of one of the following*

1. *Exactly two vertices that are strongly associated to abc.*

2. *Exactly one vertex that is strongly associated to abc and vertices $w_1, w_2, \ldots, w_m$ such that each w_i is weakly associated to abc and $|\cup_i s_{abc}(w_i)| = 1$. (It is possible that $m = 0$, i.e., no such w_i exists).*

Proof. If there is no vertex strongly associated to abc, then since abc is thick, we must have 3 distinct vertices v_1, v_2, v_3 such that $v_1ab, v_2bc, v_3ca \in E(H)$, a linear cycle, a contradiction. So there must be a vertex strongly associated to abc.

Now we show that if there are two vertices p, q strongly associated to a hyperedge $abc \in E(H)$, then there are no other vertices associated to abc. Suppose by contradiction that there are such vertices. Then, among these vertices there is a vertex r such that $|s_{abc}(p) \cup s_{abc}(q) \cup s_{abc}(r)| = 3$ since abc is thick. Now consider the bipartite graph whose two color classes are $\{p, q, r\}$ and $\{\{a, b\}, \{b, c\}, \{c, a\}\}$ where $v \in \{p, q, r\}$ is connected to $\{x, y\} \in \{\{a, b\}, \{b, c\}, \{c, a\}\}$ if $vxy \in E(H)$. It can be easily checked that Hall's condition holds for the color class $\{p, q, r\}$ and so there exists a matching between the two color classes, but this produces a linear cycle (of size 3) in H, a contradiction.

So the only remaining possibility is that abc has exactly one vertex which is strongly associated to it and maybe some other vertices $w_1, w_2, \dots, w_m$ that are weakly associated to it. We only have to show that $|\cup_i s_{abc}(w_i)| = 1$. Suppose by contradiction that there are vertices w_i and w_j such that their supports in abc are different. Let $s_{abc}(w_i) = \{\{a,b\}\}$ and $s_{abc}(w_j) = \{\{b,c\}\}$ without loss of generality. Then, since abc is thick, there is a vertex v such that $v \neq w_i$, $v \neq w_j$ and $acv \in E(H)$. Now, acv, abw_i, bcw_j is a linear cycle, a contradiction. □

Lemma 2.6 (Main Lemma)**.** *Let T be a linear tree. Then there exists a coloring of $V(T)$, such that the hypergraph induced by $V(T)$ is properly colored and for each vertex $v \in V(H) \setminus V(T)$ where v is strongly associated to some hyperedge of T, there exists a coloring of v such that hyperedges vab with $a, b \in V(T)$ are properly colored, and for each remaining vertex $v \in V(H) \setminus V(T)$ the hyperedges vab with $a, b \in V(T)$ are properly colored regardless of the color of v.*

Before we prove this lemma, we will show how to prove Theorem 2.2 using it.

Observation 2.7. Let $w \in V(T)$. Notice that the above lemma holds even if we add the extra condition that the color of w is given.

Now we prove our main theorem using this lemma.

Proof of Theorem 2.2. Let T_1 be any skeleton of H. Then there exists a coloring of T_1 given by Lemma 2.6. Let $U_1 \subseteq V(H) \setminus V(T_1)$ be the set of all vertices such that each $u \in U_1$ is strongly associated to some hyperedge of T_1. If $|U_1| = 0$, then by Lemma 2.6 all the vertices of $V(H) \setminus V(T_1)$ can be colored arbitrarily such that the hyperedges vab with $a, b \in V(T_1)$ are properly colored. Also, since T_1 is a skeleton, there are no hyperedges vxy where $v \in V(T_1)$ and $x, y \in V(H) \setminus V(T_1)$. Therefore, the vertices of $V(H) \setminus V(T_1)$ can be colored independently from vertices of $V(T_1)$ and so we have the same problem for the subhypergraph induced by $V(H) \setminus V(T_1)$. So we can assume that $|U_1| \neq 0$. Now let us define a sequence of linear trees $T_1, T_2, \dots, T_i, T_{i+1}, \dots, T_m$ recursively as follows: Let $U_i \subseteq V(H) \setminus \cup_{j=1}^{i} V(T_j)$ be the set of vertices where each $u \in U_i$ is strongly associated to some hyperedge of $\cup_{j=1}^{i} T_j$ and let T_{i+1} be the skeleton in the subhypergraph induced by $V(H) \setminus \cup_{j=1}^{i} V(T_j)$ which contains at least one vertex from U_i (we continue this procedure as long as $|U_i| \neq 0$; so $|U_m| = 0$). In fact, we will show that $|V(T_{i+1}) \cap U_i| = 1$. Let H_i denote the subhypergraph of H induced by $\cup_{j=1}^{i} V(T_j)$.

Claim 2.8. *For each $1 \leq i \leq m-1$, there is a linear path in H_i between any two vertices $u, v \in V(H_i)$. Moreover, $V(T_{i+1}) \cap U_i$ consists of only one vertex and this vertex can be strongly associated to hyperedge(s) of T_s for exactly one $1 \leq s \leq i$.*

Proof of Claim 2.8. We prove the claim by induction on i. For $i = 1$, the statement is trivial. Assume the statement is true for $i = k$. First we will show that there is a linear path between $u \in V(T_{k+1}) \cap U_k$ and any $v \in V(H_k)$. Let $abc \in E(T_s)$ (for some $1 \leq s \leq k$) be the hyperedge in $\cup_{j=1}^{k} T_j$ that is strongly associated to u. Consider the shortest linear path $\mathcal{P}_1$ containing v and a vertex of $\{a, b, c\}$ (in case, $v \in \{a, b, c\}$, $\mathcal{P}_1$ consists of just v). Clearly,

$\mathcal{P}_1$ cannot contain all 3 of the vertices a, b, c since it is a shortest path. If $\mathcal{P}_1$ contains only one vertex from $\{a, b, c\}$, say a w.l.o.g, then since u is strongly associated to abc, either uac or uab is in H_{k+1}, which together with $\mathcal{P}_1$ gives us a linear path from u to v as desired. If $\mathcal{P}_1$ contains two vertices of $\{a, b, c\}$, say a, b w.l.o.g, then either uac or ubc is in H_{k+1} which together with $\mathcal{P}_1$ gives us a linear path from u to v. Notice that this path contains only one vertex from T_{k+1}. Since there is a linear path between every 2 vertices of T_{k+1} we have a linear path between any vertex of T_{k+1} and any vertex of H_k. By the induction hypothesis there is a linear path between any two vertices of H_k and so we have proved the first part of the claim.

Now assume by contradiction that there are two vertices $u, u' \in V(T_{k+1}) \cap U_k$. Let pqr be the hyperedge in $\cup_{j=1}^{k} T_j$ that is strongly associated to u'. Consider the shortest linear path $\mathcal{P}_2$ containing u and a vertex of $\{p, q, r\}$. By the same argument as before, we can assume that $\mathcal{P}_2$ contains at most two vertices of $\{p, q, r\}$ and there is a hyperedge e such that $\mathcal{P}_2 \cup e$ is a linear path between u and u' which doesn't contain any other vertices of T_{k+1}. However, $\mathcal{P}_2 \cup e$ together with the linear path between u and u' in T_{k+1} gives us a linear cycle, a contradiction.

So $V(T_{k+1}) \cap U_k$ consists of only vertex, say u. If u is strongly associated to two hyperedges $h_1 \in T_r$ and $h_2 \in T_s$ (where $r \neq s$ and $r, s \leq k$), then the shortest linear path $\mathcal{P}$ between h_1 and h_2 consists of at least one hyperedge. By a similar argument as before, there are hyperedges e_1, e_2 containing u such that $\mathcal{P}$, e_1 and e_2 form a linear cycle, a contradiction. □

We will show that for each $1 \leq k \leq m$, H_k is properly colored such that each T_i, $i \leq k$ is colored according to Lemma 2.6. For $k = 1$ the above statement is trivially true. Let us assume that the statement is true for k and show that it is true for $k + 1$.

By the above claim $V(T_{k+1}) \cap U_k$ consists of only one vertex u and this vertex is strongly associated to hyperedge(s) of T_s for exactly one $1 \leq s \leq k$. Also, it is easy to see that if $uab \in H_{k+1}$ and $a, b \in V(H_k)$ then $a, b \in V(T_i)$ for some $i \leq k$. If $i = s$ and $a, b \in V(T_s)$, then we know by Lemma 2.6 that there exists a color for u, say c such that hyperedges uab are properly colored. Let us color u by c. If $i \neq s$, and $a, b \in V(T_i)$ then regardless of the color of u the hyperedges uab are colored properly due to Lemma 2.6. Since the set of vertices that are strongly associated to hyperedges of T_{k+1} is disjoint from $V(H_k)$ (the already colored part), we can apply Lemma 2.6 to color T_{k+1} such that u is still colored with c by Observation 2.7. Therefore, we have shown that H_{k+1} is properly colored such that each T_i, $i \leq k + 1$ is colored according to Lemma 2.6, as desired and so we have statement for H_m by induction.

In the remaining vertices, namely $V(H) \setminus V(H_m)$, since there are no strongly associated vertices, by Lemma 2.6 they can be colored independently from H_m and we now have a smaller vertex set: $V(H) \setminus V(H_m)$ to color. Therefore, by induction on number of vertices we may color H properly. □

2.2.1 Proof of Lemma 2.6 (Main Lemma)

let $abc \in E(T)$ and let $u \in V(T)$ be a vertex that's strongly associated to abc. Then there is a hyperedge uvw of the skeleton such that $|\{u,v,w\} \cap \{a,b,c\}| = 1$ because otherwise w.l.o.g there is a linear path $\mathcal{P}$ in T between u and a which doesn't contain b and c. Since u is strongly associated to abc, either uab or uac is a hyperedge of H. This hyperedge together with $\mathcal{P}$ produces a linear cycle in H, a contradiction.

We identify some sets of vertices of size 5 which play an important role in the forthcoming proof.

Definition 2.9. *Let $h_1 = abc$, $h_2 = bde$ where $h_1, h_2 \in E(T)$. If there is no hyperedge $h \in H$ such that $|h \cap (h_1 \cup h_2)| = 2$, then the set of vertices $\{a,b,c,d,e\}$ is called a* Special Block *of T.*

Claim 2.10. *Let $h_1 = abc$, $h_2 = bde$ where $h_1, h_2 \in E(T)$ are thick hyperedges. If $abe, cbd \in E(H)$ or $abd, cbe \in E(H)$, then $\{a,b,c,d,e\}$ is a Special Block.*

Proof of Claim 2.10. It's easy to see that if $\{x,y\} \in \{a,c,d,e\}$ then either h_1, h_2 and xyz or l_1, l_2 and xyz will create linear cycle. So the only cases that are left to be considered are $\{x,y\} = \{d,b\}$ or $\{x,y\} = \{e,b\}$. Since $\{d,e\}$ is a thick pair either dea or dec is a hyperedge in H. W.l.o.g. let's say $dec \in E(H)$. Then in either of the two remaining cases, xyz along with abc and dec will create a linear cycle, a contradiction. □

Claim 2.11. *Let $h_1, h_2 \in E(T)$ be thick hyperedges. If there are two vertices of h_2 which are strongly associated to h_1, then $h_1 \cup h_2$ is a Special Block.*

Proof of Claim 2.11. We know that $|h_1 \cap h_2| = 1$ since a vertex of h_2 is strongly associated to h_1. Let $h_1 = abc$ and $h_2 = dbe$. So d and e are strongly associated to h_1. Assume by contradiction that there exists a hyperedge $xyz \in H$ such that $\{x,y\} \subset \{a,b,c,d,e\}$ and $z \notin \{a,b,c,d,e\}$. First let us observe that $\{x,y\} \not\subset \{a,b,c\}$ because the hyperedge abc already has two vertices d, e strongly associated to it and hence cannot have any other vertex associated to it due to Lemma 2.5. So if we consider the bipartite graph whose color classes are $\{d,e\}$ and $\{\{a,b\},\{b,c\}\}$ where $v \in \{d,e\}$ is connected to $\{x,y\} \in \{\{a,b\},\{b,c\}\}$ if $vxy \in E(H)$, it's easy to see that Hall's condition holds for this bipartite graph. Hence there is a matching. So either $abe, cbd \in E(H)$ or $abd, cbe \in E(H)$. Now, by applying Claim 2.10, we can conclude that $\{a,b,c,d,e\}$ is a Special Block. □

If a hyperedge h is strongly associated to a vertex of another hyperedge h', then it is easy to see that there is a vertex in h which is associated to h'. Therefore, the above claim implies that vertices of $h_1 \cup h_2$ can't be strongly associated to any hyperedge of $E(T) \setminus \{h_1, h_2\}$.

Since the hypergraph induced on $\{a,b,c,d,e\}$ is not K_5^3, it is easy to see that there is a proper coloring $c : \{a,b,c,d,e\} \mapsto \{1,2\}$.

Claim 2.12. *Assume that $h_1 = abc$, $h_2 = bde$ and $\{a,b,c,d,e\}$ is a Special Block of T. Let T_a, T_b, T_c, T_d, T_e be maximal linear subtrees of T such that $V(T_x) \cap \{a,b,c,d,e\} = \{x\}$ where $x \in \{a,b,c,d,e\}$. Then, if Lemma* 2.6 *holds for each T_x, where $x \in \{a,b,c,d,e\}$ and coloring $c : \{a,b,c,d,e\} \mapsto \{1,2\}$ is given, then it holds for T as well.*

Observation 2.13. It is easy to see that $V(T_x) \cap V(T_y) = \emptyset$ for any distinct $x, y \in \{a, b, c, d, e\}$ and $\cup_{x \in \{a,b,c,d,e\}} E(T_x) \cup \{h_1, h_2\} = E(T)$.

Proof of Claim 2.12. First we show that the hypergraph induced on $V(T)$ is properly colored. Let v be a vertex which is strongly associated to a hyperedge h of T_x for some $x \in \{a, b, c, d, e\}$. If v is in T, we know that there is a hyperedge h' of T which contains v such that $|h \cap h'| = 1$. Since $V(T_x) \cap V(T_y) = \emptyset$, $h' \in E(T_x) \cup \{h_1, h_2\}$. But we showed that the vertices of $h_1 \cup h_2$ can't be strongly associated to any hyperedge of $E(T) \setminus \{h_1, h_2\}$. So $v \in V(T_x)$. Since we assumed that Lemma 2.6 holds for T_x, by using Observation 2.7 for T_x, we have that the vertices of $V(T) \setminus V(T_x)$ and the vertices of $V(T_x)$ can be colored independently. This implies that the hypergraph induced by $V(T)$ is properly colored.

Let $v \in V(H) \setminus V(T)$. First assume that v is not strongly associated to any hyperedge of T and let $p, q \in V(T)$ be arbitrary. We have to show that vpq is properly colored regardless of the color of v. If $p, q \in T_x$ for some $x \in \{a, b, c, d, e\}$ then we are done because we assumed Lemma 2.6 holds for T_x. So, let $p \in T_x$ and $q \in T_y$ for some distinct $x, y \in \{a, b, c, d, e\}$. Since both p and q can't be in S (by definition of S), the smallest linear path between p and q in T has 2 hyperedges. This linear path, together with vpq forms a linear cycle, a contradiction.

Now assume that v is strongly associated to a hyperedge of T. If v is strongly associated to hyperedges h_x, h_y of T such that $h_x \in E(T_x)$ and $h_y \in E(T_y)$, then it is easy find a linear cycle using the minimal linear path in T between h_x and h_y. This implies that there is a unique $x \in \{a, b, c, d, e\}$ such that v is strongly associated to hyperedge(s) of only T_x. As we showed in the previous paragraph if $vpq \in E(H)$ then both p and q are in T_y for some $y \in \{a, b, c, d, e\}$. If $y \neq x$, then we know that hyperedges vpq are properly colored regardless of the color of v by applying Lemma 8 to T_y. If $y = x$, then by applying Lemma 8 to T_y again, there is a coloring of v such that hyperedges vpq are properly colored, as desired. □

So applying Claim 2.12 recursively, it suffices to prove Lemma 2.6 for a linear subtree T of H which has no Special Block. So from now on, we may assume that there is no Special Block in T.

We will now construct a special graph G_T by following the steps in the *Construction* below, one after another. This graph will be connected, and its vertex set and edge set satisfy: $V(G_T) = V(T)$ and if $ab \in E(G_T)$ then there exists a vertex $x \in V(T)$ such that $abx \in E(T)$. We will then show later that this graph G_T is actually a tree and that a proper 2-coloring of G_T will give us a proper 2-coloring of the hypergraph induced on $V(T)$ as demanded by Lemma 2.6.

Construction. *1. For every two hyperedges $abc, ebd \in E(T)$, where abc is a thick hyperedge which is strongly associated to the vertex e of ebd then,*

(a) add eb to $E(G_T)$.

(b) add ac to $E(G_T)$ if $ace \in E(H)$ (note that they may have been already added).

2. For every $abc \in E(T)$, if abc is a hyperedge of T and vab is a hyperedge of H such that v is weakly associated to abc, then add ab to $E(G_T)$.

3. *For every two hyperedges $abc, ebd \in E(T)$ which are stongly associated to a vertex $v \in V(H) \setminus V(T)$, if acv (respectively edv) is a hyperedge of H, then add ac to $E(G_T)$ (respectively $ed \in E(G_T)$).*

4. *After completing the above steps, for every hyperedge $abc \in E(T)$ we do the following. If abc is thick, and less than two of the three pairs ab, bc, ca are in $E(G_T)$ we add some more pairs arbitrarily so that $E(G_T)$ has exactly two pairs from ab, bc, ca. If abc is not thick, we add pairs from ab, bc, ca in a way that the only remaining pair is not thick.*

Now we claim the following.

Claim 2.14. *G_T is a tree and so it can be properly colored.*

Before we prove the above claim, we will show that it implies Lemma 2.6.

First let us prove that the proper coloring of G_T gives us a proper coloring of the subhypergraph induced by $V(T)$. Since $V(G_T) = V(T)$, a proper coloring of G_T gives us a proper coloring of the hyperedges of T. Therefore, it suffices to prove that for every hyperedge $abc \in E(T)$, the hyperedges xyv where $x, y \in \{a, b, c\}$ and $v \in V(T) \setminus \{a, b, c\}$ are properly colored. If abc is not thick, then it is easy to see that xy (which has to be a thick pair) must be in G_T (due to point 4 in the construction of G_T) which means that x and y have different colors and so the hyperedge xyv is properly colored, as desired. If abc is thick, then v must be associated to abc. If v is weakly associated to abc, then by the construction of G_T (point 2), xy must be in G_T and so xyv is properly colored again. If v is strongly associated to abc, then v belongs to a neighboring hyperedge of abc in T. W.l.o.g assume that $vbw \in E(T)$. By the above construction of G_T, we have $bv, ac \in E(G_T)$. So b and v have different colors and a and c have different colors. Therefore, all the hyperedges vxy are properly colored. So the subhypergraph induced by $V(T)$ is properly colored.

Now let $v \in V(H) \setminus V(T)$. We will show that v satisfies the properties of Lemma 2.6. If v is not strongly associated to any hyperedge of T, then for every $xyv \in E(H)$, $xy \in E(G_T)$ and so v can be colored arbitrarily. So assume that v is strongly associated to hyperedges $h_1, h_2, \ldots, h_k$ of T. We consider two cases. If $k \geq 2$, then we claim that $|h_i \cap h_j| \neq \emptyset$ for every $i, j \in \{1, 2, \ldots, k\}$ because otherwise we can find a linear cycle using the shortest linear path between h_i and h_j and v. Since h_i are hyperedges of a linear tree, and every two of them have a common point, there is a vertex o such that $\cap_i h_i = \{o\}$. Let us use different colors for v and o. If $xy \notin h_i$ for any i, then as we saw before xyv is properly colored independent of the color of v. So $xy \in h_i$ for some i. If $o \in \{x, y\}$, then since o and v are colored differently, xyv is colored properly. If $o \notin \{x, y\}$, then by the construction of G_T (see point 3), xy is in G_T and so xyv is properly colored, as desired. So the only remaining case is if $k = 1$. In this case, the hyperedge h_1 has two vertices of the same color and if we color v differently from this color, hyperedges vxy are properly colored. This completes the proof of Lemma 2.6.

Proof of Claim 2.14. Assume by contradiction that G_T has a cycle. Since T is a linear tree, this cycle has to a triangle abc where $abc \in E(T)$ is a thick hyperedge. First observe that none of the pairs ab, bc, ca were added during point 4 of the construction of G_T. We now consider different cases for how abc could be formed.

Case 1. One of the pairs ab, bc, ca was added by Construction 1b.

W.l.o.g generality let the pair added by Construction 1b was ac. Then, there exists a hyperedge $bde \in E(T)$ such that d is strongly associated to abc and $acd \in E(H)$. So either abd or bcd is in $E(H)$. Clearly, there is no $w \notin \{a, b, c, d, e\}$ such that wab or wbc is a hyperedge of H for otherwise we have a linear cycle. So the only vertices that can be associated to abc are d and e. Since abc is thick, ab, bc are thick pairs. If either bce or abe is in $E(H)$, then the conditions of Hall's theorem hold for the bipartite graph whose color classes are $\{ab, bc\}$ and $\{d, e\}$ where $xy \in \{ab, bc\}$ is connected to $z \in \{d, e\}$ if and only if $xyz \in E(H)$. So there is a matching and by Claim 2.10, we have a contradiction since we assumed there is no Special Block of T. So assume that $bce, abe \notin E(H)$. So the only hyperedges (besides abc) containing ab and bc are abd and bcd which implies that ab and bc were not added by Construction 1b, 2 and 3. So both ab and bc were added by Construction 1a. Assume that bc was added because either b or c was strongly associated to a hyperedge h'. This means that h' is thick and $h' = dbe$ because otherwise we have $wbc \in E(H)$ for some $w \notin \{a, b, c, d, e\}$, a contradiction. So c is strongly associated to bde. Similarly, a is strongly associated to bde. So by Claim 2.11, $\{a, b, c, d, e\}$ is a Special Block, a contradiction.

So from now on, we can assume that Construction 1b was never used to add the pairs ab, bc, ca.

Case 2. One of the pairs ab, bc, ca was added by Construction 3.

W.l.o.g let us say ac was added by Construction 3. Then, there is a hyperedge $bde \in E(T)$ and $v \in V(H) \setminus V(T)$ such that v is strongly associated to both hyperedges abc, bed and $acv \in E(H)$. Since ab is a thick-pair, there is a vertex $w \notin \{a, b, c\}$ such that $abw \in E(H)$. If $w \notin \{a, b, c, d, e, v\}$ then since $acv, wab \in E(H)$ and one of $bev, bdv \in E(H)$, they form a linear cycle, a contradiction. If $w = e$, then since $abe, acv \in E(H)$ and one of $bdv, dev \in E(H)$, we have a linear cycle again, a contradiction. Similarly $w \neq d$. Therefore, $w = v$. So the only hyperedge besides abc which contains ab, is abv. Similarly, the only hyperedge besides abc which contains bc is bcv. This implies that ab and bc were not added by Construction 1, 2 and 4. Also, it's easy to see that they were not added by Construction 3, otherwise v would have been strongly associated to a hyperedge of T which is not a neighbor of ebd, which is a contradiction.

So the only reminaing case is when ab, bc, ca are added by Construction 1a or 2.

Case 3. ab, bc, ca were added by Construction 1a or 2.

Two of the pairs ab, bc, ca cannot be added by Construction 2 due to Lemma 2.5. Therefore, we have two subcases: Either exactly one of ab, bc, ca was added by Construction 2 and the other two were added by Construction 1a or all of them were added by Construction 1a.

Assume that all of the pairs ab, bc, ca were added by Construction 1a. Let $xy \in \{ab, bc, ca\}$. Let us say xy was added because there is a thick hyperedge $h_{xy} \in E(T)$ which is strongly associated to either x or y. If any two of the there hyperedges h_{ab}, h_{bc}, h_{ca} are the same, then by Claim 2.11, we have a Special Block in T, a contradiction. Therefore, $h_{ab} \neq h_{bc} \neq h_{ca}$. But then, we have hyperedges $abv_1, acv_2, bcv_3 \in E(H)$ where $v_1 \in h_{ab}, v_2 \in h_{bc}, v_3 \in h_{ac}$ which form a linear cycle, a contradiction.

Now assume that one of the pairs ab, bc, ca was added by Construction 2 and the other two were added by Construction 1a. W.l.o.g assume that ab and bc were added by Construction 1a and ca by Construction 2. Let us say ab (respectively bc) was added because there is a thick hyperedge $h_{ab} \in E(T)$ (respectively $h_{bc} \in E(T)$) which is strongly associated to either a or b (respectively b or c). So there are vertices $v_1 \in h_{ab}$ and $v_2 \in h_{bc}$ such that $abv_1, bcv_2 \in E(H)$. If $h_{ab} = h_{bc}$, then by Claim 2.11 we have a Special Block in T, a contradiction. So $h_{ab} \neq h_{bc}$. Let us say ac was added because there is a vertex w weakly associated to abc such that $wac \in E(H)$. If $w \neq v_1$ and $w \neq v_2$, then we have a linear cycle, namely acw, abv_1, bcv_2, a contradiction. So let us assume w.l.o.g that $w = v_1$. Let $h_{ab} = v_1ex$ where x is either a or b. If $x = b$, then h_{ab}, v_1ac, bcv_2 is a linear cycle, a contradiction. If $x = a$, then clearly b is strongly associated to $h_{ab} = v_1xe$. So either the hyperedge $abe \in E(H)$ or $bev_1 \in E(H)$. This hyperedge together with acv_1 and bcv_2 gives us a linear cycle, a contradiction. □

2.3 Proof of Theorem 2.4

Let H be a 3-uniform hypergraph without any linear cycles. First let us assume that there are no vertices $u, v \in V(H)$ such that for every $x \in V(H) \setminus \{u, v\}$, $uvx \in E(H)$ and show that Theorem 2.4 holds in this case whenever $|V(H)| \geq 6$.

We distinguish some cases.

Case 1. *There are no vertices $u, v \in V(H)$ such that $uvx \in E(H)$ for every $x \in V(H)$ and $|V(H)| \geq 6$.*

Let $\mathcal{P} = \{p_0q_0p_1, p_1q_1p_2, p_2q_2p_3, \ldots, p_{k-1}q_{k-1}p_k\}$ be a longest linear path of H such that $p_0q_0p_1$ is the first and $p_{k-1}q_{k-1}p_k$ is the last hyperedge of the path. Consider a skeleton containing $\mathcal{P}$. The set of hyperedges of this skeleton incident on p_1 (respectively p_{k-1}) except $p_1q_1p_2$ is called as a *windmill* at p_1 (respectively p_{k-1}) and the size of this set is called the size of the windmill. Thus there are two windmills corresponding to $\mathcal{P}$ and the skeleton containing it. Among all the skeletons of maximum size which contain $\mathcal{P}$, let us take a skeleton T such that the size of the smaller windmill is minimum. W.l.o.g. we may assume that the smaller windmill is at p_1.

Lemma 2.15. *Any hyperedge $abc \in E(T)$ is strongly associated to at most one vertex of $V(H) \setminus V(T)$.*

Proof. Suppose by contradiction that $abc \in E(T)$ is strongly associated to two vertices $v_1, v_2 \in V(H) \setminus V(T)$. Consider the bipartite graph whose color classes are $\{v_1, v_2\}$ and $\{ab, bc, ca\}$ where $v \in \{v_1, v_2\}$ and $xy \in \{ab, bc, ca\}$ are adjacent iff $vxy \in E(H)$. Then it can be easily seen that there is a matching saturating $\{v_1, v_2\}$ between the two color classes. If we replace abc by the two hyperedges corresponding to this matching we will get a skeleton of bigger size contradicting the fact that T has maximum size. □

We have the following corollary of the above lemma.

Corollary 2.16. *Let $|V(H) \setminus V(T)| = t$ and let degree of $v \in V(T)$ in the subhypergraph of H induced by $V(T)$ be $d_T(v)$. Then the degree of any vertex $v \in V(T)$ which is in exactly one hyperedge of T, is at most $d_T(v) + t + 1$.*

Let us call the subtree of T which contains the hyperedges of T incident to v as a *star* of T at $v \in V(T)$. Considering the pairs covered by the hyperedges of T as a graph $G(T)$, for any $v \in V(T)$ the pairs $\{x, y\}$ that are at equal distance from v in $G(T)$ are called pairs opposite to v. Clearly, every hyperedge of T has exactly one pair opposite to v. We have the following simple lemmas which are stated without proofs.

Lemma 2.17. *Let $v \in V(T)$ and $vab \in E(H)$ be such that $\{a, b\}$ does not intersect the star at $v \in V(T)$. Then $\{a, b\}$ is a pair opposite to v in T.*

Lemma 2.18. *Let $p_0q_0x \in E(H)$ and let us consider the linear path between x and p_0. Let $\mathcal{P}'$ be the subpath of this linear path without the starting and ending hyperedges (i.e., not including the two hyperedges which contain p_0 and x). Then, for any $y, z \in V(\mathcal{P}')$, we have $p_0yz \notin E(H)$.*

Case 1.1. *The size of the smaller windmill is at least* 2.

We will show that the degree of p_0 is at most $n - 2$. If x is in $V(T) \setminus \{p_1, p_0, q_0\}$, then we claim that $p_0q_0x \notin E(H)$ because if x is in the windmill around p_1 then the linear path $\mathcal{P}$ can be extended. If x is not in the windmill around p_1 then by replacing the hyperedge $p_0q_0p_1$ with p_0q_0x will decrease the size of the smaller windmill contradicting the assumption that the size of the smaller windmill is minimum.

The hyperedges containing p_0 are of the following two types. We will count them separately.

First, let us count the number of hyperedges of the type p_0p_1x where $x \in V(T) \setminus \{q_0\}$. Since p_0p_1 can't be opposite to any $x \in V(T) \setminus \{q_0\}$, by Lemma 2.17, p_0p_1 must intersect the star at x. This means that x should be contained in the star at p_1. So the number of hyperedges of the type p_0p_1x where $x \in V(T) \setminus \{q_0\}$ is $2w_1$ where w_1 is the size of the windmill at p_1. Let w_2 be the size of the windmill at p_{k-1} (So $w_1 \leq w_2$).

Now, let us count the number of hyperedges of the type p_0xy where $x, y \in V(T) \setminus \{p_1, q_0\}$. Since xy doesn't intersect the star at p_0, by Lemma 2.17, xy is opposite to p_0. If xy is a pair of the hyperedge of either windmill then we can extend $\mathcal{P}$ by p_0xy, a contradiction. So the number of such xy pairs is at most $\frac{V(T)-(2w_1+1)-2w_2}{2} = \frac{(n-t)-(2w_1+1)-2w_2}{2}$.

Then the total degree of p_0 in the subhypergraph induced by $V(T)$,

$$d_T(p_0) \leq 1 + 2w_1 + \frac{(n-t) - (2w_1 + 1) - 2w_2}{2}.$$

Thus by Corollary 2.16, the degree of p_0 is at most

$$1 + 2w_1 + \frac{(n-t) - (2w_1 + 1) - 2w_2}{2} + t + 1 = \frac{n + t + 2w_1 - 2w_2 + 3}{2} \leq \frac{n + t + 3}{2}.$$

So we are done unless $\frac{n+t+3}{2} \geq n - 1$, which simplifies to $n - t = |V(T)| \leq 5$ and this is considered in Case 1.3.

Case 1.2. *The size of the smaller windmill is* 1.

There are three types of hyperedges in H that contain p_0: hyperedges of the type p_0q_0x, p_0yz and p_0p_1w. We always consider the hyperedge $p_0q_0p_1$ as of the type p_0q_0x. Let r be the number of hyperedges in H of the type p_0q_0x where $x \in V(H) \setminus \{p_0, q_0\}$ and let s be the number of hyperedges in H of the type p_0yz where $y, z \in V(H) \setminus \{p_0, q_0, p_1\}$.

Lemma 2.19. *$r + s \leq n - 2$ and if equality holds then $p_0p_kq_{k-1} \in E(H)$.*

Proof. First we claim that $r + s \leq n - s$. Since $\{y, z\}$ doesn't intersect the star at p_0, by Lemma 2.17, the pair $\{y, z\}$ is opposite to p_0. We claim that if $p_0yz \in E(H)$ then the pair $\{y, z\}$ must be contained in the linear path $\mathcal{P}$. It is easy to see that since $\{y, z\}$ is opposite to p_0, either both y and z are contained in $\mathcal{P}$ or both of them are not in $\mathcal{P}$. In the latter case, $\mathcal{P}$ can be extended by adding the hyperedge p_0yz, contradicting the maximality of $\mathcal{P}$.

Now consider the pair $\{y_1, z_1\}$ closest to p_0 such that $p_0y_1z_1 \in E(H)$. By Lemma 2.18, the farthest $x \in \mathcal{P}$ from p_0 such that $p_0q_0x \in E(H)$ can be either y_1 or z_1 but no later. This means that every vertex in $V(H) \setminus \{p_0, q_0\}$ belongs to at most one hyperedge of the type p_0q_0x or p_0yz except y_1, z_1. So $r + 2s \leq n - 2 + 2 = n$, as desired.

Since $r + s \leq n - s$, we are done if $s \geq 2$ and so we can assume $s \leq 1$. By our assumption that there are no vertices $u, v \in V(H)$ such that for every $x \in V(H) \setminus \{u, v\}$, $uvx \in E(H)$, we have $r \leq n - 3$. So, $r + s \leq n - 3 + 1 = n - 2$, as desired. If $r + s = n - 2$, then we must have $s \geq 1$. That is, there exists an edge of the type p_0yz where $y, z \in V(H) \setminus \{p_0, q_0, p_1\}$. Since the pair $\{y, z\}$ must be opposite to p_0 and is contained in $\mathcal{P}$, if $\{y, z\} \neq p_kq_{k-1}$ then by Lemma 2.18, $p_0q_0p_k, p_0q_0q_{k-1} \notin E(H)$. So the vertices p_k, q_{k-1} do not belong to a hyperedge of the type p_0q_0x or p_0yz. So, by the same argument as before, $r + 2s \leq n - 4 + 2 = n - 2$ which is a contradiction since we assumed $r + s = n - 2$ and $s \geq 1$. □

Case 1.2.1. *There is a hyperedge of type $p_0q_0x \in E(H)$ where $x \in V(T) \setminus \{p_0, p_1, p_2, q_0, q_1\}$.*

In this case, we claim that number of hyperedges of the type p_0p_1y in H where $y \in V(H) \setminus \{p_0, q_0, p_1\}$ is at most 1 and if such a hyperedge exists then y is either p_2 or q_1. Assume by contradiction that $p_0p_1y' \in E(H)$ where $y' \neq q_0$. Let $\mathcal{P}_1$ be a linear path in T between (and including) x and p_1. If $y' \notin \mathcal{P}_1$, then p_0q_0x, p_0p_1y' and $\mathcal{P}_1$ form a linear cycle. So $y' \in \mathcal{P}_1$. Since $\{p_0, p_1\}$ cannot be an opposite pair of any vertex on $\mathcal{P}_1$ except q_0, by Lemma 2.17, $\{p_0, p_1\}$ must intersect the star at y'. So y' is either p_2 or q_1. If both hyperedges $p_0p_1p_2$ and $p_0p_1q_1$ are in H then p_0q_0x, $\mathcal{P}_1 \setminus \{p_1p_2q_1\}$ and one of these two hyperedges form a linear cycle. Therefore the desired claim follows.

If both hyperedges $p_0p_1p_2$, $p_0p_1q_1$ are not in H, the degree of p_0 is $r + s$ and by Lemma 2.19, $r + s \leq n - 2$ and so Theorem 2.4 holds. Therefore, from now on, we may assume that exactly one of the two hyperedges $p_0p_1p_2$, $p_0p_1q_1$ is in H. If $r + s$ is strictly less than $n - 2$ then degree of p_0 is at most $n - 2$ and Theorem 2.4 holds again. So we also assume that $r + s = n - 2$. By Lemma 2.19 if $r + s = n - 2$, then $p_0p_kq_{k-1} \in E(H)$. It follows that the size of the windmill at p_{k-1} is 1 because if it is more than 1, then the linear path $\mathcal{P}$ can be extended by adding $p_0p_kq_{k-1}$ to it. Therefore the size of the windmills at p_{k-1} and p_1 are both 1. By symmetry, if we define r' and s' for p_k as we defined r and s for p_0, Lemma 2.19

holds for them. Since a hyperedge of the type $p_kq_{k-1}x$ exists, namely $p_kq_{k-1}p_0$, by the same argument as before we can assume that $r' + s' = n - 2$ and so $p_0q_0p_k \in E(H)$. By Lemma 2.18 for p_k, it is easy to see that $s' \leq 1$. So $r' \geq n - 3$. We know that $p_0p_1y \in E(H)$ where y is either p_2 or q_1. Now $p_0q_0p_k$, p_0p_1y and either $p_kq_{k-1}y$ or $p_kq_{k-1}p_1$ (one of them exists because $r' \geq n - 3$) form a linear cycle, a contradiction.

Case 1.2.2. *There is no hyperedge of type $p_0q_0x \in E(H)$ where $x \in V(T) \setminus \{p_0, p_1, p_2, q_0, q_1\}$.*

Let d_0 be the degree of p_0 in the subhypergraph of H induced by $\{p_0, p_1, p_2, q_0, q_1\}$. Clearly $d_0 \leq 6$. If $p_kq_{k-1}p_0 \in E(H)$ (so the size of the windmill at p_{k-1} is 1), then by symmetry (by looking at p_k instead of p_0) we are done by the previous case. So we can assume that $p_kq_{k-1}p_0 \notin E(H)$.

If there is a vertex $v \in V(H) \setminus V(T)$ which is strongly associated to $p_0q_0p_1$, then we claim that $d_0 \leq 4$ because if either $p_0q_0p_2$ or $p_0q_0q_1$ is in H, then it is easy to check that we have a linear cycle. Let $|V(H) \setminus V(T)| = t$. So the degree of p_0 in the subhypergraph of H induced by T, $d_T(p_0) \leq d_0 + \frac{n-t-7}{2}$ (here we used $p_kq_{k-1}p_0 \notin E(H)$). By Corollary 2.16, degree of p_0 is at most

$$d_0 + \frac{n-t-7}{2} + t + 1 \leq \frac{n+t+3}{2}.$$

Then, Theorem 2.4 holds unless $\frac{n+t+3}{2} \geq n - 1$ which simplifies to $n - t \leq 5$ and this is considered in Case 1.3.

If there is no vertex $v \in V(H) \setminus V(T)$ which is strongly associated to $p_0q_0p_1$, then degree of p_0 is at most $d_T(p_0) + t$. And, $d_T(p_0) \leq d_0 + \frac{n-t-7}{2}$. So, degree of p_0 is at most

$$d_0 + \frac{n-t-7}{2} + t \leq d_0 + \frac{n+t-7}{2},$$

and Theorem 2.4 holds unless $d_0 + \frac{n+t-7}{2} \geq n - 1$ which simplifies to $d_0 \geq \frac{n-t+5}{2}$. If $n - t > 7$ then $d_0 > 6$ which is impossible. So we may assume $n - t \leq 7$. The case $n - t \leq 5$ is considered in Case 1.3. Since $n - t$ is odd (the number of vertices in the skeleton is odd) we only have to deal with the case when $n - t = 7$. In this case the size of the skeleton T is 3 and since the size of the smaller windmill is 1, T consists only of a linear path of size 3. In this case, $d_0 = 6$. By the same argument, the degree of q_0 in the subhypergraph induced by $\{p_0, p_1, p_2, q_0, q_1\}$ is 6. By symmetry the degree of p_3 in the subhypergraph of H induced by $\{p_3, q_2, p_2, q_1, p_1\}$ is also 6 and so $p_0p_1q_1, q_0p_1p_2, p_3p_2q_1 \in E(H)$ form a linear cycle, a contradiction.

Case 1.3. $|V(T)| \leq 5$ *for a skeleton T of H.*

Let T be a skeleton of H where $|V(T)| \leq 5$ and we want to show that Thoerem 2.4 holds. Since $|V(T)|$ is odd, either $|V(T)| = 3$ or $|V(T)| = 5$.

First assume $|V(T)| = 3$ and let T consist of one hyperedge abc. Consider the trace graph G_a where $\{x, y\} \in E(G_a)$ if and only if $axy \in E(H)$. Now notice that if there are two edges $pq, rs \in E(G_a)$ that are disjoint then $apq, ars \in E(H)$ form a skeleton on 5 vertices, a contradiction. So every two edges of G_a have a common vertex. It is easy to see that the set of edges of such a graph is either a star (a graph where all the edges have a common vertex) or a triangle. Notice that there may be some isolated vertices in the graph. Since $|V(G_a)| = |V(H)| - 1 \geq 5$, we have $|E(G_a)| \leq |V(G_a)| - 1$. So the degree of a in H is $|E(G_a)| \leq |V(G_a)| - 1 = |V(H)| - 2$ as desired.

Now let $|V(T)| = 5$ and $E(T) = \{a_1a_2b, c_1c_2b\}$. Since $V(H) \geq 6$, $|V(H) \setminus V(T)| \neq \emptyset$ We consider two cases.

Case 1.3.1. *There is no vertex in $V(H)\setminus V(T)$ which is strongly associated to any hyperedge of T.*

Since H is connected and T is a skeleton, there must be an edge xyv where $x, y \in V(T)$ and $v \in V(H)\setminus V(T)$. By assumption we know that v is not strongly associated to any edge of T. So the degree of v in the subhypergraph induced on $V(T) \cup \{v\}$ is at most 2. Now consider the trace graph G_v on the vertex set $V(G_v) := V(H) \setminus (V(T) \cup \{v\})$ where $ab \in E(G_v)$ if and only if $abv \in E(H)$. Now notice that if there are two edges $pq, rs \in E(G_v)$ that are disjoint then $vpq, vrs, vxy \in E(H)$ form a skeleton on 7 vertices, a contradiction. So every two edges of G_v have a common vertex and so $E(G_v)$ is either a triangle or a star. In either case, $|E(G_v)| \leq |V(G_v)|$. So the degree of v is at most $2+|E(G_v)| \leq 2+|V(G_v)| = 2+n-6 = n-4$.

Case 1.3.2. *There is a vertex $v \in V(H) \setminus V(T)$ which is strongly associated to a hyperedge of T.*

Assume without loss of generality that v is strongly associated to a_1a_2b. So $vba_i \in E(H)$ for some $i \in \{1, 2\}$ which implies that there is no hyperedge vxy with $x, y \in V(H) \setminus (V(T) \cup \{v\})$ because otherwise vxy, vba_i, bc_1c_2 form a skeleton on 7 vertices. If v is not strongly associated to bc_1c_2, then the degree of v is at most $1+3 = 4$ and we are done since we assumed $V(H) \geq 6$. Therefore, we may assume v is strongly associated to bc_1c_2 and so $vbc_j \in E(H)$ for some $j \in \{1, 2\}$. Now it is easy to see that there are no hyperedges $a_1a_2c_k$ and $c_1c_2a_k$ for any $k \in \{1, 2\}$ because otherwise we have a linear cycle. If any of the vertices $\{a_1, a_2, c_1, c_2\}$ have degree at most 2 in the subhypergraph induced by $V(T)$, then by Corollary 2.16, the degree of this vertex in H is at most $2 + t + 1 = t + 3$ where $V(H) \setminus V(T) = t$. and we are done because $V(H) = t + 5$. So we may assume that all of the vertices $\{a_1, a_2, c_1, c_2\}$ have degree at least 3 in the subhypergraph induced by $V(T)$. It is easy to see that the only way this degree condition is met for the vertex a_i is if $a_ibc_1, a_ibc_2 \in E(H)$ for each $i \in \{1, 2\}$. This implies that $a_1a_2v, c_1c_2v \notin E(H)$ because otherwise we have a linear cycle. So the degree of v is at most 4 and we are done because $V(H) \geq 6$.

Case 2. *If there are vertices $u, v \in V(H)$ such that $uvx \in E(H)$ for every $x \in V(H)$.*

If we assume by contradiction that Theorem 2.4 does not hold, then by the previous section, whenever $|V(H)| \geq 6$ we know that there are vertices $u, v \in V(H)$ such that $uvx \in E(H)$ for every $x \in V(H)$.

Lemma 2.20. *Let H be a 3-uniform linear cycle free hypergraph. If the degree of every vertex in H is at least $V(H) - 1$ where $|V(H)| \geq 6$, then there is a subhypergraph H_0 where degree of every vertex in H_0 is at least $V(H_0) + 1$ and $V(H_0) = V(H) - 4$.*

Proof. Let $|V(H)| = n$. Let $u, v \in V(H)$ such that $uvw \in E(H)$ for every $w \in V(H)$. Since degree of u is at least $n-1$, there must a hyperedge xyu where $x, y \in V(H) \setminus \{u, v\}$. If there is a hyperedge $xab \in E(H)$ where $a, b \in V(H) \setminus \{u, v, x, y\}$, then the hyperedges, uva, xab and xyu form a linear cycle, a contradiction. Therefore, $y \in \{a, b\}$.

Consider the trace graph $G_{u,v}$ where $\{p, q\} \in E(G_{u,v})$ if and only if either $pqu \in E(H)$ or $pqv \in E(H)$. Let the degree of x in $G_{u,v}$ be d and let the corresponding edges be $xy_1, xy_2, \dots, xy_d$. If $d \geq 2$ and $xy_iu, xy_jv \in E(H)$ where $i \neq j$, then xy_iu, xy_jv and uva where $a \in \{u, v, y_i, y_j, x\}$ form a linear cycle. So if $d \geq 2$, then either $xy_iu \in E(H)$ for every $1 \leq i \leq d$ or $xy_iv \in E(H)$ for every $1 \leq i \leq d$. W.l.o.g assume the former. Consider the case when $d \geq 3$. We know that if $xab \in E(H)$ with $a, b \in V(H) \setminus \{u, v, x\}$ then $y_i \in \{a, b\}$. So it follows that $y_1, y_2, \dots, y_d \in \{a, b\}$, which is impossible when $d \geq 3$. Therefore, $xab \notin E(H)$ where $a, b \in V(H) \setminus \{u, v, x\}$ and so the degree of x is $d+1 \leq n-3+1 = n-2$, a contradiction. Now consider the case when $d = 2$. In this case, we can have $xy_1y_2 \in E(H)$. So the degree of x is at most $d + 2 = 4$ a contradiction since $n \geq 6$. Therefore we conclude that $d = 1$. In this case we claim that $xy_1a \in E(H)$ for every $a \in V(H) \setminus \{x, y_1\}$ because otherwise degree of x is at most $n - 2$. Let the subhypergraph induced by $V(H) \setminus \{u, v, x, y_1\}$ be H_0. It is easy to see that if $abu \in E(H)$ for $a, b \in V(H_0)$ then the hyperedges abu, uvx, xya form a linear cycle, a contradiction. Similarly, $abv, abx, aby \notin E(H)$. So the degree of a vertex in H_0 is at least $n - 1 - 2 = V(H_0) + 1$, as desired. □

Actually, we will use the following simple corollary obtained by repeated applications of the lemma above.

Corollary 2.21. *If H_l is a subhypergraph of H where degree of each vertex in $V(H_l)$ is at least $|V(H_l)| + n_l$, where $n_l \geq -1$, then it has a subhypergraph H_{l+1} such that the degree of every vertex in $V(H_{l+1})$ is at least $|V(H_{l+1})| + n_{l+1}$, where $n_{l+1} = n_l + 2$ and $|V(H_{l+1})| = |V(H_l)| - 4$.*

Assume by contradiction that Theorem 2.4 does not hold. That is, there is a hypergraph $H := H_1$ on n vertices where degree of every vertex is at least $n - 1$ and $n \geq 10$. Then by using Corollary 2.21, there is an l such that $|V(H_l)| \leq 5$ and the degree of every vertex in H_l is at least $|V(H_l)| + 3$ (notice that since $n \geq 10$, we must have $l \geq 3$), which is impossible.

2.4 Proof of Theorem 2.5

We call a hypergraph *mixed* if it can contain hyperedges of both sizes 2 and 3. A linear cycle in a mixed hypergraph is still defined according to Definition 2.1. We will in fact prove our theorem for mixed hypergraphs (which is clearly a bigger class of hypergraphs than 3-uniform hypergraphs). More precisely, we will prove the following stronger theorem.

Theorem 2.22. *If H is a mixed hypergraph, then its vertex set $V(H)$ can be covered by at most $\alpha(H)$ edge-disjoint linear cycles (where we accept a single vertex or a hyperedge as a linear cycle).*

Proof. We prove the theorem by induction on $\alpha(H)$. If $|V(H)| = 1$ or 2, then the statement is trivial. If $|V(H)| \geq 3$ and $\alpha(H) = 1$, then H contains all possible edges of size 2 and there is a Hamiltonian cycle consisting only of edges of size 2, which is of course a linear cycle covering $V(H)$.

Let $\alpha(H) > 1$. If $E(H) = \emptyset$, then $\alpha(H) = V(H)$ and the statement of our theorem holds trivially since we accept each vertex as a linear cycle. If $E(H) \neq \emptyset$, then let P be a longest linear path in H consisting of hyperedges $h_0, h_1, \ldots, h_l$ ($l \geq 0$). If h_i is of size 3, then let $h_i = v_i v_{i+1} u_{i+1}$ and if it is of size 2, then let $h_i = v_i v_{i+1}$. A linear subpath of P starting at v_0 (i.e., a path consisting of hyperedges $h_0, h_1, \ldots, h_j$ for some $j \leq l$) is called an *initial segment* of P. Let C be a linear cycle in H which contains the longest initial segment of P. If there is no linear cycle containing h_0, then we simply let $C = h_0$.

Let us denote the subhypergraph of H induced on $V(H) \backslash V(C)$ by $H \backslash C$. Let $R = \{v_k u_k \mid \{v_k, u_k\} \subseteq V(P) \setminus V(C) \text{ and } v_0 v_k u_k \in E(H)\}$ be the set of *red* edges. Let us construct a new hypergraph H' where $V(H') = V(H) \setminus V(C)$ and $E(H') = E(H \setminus C) \cup R$. We will show that $\alpha(H') < \alpha(H)$ and any linear cycle cover of H' can be extended to a linear cycle cover of H by adding C and extending the red edges by v_0.

The following claim shows that the independence number of H' is smaller than the independence number of H. This fact will later allow us to apply induction.

Claim 2.23. *If I is an independent set in H', then $I \cup v_0$ is an independent set in H.*

Proof. Suppose by contradiction that $h \subseteq (I \cup v_0)$ for some $h \in E(H)$. Then, clearly $v_0 \in h$ because otherwise I is not an independent set in H'. Now let us consider different cases depending on the size of $h \cap (V(P) \setminus V(C))$. If $|h \cap (V(P) \setminus V(C))| = 0$ then, by adding h to P, we can produce a longer path than P, a contradiction. If $|h \cap (V(P) \setminus V(C))| = 1$, let $h \cap (V(P) \setminus V(C)) = \{x\}$. Then the linear subpath of P between v_0 and x together with h forms a linear cycle which contains a larger initial segment of P than C, a contradiction. If $|h \cap (V(P) \setminus V(C))| = 2$, then let $h \cap (V(P) \setminus V(C)) = \{x, y\}$. Let us take smallest i and j such that $x \in h_i$ and $y \in h_j$ (i.e., if $x \in h_i \cap h_{i+1}$ then let us take h_i). If $i \neq j$, say $i < j$ without loss of generality, then the linear subpath of P between v_0 and x together with h forms a linear cycle with longer initial segment of P than C, a contradiction. Therefore, $i = j$ but in this case, $\{x, y\}$ is a red edge and so at most one of them can be contained in I, contradicting the assumption that $h = v_0 xy \subseteq (I \cup v_0)$. Hence, $I \cup v_0$ is an independent set in H, as desired. □

The following claim will allow us to construct linear cycles in H from red edges.

Claim 2.24. *The set of hyperedges of every linear cycle in H' contains at most one red edge.*

Proof. Suppose by contradiction that there is a linear cycle C' in H' containing at least two hyperedges which are red edges. Then there is a linear subpath P' of C' consisting of hyperedges $h'_0, h'_1, \ldots, h'_m$ such that $h'_0 := v_s u_s$ and $h'_m := v_t u_t$ (where $s > t$) are red edges but h'_k is not a red edge for any $1 \leq k \leq m-1$. Let us first take the smallest i such that $V(P') \cap h_i \neq \emptyset$ and then the smallest j such that $h'_j \cap h_i \neq \emptyset$. It is easy to see that $|V(P') \cap h_i| \leq 2$ (since i was smallest). If $\left|h'_j \cap h_i\right| = 1$, then the linear cycle consisting of hyperedges $h'_1, \ldots, h'_j$ and $h_i, h_{i-1}, \ldots, h_0$ and $v_0 v_s u_s$ contains a larger initial segment of P than C (as $h'_j \cap h_i \in V(P) \setminus V(C)$), a contradiction. If $\left|h'_j \cap h_i\right| = 2$, then notice that $\left|h'_{j+1} \cap h_i\right| = 1$. Now the linear cycle consisting of the hyperedges $h'_{m-1}, h'_{m-2}, \ldots, h'_{j+1}$ and $h_i, h_{i-1}, \ldots, h_0$ and $v_0 v_t u_t$ contains a larger initial segment of P than C, a contradiction. □

By Claim 2.23, $\alpha(H') \leq \alpha(H) - 1$. So by induction hypothesis, $V(H')$ can be covered by at most $\alpha(H) - 1$ edge-disjoint linear cycles (where we accept a single vertex or a hyperedge as a linear cycle). Now let us replace each red edge $\{x, y\}$ with the hyperedge xyv_0 of H. Claim 2.24 ensures that in each of these linear cycles, at most one of the hyperedges is a red edge. Therefore, it is easy to see that after the above replacement, linear cycles of H' remain as *linear* cycles in H and they cover $V(H') = V(H) \setminus V(C)$. Now the linear cycle C, together with these linear cycles give us at most $\alpha(H) - 1 + 1 = \alpha(H)$ edge-disjoint linear cycles covering $V(H)$, completing the proof. □

Chapter 3

Asymptotics for Turán numbers of cycles in 3-uniform hypergraphs

3.1 Introduction

Recall that a Berge cycle of length $k \geq 2$, denoted by C_k, is an alternating sequence of distinct vertices and distinct edges of the form $v_1, h_1, v_2, h_2, \ldots, v_k, h_k$ where $v_i, v_{i+1} \in h_i$ for each $i \in \{1, 2, \ldots, k-1\}$ and $v_k, v_1 \in h_k$. (Note that if a hypergraph does not contain a Berge-C_2, then it is linear.) This definition of a hypergraph cycle is the classical definition due to Berge. More generally, if $F = (V(F), E(F))$ is a graph and $\mathcal{Q} = (V(\mathcal{Q}), E(\mathcal{Q}))$ is a hypergraph, then we say $\mathcal{Q}$ is *Berge-F* if there is a bijection $\phi : E(F) \to E(\mathcal{Q})$ such that $e \subseteq \phi(e)$ for all $e \in E(F)$. In other words, given a graph F we can obtain a Berge-F by replacing each edge of F with a hyperedge that contains it.

The systematic study of the Turán numbers of Berge cycles started with the study of Berge triangles by Győri [47], and continued with the study of Berge five cycles by Bollobás and Győri [9] who showed the following.

Theorem 3.1 (Bollobás, Győri [9]). *We have,*

$$(1 + o(1))\frac{n^{3/2}}{3\sqrt{3}} \leq ex_3(n, C_5) \leq \sqrt{2}n^{3/2} + 4.5n.$$

The following construction of Bollobás and Győri shows the lower bound in Theorem 3.1.

Bollobás-Győri Example. Take a C_4-free bipartite graph G_0 with $n/3$ vertices in each part and $(1 + o(1))(n/3)^{3/2}$ edges. In one part, replace each vertex u of G_0 by a pair of two new vertices u_1 and u_2, and add the triple u_1u_2v for each edge uv of G_0. It is easy to check that the resulting hypergraph H does not contain a Berge cycle of length 5. Moreover, the number of hyperedges in H is the same as the number of edges in G_0.

In this chapter, we improve Theorem 3.1 as follows.

Theorem 3.2 (E., Győri, Methuku [29]). *We have,*

$$\mathrm{ex}_3(n, C_5) < 0.254n^{3/2} + O(n).$$

Roughly speaking, our proof idea is to analyze the structure of Berge-C_5-free hypergraphs, and use this structure to efficiently limit the number of paths of length 3 in the 2-shadow; which is then combined with the lower bound on the number of paths of length 3 provided by the Blakley-Roy inequality [7].

Naturally, we also consider forbidding Berge k-cycles in linear hypergraphs. For $k \geq 2$, Füredi and Özkahya [40] showed $\mathrm{ex}_3^{\mathrm{lin}}(n, C_{2k+1}) \leq 2kn^{1+1/k} + 9kn$. In fact it is shown in [46, 40] that $\mathrm{ex}_3(n, C_{2k+1}) \leq O(n^{1+1/k})$. For the even case it is easy to show $\mathrm{ex}_3^{\mathrm{lin}}(n, C_{2k}) \leq \mathrm{ex}(n, C_{2k}) = O(n^{1+1/k})$ by selecting a pair from each hyperedge of a C_{2k}-free 3-uniform linear hypergraph. A (Berge) path of length k is an alternating sequence of distinct vertices and distinct edges of the form $v_0, h_0, v_1, h_1, v_2, h_2, \ldots, v_{k-1}, h_{k-1}, v_k$ where $v_i, v_{i+1} \in h_i$ for each $i \in \{0, 1, 2, \ldots, k-1\}$. Below we concentrate on the linear Turán numbers of C_3, C_4 and C_5.

As discussed in Chapter 1, determining $\mathrm{ex}_3^{\mathrm{lin}}(n, C_3)$ was settled by Ruzsa and Szemerédi [68], showing that $n^{2-\frac{c}{\sqrt{\log n}}} < \mathrm{ex}_3^{\mathrm{lin}}(n, C_3) = o(n^2)$ for some constant $c > 0$.

Only a handful of results are known about the asymptotic behaviour of Turán numbers for hypergraphs. In this chapter, we focus on determining the asymptotics of $\mathrm{ex}_3^{\mathrm{lin}}(n, C_5)$ by giving a new construction, and a new proof of the upper bound which introduces some important ideas. We also determine the asymptotics of $\mathrm{ex}_3^{\mathrm{lin}}(n, C_4)$ and construct 3-uniform linear hypergraphs avoiding linear cycles of given odd length(s).

The following is one of the main results in this chapter.

Theorem 3.3 (E., Győri, Methuku [30]).

$$\mathrm{ex}_3^{\mathrm{lin}}(n, C_5) = \frac{1}{3\sqrt{3}}n^{3/2} + O(n).$$

To show the lower bound in the above theorem we give the following construction. For the sake of convenience we usually drop floors and ceilings of various quantities in the construction below, and in the rest of the chapter, as it does not effect the asymptotics.

Construction of a C_5-free linear hypergraph H: For each $1 \leq t \leq \sqrt{n/3}$, let $L_t = \{l_1^t, l_2^t, \ldots, l_{\sqrt{n/3}}^t\}$ and $R_t = \{r_1^t, r_2^t, \ldots, r_{\sqrt{n/3}}^t\}$. Let $B = \{v_{i,j} \mid 1 \leq i, j \leq \sqrt{n/3}\}$. The vertex set of H is $V(H) = \bigcup_{i=1}^{\sqrt{n/3}} (L_i \cup R_i) \cup B$ and the edge set of H is $E(H) = \{v_{i,j} l_i^t r_j^t \mid v_{i,j} \in B \text{ and } 1 \leq t \leq \sqrt{n/3}\}$.

Clearly $|V(H)| = n$ and $|E(H)| = \frac{n^{3/2}}{3\sqrt{3}}$ and H is linear. It is easy to check that H is C_5-free but this is proved in a more general setting in Theorem 3.5.

Lazebnik and Verstraëte [59] showed that

$$\mathrm{ex}_3^{\mathrm{lin}}(n, \{C_3, C_4\}) = \frac{n^{3/2}}{6} + O(n). \quad (3.1)$$

This was remarkable especially considering the fact that the asymptotics for the corresponding extremal function for graphs $\mathrm{ex}(n, \{C_3, C_4\})$ is not known and is a long standing problem of Erdős [18]. Erdős and Simonovits [19] conjectured that $\mathrm{ex}(n, \{C_3, C_4\}) = \mathrm{ex}_{\mathrm{bip}}(n, C_4)$ while Allen, Keevash, Sudakov, and Verstraëte [1] conjectured that this is not true.

In this chapter we strengthen the above mentioned result of Lazebnik and Verstraëte [59], by showing that their upper bound in (3.1) still holds even if the C_3-free condition is dropped. This shows $\mathrm{ex}_3^{\mathrm{lin}}(n, C_4) \sim \mathrm{ex}_3^{\mathrm{lin}}(n, \{C_3, C_4\})$, as detailed below.

Theorem 3.4 (E., Győri, Methuku [30])**.**

$$\mathrm{ex}_3^{\mathrm{lin}}(n, C_4) \leq \frac{1}{6} n\sqrt{n+9} + \frac{n}{2} = \frac{n^{3/2}}{6} + O(n).$$

The lower bound $\mathrm{ex}_3^{\mathrm{lin}}(n, C_4) \geq \frac{1}{6}n^{3/2} - \frac{1}{6}\sqrt{n}$ follows from (3.1). (Note that the construction from [59] showing this lower bound is C_3-free as well.) Therefore,

$$\mathrm{ex}_3^{\mathrm{lin}}(n, C_4) = \frac{n^{3/2}}{6} + O(n).$$

The last result of this chapter shows strong connection between Turán numbers of even cycles in graphs and linear Turán numbers of linear cycles of odd length in 3-uniform hypergraphs. This is explained below, after introducing some definitions.

A *linear cycle* C_k^{lin} of length $k \geq 3$ is an alternating sequence $v_1, h_1, v_2, h_2, ..., v_k, h_k$ of distinct vertices and distinct hyperedges such that $h_i \cap h_{i+1} = \{v_{i+1}\}$ for each $i \in \{1, 2, \dots, k-1\}$, $h_1 \cap h_k = \{v_1\}$ and $h_i \cap h_j = \emptyset$ if $1 < |j-i| < k-1$. (A *linear path* can be defined similarly.) The vertices $v_1, v_2, \dots, v_k$ are called the *basic vertices* of C_k^{lin} and the graph with the edge set $\{v_1v_2, v_2v_3, \dots, v_{k-1}v_k, v_kv_1\}$ is called the *basic cycle* of C_k^{lin}.

Let $\mathcal{C}_k$ and $\mathcal{C}_k^{\mathrm{lin}}$ denote the set of (Berge) cycles C_l and the set of linear cycles C_l^{lin}, respectively, where l has the *same parity* as k and $2 \leq l \leq k$. In particular, in Theorem 3.5 we will be interested in the sets $\mathcal{C}_{2k-2} = \{C_2, C_4, C_6, \dots, C_{2k-2}\}$ and $\mathcal{C}_{2k+1}^{\mathrm{lin}} = \{C_3^{\mathrm{lin}}, C_5^{\mathrm{lin}}, \dots, C_{2k+1}^{\mathrm{lin}}\}$. Note that the (Berge) cycle C_2 corresponds to two hyperedges that share at least 2 vertices, so a hypergraph is linear if and only if it is C_2-free. In particular, for graphs (i.e., 2-uniform hypergraphs) the C_2-free condition does not impose any restriction, and there is no difference between a (Berge) cycle C_l and a linear cycle C_l^{lin}.

Recall that Bondy and Simonovits [10] showed that for $k \geq 2$, $\mathrm{ex}(n, C_{2k}) \leq c_k n^{1+\frac{1}{k}}$ for all sufficiently large n. Improvements to the constant factor c_k are made in [72, 66, 13]. The *girth* of a graph is the length of a shortest cycle contained in the graph. For $k = 2, 3, 5$, constructions of C_{2k}-free graphs on n vertices with $\Omega(n^{1+\frac{1}{k}})$ edges are known: Benson [6] and

Singleton [69] constructed a bipartite $\mathcal{C}_6$-free graph with $(1+o(1))(n/2)^{4/3}$ edges and Benson [6] constructed a bipartite $\mathcal{C}_{10}$-free graph with $(1+o(1))(n/2)^{6/5}$ edges. For $k \notin \{2,3,5\}$ it is not known if the order of magnitude of $\mathrm{ex}(n, C_{2k})$ is $\Theta(n^{1+\frac{1}{k}})$. The best known lower bound is due to Lazebnik, Ustimenko and Woldar [60], who showed that there exist graphs of girth more than $2k+1$ containing $\Omega(n^{1+\frac{2}{3k-3+\epsilon}})$ edges where $k \geq 2$ is fixed, $\epsilon = 0$ if k is odd and $\epsilon = 1$ if k is even.

Recently Collier-Cartaino, Graber and Jiang [14] showed that for all $l \geq 3$, $\mathrm{ex}_3^{\mathrm{lin}}(n, C_l^{\mathrm{lin}}) \leq O(n^{1+\frac{1}{\lfloor l/2 \rfloor}})$. In fact, they proved the same upper bound for all r-uniform hypergraphs with $r \geq 3$. However, it is not known if C_l^{lin}-free linear 3-uniform hypergraphs on n vertices with $\Omega(n^{1+\frac{1}{\lfloor l/2 \rfloor}})$ hyperedges exist. It is mentioned in [14] that the best known lower bound

$$\mathrm{ex}_3^{\mathrm{lin}}(n, C_l^{\mathrm{lin}}) \geq \Omega(n^{1+\frac{1}{l-1}}), \tag{3.2}$$

was observed by Verstraëte, by taking a random subgraph of a Steiner triple system.

If $l = 2k+1$ is odd, then we are able to construct a $\mathcal{C}_{2k+1}^{\mathrm{lin}}$-free 3-uniform linear hypergraph on n vertices with $\Omega(n^{1+\frac{1}{k}})$ hyperedges whenever a $\mathcal{C}_{2k-2}$-free graph with $\Omega(n^{1+\frac{1}{k-1}})$ edges exists. More precisely, we show:

Theorem 3.5 (E., Győri, Methuku [30]). *Let* $\mathrm{ex}_{\mathrm{bip}}(n, \mathcal{C}_{2k-2}) \geq (1+o(1))c\left(\frac{n}{2}\right)^{\alpha} = \Omega(n^{\alpha})$ *for some* $c, \alpha > 0$*. Then,*

$$\mathrm{ex}_3^{\mathrm{lin}}(n, \mathcal{C}_{2k+1}^{\mathrm{lin}}) \geq (1+o(1))\frac{\alpha c}{4\alpha - 2} \cdot \left(\frac{\alpha - 1}{c(2\alpha - 1)}\right)^{1-\frac{1}{\alpha}} n^{2-\frac{1}{\alpha}} = \Omega(n^{2-\frac{1}{\alpha}}).$$

If $2k-2 = 2$, then by definition $\mathcal{C}_{2k-2} = \{C_2\}$, so in this case the $\mathcal{C}_{2k-2}$-free condition does not impose any restriction. Thus in order to bound $\mathrm{ex}_{\mathrm{bip}}(n, \mathcal{C}_2)$ from below, one can take a complete balanced bipartite graph. Therefore, using $c = 1$ and $\alpha = 2$ in the above theorem, we get $\mathrm{ex}_3^{\mathrm{lin}}(n, \mathcal{C}_5^{\mathrm{lin}}) \geq (1+o(1))\frac{n^{3/2}}{3\sqrt{3}}$. Since a 3-uniform linear hypergraph which is both C_3^{lin}-free and C_5^{lin}-free is (Berge) C_5-free, this also provides the desired lower bound in Theorem 3.3. As we mentioned before, in the cases $2k-2 = 4, 6, 10$, it is known that $c = 1$ and $\alpha = 1 + \frac{1}{k-1}$ by the work of Benson and Singleton and for all $k \geq 2$, it is known that $\alpha = 1 + \frac{2}{3k-6+\epsilon}$ by the work of Lazebnik, Ustimenko and Woldar, where $\epsilon = 0$ if k is odd and $\epsilon = 1$ if k is even; so substituting these in Theorem 3.5 and combining it with the upper bound of Collier-Cartaino, Graber and Jiang, we get the following corollary.

Corollary 3.1. *For* $k = 2, 3, 4, 6$, *we have* $\mathrm{ex}_3^{\mathrm{lin}}(n, \mathcal{C}_{2k+1}^{\mathrm{lin}}) \geq (1+o(1))\frac{k}{2}(\frac{n}{k+1})^{1+\frac{1}{k}}$.

Therefore, in these cases,

$$\mathrm{ex}_3^{\mathrm{lin}}(n, \mathcal{C}_{2k+1}^{\mathrm{lin}}) = \Theta(n^{1+\frac{1}{k}}).$$

Moreover, for $k \geq 2$, *we have*

$$\mathrm{ex}_3^{\mathrm{lin}}(n, \mathcal{C}_{2k+1}^{\mathrm{lin}}) \geq \Omega(n^{1+\frac{2}{3k-4+\epsilon}}),$$

where $\epsilon = 0$ *if* k *is odd and* $\epsilon = 1$ *if* k *is even.*

The above corollary provides an improvement of the lower bound in (3.2) for linear cycles of odd length.

Structure of the chapter: In the next section we introduce some notation that is used through out the chapter. In Section 3.2 we prove the theorem 3.2 In Section 3.3, we prove the upper bound of Theorem 3.3 and in Section 3.4, we prove Theorem 3.4. Finally, in Section 3.5 we prove Theorem 3.5.

Notations and Definitions

We introduce some important notations and definitions used throughout the chapter.

- Length of a path is the number of edges in the path.
- For convenience, an edge $\{a, b\}$ of a graph or a pair of vertices a, b is referred to as ab. A hyperedge $\{a, b, c\}$ is written simply as abc.
- For a hypergraph H (or a graph G), for convenience, we sometimes use H (or G) to denote the edge set of the hypergraph H (or G respectively). Thus the number of edges in H is $|H|$.
- Given a graph G and a subset of its vertices S, let the subgraph of G induced by S be denoted by $G[S]$.
- For a hypergraph H, let $\partial H = \{ab \mid ab \subset e \in E(H)\}$ denote its *2-shadow* graph.
- The *first neighborhood* of v in H is defined as

  $$N_1^H(v) = \{x \in V(H) \setminus \{v\} \mid v, x \in h \text{ for some } h \in E(H)\}$$

  and the *second neighborhood* of v in H is

  $$N_2^H(v) = \{x \in V(H) \setminus (N_1^H(v) \cup \{v\}) \mid \exists h \in E(H) \text{ such that } x \in h \text{ and } h \cap N_1^H(v) \neq \emptyset\}.$$

- For a hypergraph H and $v \in V(H)$, we denote the degree of v in H by $d(v)$. We write $d^H(v)$ instead of $d(v)$ when it is important to emphasize the underlying hypergraph.
- For a hypergraph H and a pair of vertices $u, v \in V(H)$, let $\text{codeg}(v, u)$ denote the number of hyperedges of H containing the pair $\{u, v\}$.

3.2 Proof of Theorem 3.2

Let H be a hypergraph on n vertices without a Berge 5-cycle and let $G = \partial H$ be the 2-shadow of H. First we introduce some definitions.

Definition 3.2. *A pair $xy \in \partial H$ is called* thin *if* $\text{codeg}(xy) = 1$*, otherwise it is called* fat*. We say a hyperedge $abc \in H$ is* thin *if at least two of the pairs ab, bc, ac are thin.*

Definition 3.3. *We say a set of hyperedges (or a hypergraph) is tightly-connected if it can be obtained by starting with a hyperedge and adding hyperedges one by one, such that every added hyperedge intersects with one of the previous hyperedges in 2 vertices.*

Definition 3.4. *A* block *in H is a maximal set of tightly-connected hyperedges.*

Definition 3.5. *For a block B, a maximal subhypergraph of B without containing thin hyperedges is called the* core *of the block.*

Let K_4^3 denote the complete 3-uniform hypergraph on 4 vertices. A crown of size k is a set of $k \geq 1$ hyperedges of the form $abc_1, abc_2, \ldots, abc_k$. Below we define 2 specific hypergraphs:

- Let F_1 be a hypergraph consisting of exactly 3 hyperedges on 4 vertices (i.e., K_4^3 minus an edge).
- For distinct vertices a, b, c, d and o, let F_2 be the hypergraph consisting of hyperedges oab, obc, ocd and oda.

Lemma 3.6. *Let B be a block of H, and let $\mathcal{B}$ be a core of B. Then $\mathcal{B}$ is either $\emptyset, K_4^3, F_1, F_2$ or a crown of size k for some $k \geq 1$.*

Proof. If $\mathcal{B} = \emptyset$, we are done, so let us assume $\mathcal{B} \neq \emptyset$. Since B is tightly-connected and it can be obtained by adding thin hyperedges to $\mathcal{B}$, it is easy to see that $\mathcal{B}$ is also tightly-connected. Thus if $\mathcal{B}$ has at most two hyperedges, then it is a crown of size 1 or 2 and we are done. Therefore, in the rest of the proof we will assume that $\mathcal{B}$ contains at least 3 hyperedges.

If $\mathcal{B}$ contains at most 4 vertices then it is easy to see that $\mathcal{B}$ is either K_4^3 or F_1. So assume that $\mathcal{B}$ has at least 5 vertices (and at least 3 hyperedges). Since $\mathcal{B}$ is not a crown, there exists a tight path of length 3, say abc, bcd, cde. Since abc is in the core, one of the pairs ab or ac is fat, so there exists a hyperedge $h \neq abc$ containing either ab or ac. Similarly there exists a hyperedge $f \neq cde$ and f contains ed or ec. If $h = f$ then $\mathcal{B} \supseteq F_2$. However, it is easy to see that F_2 cannot be extended to a larger tightly-connected set of hyperedges without creating a Berge 5-cycle, so in this case $\mathcal{B} = F_2$. If $h \neq f$ then the hyperedges h, abc, bcd, cde, f create a Berge 5-cycle in H, a contradiction. This completes the proof of the lemma. □

Observation 3.7. Let B be a block of H and let $\mathcal{B}$ be the core of B. If $\mathcal{B} = \emptyset$ then the block B is a crown, and if $\mathcal{B} \neq \emptyset$ then every fat pair of B is contained in $\partial\mathcal{B}$.

Edge Decomposition of $G = \partial H$. We define a decomposition $\mathcal{D}$ of the edges of G into paths of length 2, triangles and K_4's such as follows:
Let B be a block of H and $\mathcal{B}$ be its core.

If $\mathcal{B} = \emptyset$, then B is a crown-block $\{abc_1, abc_2, \ldots, abc_k\}$ (for some $k \geq 1$); we partition ∂B into the triangle abc_1 and paths ac_ib where $2 \leq i \leq k$.

If $\mathcal{B} \neq \emptyset$, then our plan is to first partition $\partial B \setminus \partial\mathcal{B}$. If $abc \in B \setminus \mathcal{B}$, then abc is a thin hyperedge, so it contains at least 2 thin pairs, say ab and bc. We claim that the pair ac is in $\partial\mathcal{B}$. Indeed, ac has to be a fat pair, otherwise the block B consists of only one hyperedge abc, so $\mathcal{B} = \emptyset$ contradicting the assumption. So by Observation 3.7, ac has to be a pair in $\partial\mathcal{B}$. For every $abc \in B \setminus \mathcal{B}$ such that ab and bc are thin pairs, add the 2-path abc to the edge decomposition $\mathcal{D}$. This partitions all the edges in $\partial B \setminus \partial\mathcal{B}$ into paths of length 2. So all we have left is to partition the edges of $\partial\mathcal{B}$.

- If $\mathcal{B}$ is a crown $\{abc_1, abc_2, \ldots, abc_k\}$ for some $k \geq 1$, then we partition ∂B into the triangle abc_1 and paths ac_ib where $2 \leq i \leq k$.
- If $\mathcal{B} = F_1 = \{abc, bcd, acd\}$ then we partition $\partial\mathcal{B}$ into 2-paths abc, bdc and cad.
- If $\mathcal{B} = F_2 = \{oab, obc, ocd, oda\}$ then we partition $\partial\mathcal{B}$ into 2-paths abo, bco, cdo and dao.
- Finally, if $\mathcal{B} = K_4^3 = \{abc, abd, acd, bcd\}$ then we partition $\partial\mathcal{B}$ as K_4, i.e., we add $\partial\mathcal{B} = K_4$ as an element of $\mathcal{D}$.

Clearly, by Lemma 3.6 we have no other cases left. Thus all of the edges of the graph G are partitioned into paths of length 2, triangles and K_4's.

Observation 3.8.

(a) If D is a triangle that belongs to $\mathcal{D}$, then there is a hyperedge $h \in H$ such that $D = \partial h$.

(b) If abc is a 2-path that belongs to $\mathcal{D}$, then $abc \in H$. Moreover ac is a fat pair.

(c) If D is a K_4 that belongs to $\mathcal{D}$, then there exists $F = K_4^3 \subseteq H$ such that $D = \partial F$.

Let $\alpha_1 |G|$ and $\alpha_2 |G|$ be the number of edges of G that are contained in triangles and 2-paths of the edge-decomposition $\mathcal{D}$ of G, respectively. So $(1 - \alpha_1 - \alpha_2)|G|$ edges of G belong to the K_4's in $\mathcal{D}$.

Claim 3.9. *We have,*

$$|H| = \left(\frac{\alpha_1}{3} + \frac{\alpha_2}{2} + \frac{2(1 - \alpha_1 - \alpha_2)}{3}\right)|G|.$$

Proof. Let B be a block with the core $\mathcal{B}$. Recall that for each hyperedge $h \in B \setminus \mathcal{B}$, we have added exactly one 2-path or a triangle to $\mathcal{D}$.

Moreover, because of the way we partitioned $\partial\mathcal{B}$, it is easy to check that in all of the cases except when $\mathcal{B} = K_4^3$, the number of hyperedges of $\mathcal{B}$ is the same as the number of elements of $\mathcal{D}$ that $\partial\mathcal{B}$ is partitioned into; these elements being 2-paths and triangles. On the other hand, if $\mathcal{B} = K_4^3$, then the number of hyperedges of $\mathcal{B}$ is 4 but we added only one element to $\mathcal{D}$ (namely K_4).

This shows that the number of hyperedges of H is equal to the number of elements of $\mathcal{D}$ that are 2-paths or triangles plus the number of hyperedges which are in copies of K_4^3 in H, i.e., 4 times the number of K_4's in $\mathcal{D}$. Since $\alpha_1 |G|$ edges of G are in 2-paths, the number of elements of $\mathcal{D}$ that are 2-paths is $\alpha_1 |G| /2$. Similarly, the number of elements of $\mathcal{D}$ that are triangles is $\alpha_2 |G| /3$, and the number of K_4's in $\mathcal{D}$ is $(1 - \alpha_1 - \alpha_2) |G| /6$. Combining this with the discussion above finishes the proof of the claim. □

The link of a vertex v is the graph consisting of the edges $\{uw \mid uvw \in H\}$ and is denoted by L_v.

Claim 3.10. $|L_v| \leq 2 |N_1(v)|$.

Proof. First let us notice that there is no path of length 5 in L_v. Indeed, otherwise, there exist vertices $v_0, v_1, \ldots, v_5$ such that $vv_{i-1}v_i \in H$ for each $1 \leq i \leq 5$ which means there is a Berge 5-cycle in H formed by the hyperedges containing the pairs $vv_1, v_1v_2, v_2v_3, v_3v_4, v_4v$, a contradiction. So by the Erdős-Gallai theorem $|L_v| \leq \frac{5-1}{2} |N_1(v)|$, proving the claim. □

Lemma 3.11. *Let $v \in V(H)$ be an arbitrary vertex, then the number of edges in $G[N_1(v)]$ is less than $8 |N_1(v)|$.*

Proof. Let G_v be a subgraph of G on a vertex set $N_1(v)$, such that $xy \in G_v$ if and only if there exists a vertex $z \neq v$ such that $xyz \in H$. Then each edge of $G[N_1(v)]$ belongs to either L_v or G_v, so $|G[N_1(v)]| \leq |L_v| + |G_v|$. Combining this with Claim 3.10, we get $|G[N_1(v)]| \leq |G_v| + 2 |N_1(v)|$. So it suffices to prove that $|G_v| < 6 |N_1(v)|$.

First we will prove that there is no path of length 12 in G_v. Let us assume by contradiction that $P = v_0, v_1, \ldots, v_{12}$ is a path in G_v. Since for each pair of vertices v_i, v_{i+1}, there is a hyperedge $v_iv_{i+1}x$ in H where $x \neq v$, we can conclude that there is a subsequence $u_0, u_1, \ldots, u_6$ of $v_0, v_1, \ldots, v_{12}$ and a sequence of distinct hyperedges $h_1, h_2, \ldots, h_6$, such that $u_{i-1}u_i \subset h_i$ and $v \notin h_i$ for each $1 \leq i \leq 6$. Since $u_0, u_3, u_6 \in N_1(v)$ there exist hyperedges $f_1, f_2, f_3 \in H$ such that $vu_0 \subset f_1$, $vu_3 \subset f_2$ and $vu_6 \subset f_3$. Clearly, either $f_1 \neq f_2$ or $f_2 \neq f_3$. In the first case the hyperedges f_1, h_1, h_2, h_3, f_2, and in the second case the hyperedges f_2, h_4, h_5, h_6, f_3 form a Berge 5-cycle in H, a contradiction.

Therefore, there is no path of length 12 in G_v, so by the Erdős-Gallai theorem, the number of edges in G_v is at most $\frac{12-1}{2} |N_1(v)| < 6 |N_1(v)|$, as required. □

3.2.1 Relating the hypergraph degree to the degree in the shadow

For a vertex $v \in V(H) = V(G)$, let $d(v)$ denote the degree of v in H and let $d_G(v)$ denote the degree of v in G (i.e., $d_G(v)$ is the degree in the shadow).

Clearly $d_G(v) \leq 2d(v)$. Moreover, $d(v) = |L_v|$ and $d_G(v) = |N_1(v)|$. So by Claim 3.10, we have

$$\frac{d_G(v)}{2} \leq d(v) \leq 2d_G(v). \tag{3.3}$$

Let $\overline{d}$ and $\overline{d}_G$ be the average degrees of H and G respectively.

Suppose there is a vertex v of H, such that $d(v) < \overline{d}/3$. Then we may delete v and all the edges incident to v from H to obtain a graph H' whose average degree is more than $3(n\overline{d}/3 - \overline{d}/3)/(n-1) = \overline{d}$. Then it is easy to see that if the theorem holds for H', then it holds for H as well. Repeating this procedure, we may assume that for every vertex v of H, $d(v) \geq \overline{d}/3$. Therefore, by (3.3), we may assume that the degree of every vertex of G is at least $\overline{d}/6$.

3.2.2 Counting paths of length 3

Definition 3.12. *A 2-path in ∂H is called* bad *if both of its edges are contained in a triangle of ∂H, otherwise it is called* good*.*

Lemma 3.13. *For any vertex $v \in V(G)$ and a set $M \subseteq N_1(v)$, let $\mathcal{P}$ be the set of the good 2-paths vxy such that $x \in M$. Let $M' = \{y \mid vxy \in \mathcal{P}\}$ then $|\mathcal{P}| < 2|M'| + 48d_G(v)$.*

Proof. Let $B_{\mathcal{P}} = \{xy \mid x \in M, y \in M', xy \in G\}$ be a bipartite graph, clearly $|B_{\mathcal{P}}| = |\mathcal{P}|$. Let $E = \{xyz \in H \mid x, y \in N_1(v), \text{codeg}(x,y) \leq 2\}$. By Lemma 3.11, $|E| \leq 2 \cdot 8|N_1(v)|$ so the number of edges of 2-shadow of E is $|\partial E| \leq 48|N_1(v)|$. Let $B = \{xy \in B_{\mathcal{P}} \mid \exists z \in V(H), xyz \in H \setminus E\}$. Then clearly,

$$|B| \geq |B_{\mathcal{P}}| - |\partial E| \geq |\mathcal{P}| - 48|N_1(v)| = |\mathcal{P}| - 48d_G(v). \tag{3.4}$$

Let $d_B(x)$ denote the degree of a vertex x in the graph B.

Claim 3.14. *For every $y \in M'$ such that $d_B(y) = k \geq 3$, there exists a set of $k-2$ vertices $S_y \subseteq M'$ such that $\forall w \in S_y$ we have $d_B(w) = 1$. Moreover, $S_y \cap S_z = \emptyset$ for any $y \neq z \in M'$ (with $d_B(y), d_B(z) \geq 3$).*

Proof. Let $yx_1, yx_2, \ldots, yx_k \in B$ be the edges of B incident to y. For each $1 \leq j \leq k$ let $f_j \in H$ be a hyperedge such that $vx_j \subset f_j$. For each $yx_i \in B$ clearly there is a hyperedge $yx_iw_i \in H \setminus E$.

We claim that for each $1 \leq i \leq k$, $w_i \in M'$. It is easy to see that $w_i \in N_1(v)$ or $w_i \in M'$ (because vx_iw_i is a 2-path in G). Assume for a contradiction that $w_i \in N_1(v)$, then since $yx_iw_i \notin E$ we have, $\text{codeg}(x_i, w_i) \geq 3$. Let $f \in H$ be a hyperedge such that $vw_i \subset f$. Now take $j \neq i$ such that $x_j \neq w_i$. If $f_j \neq f$ then since $\text{codeg}(x_i, w_i) \geq 3$ there exists a hyperedge $h \supset x_iw_i$ such that $h \neq f$ and $h \neq x_iw_iy$, then the hyperedges $f, h, x_iw_iy, yx_jw_j, f_j$ form a Berge 5-cycle. So $f_j = f$, therefore $f_j \neq f_i$. Similarly in this case, there exists a hyperedge $h \supset x_iw_i$ such that $h \neq f_i$ and $h \neq x_iw_iy$, therefore the hyperedges $f_i, h, x_iw_iy, yx_jw_j, f_j$ form a Berge 5-cycle, a contradiction. So we proved that $w_i \in M'$ for each $1 \leq i \leq k$.

Claim. *For all but at most 2 of the w_i's (where $1 \leq i \leq k$), we have $d_B(w_i) = 1$.*

Proof. If $d_B(w_i) = 1$ for all $1 \leq i \leq k$ then we are done, so we may assume that there is $1 \leq i \leq k$ such that $d_B(w_i) \neq 1$.

For each $1 \le i \le k$, $w_i \in M'$ and $x_iw_i \in \partial(H \setminus E)$ (because $x_iw_iy \in H \setminus E$), so it is clear that $d_B(w_i) \ge 1$. So $d_B(w_i) > 1$. Then there is a vertex $x \in M \setminus \{x_i\}$ such that $w_ix \in B$. Let $f, h \in H$ be hyperedges with $w_ix \in h$ and $xv \in f$. If there are $j, l \in \{1, 2, \dots, k\} \setminus \{i\}$ such that x, x_j and x_l are all different from each other, then clearly, either $f \ne f_j$ or $f \ne f_l$, so without loss of generality we may assume $f \ne f_j$. Then the hyperedges $f, h, w_ix_iy, yw_jx_j, f_j$ create a Berge cycle of length 5, a contradiction. So there are no $j, l \in \{1, 2, \dots, k\} \setminus \{i\}$ such that x, x_j and x_l are all different from each other. Clearly this is only possible when $k < 4$ and there is a $j \in \{1, 2, 3\} \setminus \{i\}$ such that $x = x_j$. Let $l \in \{1, 2, 3\} \setminus \{i, j\}$. If $f_j \ne f_l$ then the hyperedges $f_j, h, w_ix_iy, yw_lx_l, f_l$ form a Berge 5-cycle. Therefore $f_j = f_l$. So we proved that $d_B(w_i) \ne 1$ implies that $k = 3$ and for $\{j, l\} = \{1, 2, 3\} \setminus \{i\}$, we have $f_j = f_l$. So if $d_B(w_i) \ne 1$ and $d_B(w_j) \ne 1$ we have $f_j = f_l$ and $f_i = f_l$, which is impossible. So $d_B(w_j) = 1$. So we proved that if for any $1 \le i \le k$, $d_B(w_i) \ne 1$ then $k = 3$ and all but at most 2 of the vertices in $\{w_1, w_2, w_3\}$ have degree 1 in the graph B, as desired. □

We claim that for any $i \ne j$ where $d_B(w_i) = d_B(w_j) = 1$ we have $w_i \ne w_j$. Indeed, if there exists $i \ne j$ such that $w_i = w_j$ then w_ix_j and w_ix_i are both adjacent to w_i in the graph B which contradicts to $d_B(w_i) = 1$. So using the above claim, we conclude that the set $\{w_1, w_2, \dots, w_k\}$ contains at least $k - 2$ distinct elements with each having degree one in the graph B, so we can set S_y to be the set of these $k - 2$ elements. (Then of course $\forall w_i \in S_y$ we have $d_B(w_i) = 1$.)

Now we have to prove that for each $z \ne y$ we have $S_y \cap S_z = \emptyset$. Assume by contradiction that $w_i \in S_z \cap S_y$ for some $z \ne y$. That is, there is some hyperedge $uw_iz \in H \setminus E$ where $u \in M$, moreover $u = x_i$ otherwise $d_B(w_i) > 1$. So we have a hyperedge $x_iw_iz \in H \setminus E$ for some $z \in M' \setminus \{y\}$. Let $j, l \in \{1, 2, \dots, k\} \setminus \{i\}$ such that $j \ne l$. Recall that $x_jv \subset f_j$ and $x_lv \subset f_l$. Clearly either $f_j \ne f_i$ or $f_l \ne f_i$ so without loss of generality we can assume $f_j \ne f_i$. Then it is easy to see that the hyperedges $f_j, x_jw_jy, yx_iw_i, w_izx_i, f_i$ are all different and they create a Berge 5-cycle ($x_jw_jy \ne yx_iw_i$ because $x_j \ne w_i$). □

For each $x \in M'$ with $d_B(x) = k \ge 3$, let S_x be defined as in Claim 3.14. Then the average of the degrees of the vertices in $S_x \cup \{x\}$ in B is $(k + |S_x|)/(k - 1) = (2k - 2)(k - 1) = 2$. Since the sets $S_x \cup x$ (with $x \in M'$, $d_B(x) \ge 3$) are disjoint, we can conclude that average degree of the set M' is at most 2. Therefore $2|M'| \ge |B|$. So by (3.4) we have $2|M'| \ge |B| > |\mathcal{P}| - 48d_G(V)$, which completes the proof of the lemma. □

Claim 3.15. *We may assume that the maximum degree in the graph G is less than $160\sqrt{n}$ when n is large enough.*

Proof. Let v be an arbitrary vertex with $d_G(v) = C\overline{d}$ for some constant $C > 0$. Let $\mathcal{P}$ be the set of the good 2-paths starting from the vertex v. Then applying Lemma 3.13 with $M = N_1(v)$ and $M' = \{y \mid vxy \in \mathcal{P}\}$, we have $|\mathcal{P}| < 2|M'| + 48d_G(v) < 2n + 48 \cdot C\overline{d}$. Since the minimum degree in G is at least $\overline{d}/6$, the number of (ordered) 2-paths starting from v is at least $d(v) \cdot (\overline{d}/6 - 1) = C\overline{d} \cdot (\overline{d}/6 - 1)$. Notice that the number of (ordered) bad 2-paths starting at v is the number of 2-paths vxy such that $x, y \in N_1(v)$. So by Lemma 3.11, this is at most $2 \cdot 8 |N_1(v)| = 16C\overline{d}$, so the number of good 2-paths is at least $C\overline{d} \cdot (\overline{d}/6 - 17)$. So $|\mathcal{P}| \geq C\overline{d} \cdot (\overline{d}/6 - 17)$. Thus we have

$$C\overline{d} \cdot (\overline{d}/6 - 17) \leq |\mathcal{P}| < 2n + 48C\overline{d}.$$

So $C\overline{d}(\overline{d}/6 - 65) < 2n$. Therefore, $6C(\overline{d}/6 - 65)^2 < 2n$, i.e., $\overline{d} < 6\sqrt{n/3C} + 390$, so $|H| = n\overline{d}/3 \leq 2n\sqrt{n/3C} + 130n$. If $C \geq 36$ we get that $|H| \leq \frac{n^{3/2}}{3\sqrt{3}} + 130n = \frac{n^{3/2}}{3\sqrt{3}} + O(n)$, proving Theorem 3.2. So we may assume $C < 36$.

Theorem 3.1 implies that

$$|H| = n\overline{d}/3 \leq \sqrt{2}n^{3/2} + 4.5n, \tag{3.5}$$

so $\overline{d} \leq 3\sqrt{2}\sqrt{n} + 13.5$. So combining this with the fact that $C < 36$, we have $d_G(v) = C\overline{d} < 108\sqrt{2}\sqrt{n} + 486 < 160\sqrt{n}$ for large enough n. □

Combining Lemma 3.13 and Claim 3.15, we obtain the following.

Lemma 3.16. *For any vertex $v \in V(G)$ and a set $M \subseteq N_1(v)$, let $\mathcal{P}$ be the set of good 2-paths vxy such that $x \in M$. Let $M' = \{y \mid vxy \in \mathcal{P}\}$ then $|\mathcal{P}| < 2|M'| + 7680\sqrt{n}$ when n is large enough.*

Definition 3.17. *A 3-path x_0, x_1, x_2, x_3 is called* good *if both 2-paths x_0, x_1, x_2 and x_1, x_2, x_3 are good 2-paths.*

Claim 3.18. *The number of (ordered) good 3-paths in G is at least $n\overline{d}_G^3 - C_0 n^{3/2}\overline{d}_G$ for some constant $C_0 > 0$ (for large enough n).*

Proof. First we will prove that the number of (ordered) 3-walks that are not good 3-paths is at most $5440n^{3/2}\overline{d}_G$.

For any vertex $x \in V(H)$ if a path yxz is a bad 2-path then zy is an edge of G, so the number of (ordered) bad 2-paths whose middle vertex is x, is at most 2 times the number of edges in $G[N_1(x)]$, which is less than $2 \cdot 8|N_1(x)| = 16d_G(x)$ by Lemma 3.11. The number of 2-walks which are not 2-paths and whose middle vertex is x is exactly $d_G(x)$. So the total number of (ordered) 2-walks that are not good 2-paths is at most $\sum_{x \in V(H)} 17d_G(x) = 17n\overline{d}_G$.

Notice that, by definition, any (ordered) 3-walk that is not a good 3-path must contain a 2-walk that is not a good 2-path. Moreover, if xyz is a 2-walk that is not a good 2-path, then the number of 3-walks in G containing it is at most $d_G(x) + d_G(z) < 320\sqrt{n}$ (for large enough n) by Claim 3.15. Therefore, the total number of (ordered) 3-walks that are not good 3-paths is at most $17n\overline{d}_G \cdot 320\sqrt{n} = 5440n^{3/2}\overline{d}_G$.

By the Blakley-Roy inequality, the total number of (ordered) 3-walks in G is at least $n\overline{d}_G^3$. By the above discussion, all but at most $5440n^{3/2}\overline{d}_G$ of them are good 3-paths, so letting $C_0 = 5440$ completes the proof of the claim. □

Claim 3.19. *Let $\{a, b, c\}$ be the vertex set of a triangle that belongs to $\mathcal{D}$. (By Observation 3.8 (a) $abc \in H$.) Then the number of good 3-paths whose first edge is ab, bc or ca is at most $8n + C_1\sqrt{n}$ for some constant C_1 and for large enough n.*

Proof. Let $S_{abc} = N_1(a) \cap N_1(b) \cap N_1(c)$. For each $\{x, y\} \subset \{a, b, c\}$, let $S_{xy} = N_1(x) \cap N_1(y) \setminus \{a, b, c\}$. For each $x \in \{a, b, c\}$, let $S_x = N_1(x) \setminus (N_1(y) \cup N_1(z) \cup \{a, b, c\})$ where $\{y, z\} = \{a, b, c\} \setminus \{x\}$.

For each $x \in \{a, b, c\}$, let $\mathcal{P}_x$ be the set of good 2-paths xuv where $u \in S_x$. Let $S'_x = \{v \mid xuv \in \mathcal{P}_x\}$. For each $\{x, y\} \subset \{a, b, c\}$, let $\mathcal{P}_{xy}$ be the set of good 2-paths xuv and yuv where $u \in S_{xy}$. Let $S'_{xy} = \{v \mid xuv \in \mathcal{P}_{xy}\}$.

Let $\{x, y\} \subset \{a, b, c\}$ and $z = \{a, b, c\} \setminus \{x, y\}$. Notice that each 2-path $yuv \in \mathcal{P}_{xy}$ ($xuv \in \mathcal{P}_{xy}$), is contained in exactly one good 3-path $zyuv$ (respectively $zxuv$) whose first edge is in the triangle abc. Indeed, since $u \in S_{xy}$, $xyuv$ (respectively $yxuv$) is not a good 3-path. Therefore, the number of good 3-paths whose first edge is in the triangle abc, and whose third vertex is in S_{xy} is $|\mathcal{P}_{xy}|$. The number of paths in $\mathcal{P}_{xy}$ that start with the vertex x is less than $2\left|S'_{xy}\right| + 7680\sqrt{n}$, by Lemma 3.16. Similarly, the number of paths in $\mathcal{P}_{xy}$ that start with the vertex y is less than $2\left|S'_{xy}\right| + 7680\sqrt{n}$. Since every path in $\mathcal{P}_{xy}$ starts with either x or y, we have $|\mathcal{P}_{xy}| < 4\left|S'_{xy}\right| + 15360\sqrt{n}$. Therefore, for any $\{x, y\} \subset \{a, b, c\}$, the number of good 3-paths whose first edge is in the triangle abc, and whose third vertex is in S_{xy} is less than $4\left|S'_{xy}\right| + 15360\sqrt{n}$.

In total, the number of good 3-paths whose first edge is in the triangle abc and whose third vertex is in $S_{ab} \cup S_{bc} \cup S_{ac}$ is at most

$$4(|S'_{ab}| + |S'_{bc}| + |S'_{ac}|) + 46080\sqrt{n}. \tag{3.6}$$

Let $x \in \{a, b, c\}$ and $\{y, z\} = \{a, b, c\} \setminus \{x\}$. For any 2-path $xuv \in \mathcal{P}_x$ there are 2 good 3-paths with the first edge in the triangle abc, namely $yxuv$ and $zxuv$. So the total number of 3-paths whose first edge is in the triangle abc and whose third vertex is in $S_a \cup S_b \cup S_c$ is $2(|\mathcal{P}_a| + |\mathcal{P}_b| + |\mathcal{P}_c|)$, which is at most

$$4(|S'_a| + |S'_b| + |S'_c|) + 46080\sqrt{n}, \tag{3.7}$$

by Lemma 3.16.

Now we will prove that every vertex is in at most 2 of the sets $S'_a, S'_b, S'_c, S'_{ab}, S'_{bc}, S'_{ac}$. Let us assume by contradiction that a vertex $v \in V(G) \setminus \{a, b, c\}$ is in at least 3 of them. We claim that there do not exist 3 vertices $u_a \in N_1(a) \setminus \{b, c\}$, $u_b \in N_1(b) \setminus \{a, c\}$ and $u_c \in N_1(c) \setminus \{a, b\}$ such that xu_xv is a good 3-path for each $x \in \{a, b, c\}$. Indeed, otherwise, consider hyperedges h_a, h'_a containing the pairs au_a and u_av respectively (since au_av is a good 2-path, note that $h_a \neq h'_a$), and hyperedges h_b, h'_b, h_c, h'_c containing the pairs bu_b, u_bv, cu_c, u_cv respectively. Then either $h'_a \neq h'_b$ or $h'_a \neq h'_c$, say $h'_a \neq h'_b$ without loss of generality. Then the hyperedges $h_a, h'_a, h'_b, h_b, abc$ create a Berge 5-cycle in H, a contradiction, proving that it is impossible to have 3 vertices $u_a \in N_1(a) \setminus \{b, c\}$, $u_b \in N_1(b) \setminus \{a, c\}$ and $u_c \in N_1(c) \setminus \{a, b\}$ with the above mentioned property. Without loss of generality let us assume that there is no vertex $u_a \in N_1(a) \setminus \{b, c\}$ such that au_av is a good 2-path – in other words, $v \notin S'_a \cup S'_{ab} \cup S'_{ac}$. However, since we assumed that v is contained in at least 3 of the sets $S'_a, S'_b, S'_c, S'_{ab}, S'_{bc}, S'_{ac}$, we can conclude that v is contained in all 3 of the sets S'_b, S'_c, S'_{bc}, i.e., there are vertices $u_b \in S_b, u_c \in S_c, u \in S_{bc}$ such that vu_bb, vu_cc, vub, vuc are good 2-paths. Using a similar argument as before, if $vu \in h$, $vu_b \in h_b$ and $vu_c \in h_c$, without loss of generality we can assume that $h \neq h_b$, so the hyperedges abc,h,h_b together with hyperedges containing uc and $u_b b$ form a Berge 5-cycle in H, a contradiction.

So we proved that

$$2\,|S'_a \cup S'_b \cup S'_c \cup S'_{ab} \cup S'_{bc} \cup S'_{ac}| \geq |S'_a| + |S'_b| + |S'_c| + |S'_{ab}| + |S'_{bc}| + |S'_{ac}|$$

This together with (3.6) and (3.7), we get that the number of good 3-paths whose first edge is in the triangle abc is at most

$$8\,|S'_a \cup S'_b \cup S'_c \cup S'_{ab} \cup S'_{bc} \cup S'_{ac}| + 92160\sqrt{n} < 8n + C_1\sqrt{n}$$

for $C_1 = 92160$ and large enough n, finishing the proof of the claim. □

Claim 3.20. *Let $P = abc$ be a 2-path and $P \in \mathcal{D}$. (By Observation 3.8 (b) $abc \in H$.) Then the number of good 3-paths whose first edge is ab or bc is at most $4n + C_2\sqrt{n}$ for some constant $C_2 > 0$ and large enough n.*

Proof. First we bound the number of 3-paths whose first edge is ab. Let $S_{ab} = N_1(a) \cap N_1(b)$. Let $S_a = N_1(a) \setminus (N_1(b) \cup \{b\})$ and $S_b = N_1(b) \setminus (N_1(a) \cup \{a\})$. For each $x \in \{a, b\}$, let $\mathcal{P}_x$ be the set of good 2-paths xuv where $u \in S_x$, and let $S'_x = \{v \mid xuv \in \mathcal{P}_x\}$. The set of good 3-paths whose first edge is ab is $\mathcal{P}_a \cup \mathcal{P}_b$, because the third vertex of a good 3-path starting with an edge ab can not belong to $N_1(a) \cap N_1(b)$ by the definition of a good 3-path.

We claim that $|S'_a \cap S'_b| \leq 160\sqrt{n}$. Let us assume by contradiction that $v_0, v_1, \ldots v_k \in S'_a \cap S'_b$ for $k > 160\sqrt{n}$. For each vertex v_i where $0 \leq i \leq k$, there are vertices $a_i \in S_a$ and $b_i \in S_b$ such that aa_iv_i, bb_iv_i are good 2-paths. For each $0 \leq i \leq k$, the hyperedge $a_iv_ib_i$ is in H, otherwise we can find distinct hyperedges containing the pairs $aa_i, a_iv_i, v_ib_i, b_ib$ and these hyperedges together with abc, would form a Berge 5-cycle in H, a contradiction. We claim that there are $j, l \in \{0, 1, \ldots, k\}$ such that $a_j \neq a_l$, otherwise there is a vertex x such that $x = a_i$ for each $0 \leq i \leq k$. Then $xv_i \in G$ for each $0 \leq i \leq k$, so we get that $d_G(x) > k > 160\sqrt{n}$ which contradicts Claim 3.15.

So there are $j, l \in \{0, 1, \ldots, k\}$ such that $a_j \neq a_l$ and $a_jv_jb_j, a_lv_lb_l \in H$. By observation 3.8 (b), there is a hyperedge $h \neq abc$ such that $ac \subset h$. Clearly either $a_j \notin h$ or $a_l \notin h$. Without loss of generality let $a_j \notin h$, so there is a hyperedge h_a with $aa_j \subset h_a \neq h$. Let $h_b \supset b_jb$, then the hyperedges $abc, h, h_a, a_jv_jb_j, h_b$ form a Berge 5-cycle, a contradiction, proving that $|S'_a \cap S'_b| \leq 160\sqrt{n}$.

Notice that $|S'_a| + |S'_b| = |S'_a \cup S'_b| + |S'_a \cap S'_b| \leq n + 160\sqrt{n}$. So by Lemma 3.16, we have

$$|\mathcal{P}_a| + |\mathcal{P}_b| \leq 2(|S'_a| + |S'_b|) + 2 \cdot 7680\sqrt{n} \leq 2(n + 160\sqrt{n}) + 2 \cdot 7680\sqrt{n} = 2n + 15680\sqrt{n}$$

for large enough n. So the number of good 3-paths whose first edge is ab is at most $2n + 15680\sqrt{n}$. By the same argument, the number of good 3-paths whose first edge is bc is at most $2n + 15680\sqrt{n}$. Their sum is at most $4n + C_2\sqrt{n}$ for $C_2 = 31360$ and large enough n, as desired. □

Claim 3.21. *Let $\{a,b,c,d\}$ be the vertex set of a K_4 that belongs to $\mathcal{D}$. Let $F = K_4^3$ be a hypergraph on the vertex set $\{a,b,c,d\}$. (By Observation 3.8 (c) $F \subseteq H$.) Then the number of good 3-paths whose first edge belongs to ∂F is at most $6n + C_3\sqrt{n}$ for some constant $C_3 > 0$ and large enough n.*

Proof. First, let us observe that there is no Berge path of length $2, 3$ or 4 between distinct vertices $x, y \in \{a,b,c,d\}$ in the hypergraph $H \setminus F$, because otherwise this Berge path together with some edges of F will form a Berge 5-cycle in H. This implies, that there is no path of length 3 or 4 between x and y in $G \setminus \partial F$, because otherwise we would find a Berge path of length $2, 3$ or 4 between x and y in $H \setminus F$.

Let $S = \{u \in V(H) \setminus \{a,b,c,d\} \mid \exists\{x,y\} \subset \{a,b,c,d\}, u \in N_1(x) \cap N_1(y)\}$. For each $x \in \{a,b,c,d\}$, let $S_x = N_1(x) \setminus (S \cup \{a,b,c,d\})$. Let $\mathcal{P}_S$ be the set of good 2-paths xuv where $x \in \{a,b,c,d\}$ and $u \in S$. Let $S' = \{v \mid xuv \in \mathcal{P}_S\}$. For each $x \in \{a,b,c,d\}$, let $\mathcal{P}_x$ be the set of good 2-paths xuv where $u \in S_x$, and let $S'_x = \{v \mid xuv \in \mathcal{P}_x\}$.

Let $v \in S'$. By definition, there exists a pair of vertices $\{x,y\} \subset \{a,b,c,d\}$ and a vertex u, such that xuv and yuv are good 2-paths.

Suppose that $zu'v$ is a 2-path different from xuv and yuv where $z \in \{a,b,c,d\}$. If $u' = u$ then $z \notin \{x,y\}$ so there is a Berge 2-path between x and y or between x and z in $H \setminus F$, which is impossible. So $u \neq u'$. Either $z \neq x$ or $z \neq y$, without loss of generality let us assume that $z \neq x$. Then $zu'vux$ is a path of length 4 in $G \setminus \partial F$, a contradiction. So for any $v \in S'$ there are only 2 paths of $\mathcal{P}_a \cup \mathcal{P}_b \cup \mathcal{P}_c \cup \mathcal{P}_d \cup \mathcal{P}_S$ that contain v as an end vertex – both of which are in $\mathcal{P}_S$ – which means that $v \notin S'_a \cup S'_b \cup S'_c \cup S'_d$, so $S' \cap (S'_a \cup S'_b \cup S'_c \cup S'_d) = \emptyset$. Moreover,

$$|\mathcal{P}_S| \leq 2\,|S'|. \tag{3.8}$$

We claim that S'_a and S'_b are disjoint. Indeed, otherwise, if $v \in S'_a \cap S'_b$ there exists $x \in S_a$ and $y \in S_b$ such that vxa and vyb are paths in G, so there is a 4-path $axvyb$ between vertices of F in $G \setminus \partial F$, a contradiction. Similarly we can prove that S'_a, S'_b, S'_c and S'_d are pairwise disjoint. This shows that the sets S', S'_a, S'_b, S'_c and S'_d are pairwise disjoint. So we have

$$|S' \cup S'_a \cup S'_b \cup S'_c \cup S'_d| = |S'| + |S'_a| + |S'_b| + |S'_c| + |S'_d|. \tag{3.9}$$

By Lemma 3.16, we have $|\mathcal{P}_a| + |\mathcal{P}_b| + |\mathcal{P}_c| + |\mathcal{P}_d| \leq 2(|S'_a| + |S'_b| + |S'_c| + |S'_d|) + 4 \cdot 7680\sqrt{n}$. Combining this inequality with (3.8), we get

$$|\mathcal{P}_S| + |\mathcal{P}_a| + |\mathcal{P}_b| + |\mathcal{P}_c| + |\mathcal{P}_d| \leq 2\,|S'| + 2(|S'_a| + |S'_b| + |S'_c| + |S'_d|) + 4 \cdot 7680\sqrt{n}. \tag{3.10}$$

Combining (3.9) with (3.10) we get

$$|\mathcal{P}_S| + |\mathcal{P}_a| + |\mathcal{P}_b| + |\mathcal{P}_c| + |\mathcal{P}_d| \leq 2\,|S' \cup S'_a \cup S'_b \cup S'_c \cup S'_d| + 30720\sqrt{n} < 2n + 30720\sqrt{n}, \tag{3.11}$$

for large enough n.

Each 2-path in $\mathcal{P}_S \cup \mathcal{P}_a \cup \mathcal{P}_b \cup \mathcal{P}_c \cup \mathcal{P}_d$ can be extended to at most three good 3-paths whose first edge is in ∂F. (For example, $auv \in \mathcal{P}_a$ can be extended to $bauv, cauv$ and $dauv$.) On the other hand, every good 3-path whose first edge is in ∂F must contain a 2-path of $\mathcal{P}_a \cup \mathcal{P}_b \cup \mathcal{P}_c \cup \mathcal{P}_d \cup \mathcal{P}_S$ as a subpath. So the number of good 3-paths whose first edge is in ∂F is at most $3\,|\mathcal{P}_a \cup \mathcal{P}_b \cup \mathcal{P}_c \cup \mathcal{P}_d \cup \mathcal{P}_S| = 3(|\mathcal{P}_S| + |\mathcal{P}_a| + |\mathcal{P}_b| + |\mathcal{P}_c| + |\mathcal{P}_d|)$ which is at most $6n + C_3\sqrt{n}$ by (3.11), for $C_3 = 92160$ and large enough n, proving the desired claim. □

3.2.3 Combining bounds on the number of 3-paths

Recall that $\alpha_1 |G|$, $\alpha_2 |G|$, $(1-\alpha_1-\alpha_2)|G|$ are the number of edges of G that are contained in triangles, 2-paths and K_4's of the edge-decomposition $\mathcal{D}$ of G, respectively. Then the number of triangles, 2-paths and K_4's in $\mathcal{D}$ is $\alpha_1 |G|/3$, $\alpha_2 |G|/2$ and $(1-\alpha_1-\alpha_2)|G|/6$ respectively. Therefore, using Claim 3.19, Claim 3.20 and Claim 3.21, the total number of (ordered) good 3-paths in G is at most

$$\frac{\alpha_1}{3}|G|(8n+C_1\sqrt{n})+\frac{\alpha_2}{2}|G|(4n+C_2\sqrt{n})+\frac{(1-\alpha_1-\alpha_2)}{6}|G|(6n+C_3\sqrt{n})\leq$$
$$\leq |G|\,n\left(\frac{8\alpha_1}{3}+2\alpha_2+(1-\alpha_1-\alpha_2)\right)+(C_1+C_2+C_3)\sqrt{n}\,|G|=$$
$$=\frac{n^2\overline{d}_G}{2}\left(\frac{5\alpha_1+3\alpha_2+3}{3}\right)+(C_1+C_2+C_3)\frac{n^{3/2}\overline{d}_G}{2}.$$

Combining this with the fact that the number of good 3-paths is at least $n\overline{d}_G^3-C_0n^{3/2}\overline{d}_G$ (see Claim 3.18), we get

$$n\overline{d}_G^3-C_0n^{3/2}\overline{d}_G\leq\frac{n^2\overline{d}_G}{2}\left(\frac{5\alpha_1+3\alpha_2+3}{3}\right)+(C_1+C_2+C_3)\frac{n^{3/2}\overline{d}_G}{2}.$$

Rearranging and dividing by $n\overline{d}_G$ on both sides, we get

$$\overline{d}_G^2\leq\left(\frac{5\alpha_1+3\alpha_2+3}{6}\right)n+\frac{(C_1+C_2+C_3)}{2}\sqrt{n}+C_0\sqrt{n}.$$

Using the fact that $(5\alpha_1+3\alpha_2+3)/6\geq 1/2$, it follows that

$$\overline{d}_G^2\leq\left(\frac{5\alpha_1+3\alpha_2+3}{6}\right)n\left(1+\frac{(C_1+C_2+C_3)+2C_0}{\sqrt{n}}\right).$$

So letting $C_4=(C_1+C_2+C_3)+2C_0$ we have,

$$\overline{d}_G\leq\sqrt{1+\frac{C_4}{\sqrt{n}}}\sqrt{\frac{5\alpha_1+3\alpha_2+3}{6}}\sqrt{n}<\left(1+\frac{C_4}{2\sqrt{n}}\right)\sqrt{\frac{5\alpha_1+3\alpha_2+3}{6}}\sqrt{n},\tag{3.12}$$

for large enough n. By Claim 3.9, we have

$$|H|\leq\frac{\alpha_1}{3}|G|+\frac{\alpha_2}{2}|G|+\frac{2(1-\alpha_1-\alpha_2)}{3}|G|=\frac{4-2\alpha_1-\alpha_2}{6}\frac{n\overline{d}_G}{2}.$$

Combining this with (3.12) we get

$$|H|\leq\left(1+\frac{C_4}{2\sqrt{n}}\right)\frac{(4-2\alpha_1-\alpha_2)}{12}\sqrt{\frac{5\alpha_1+3\alpha_2+3}{6}}n^{3/2},$$

for sufficiently large n. So we have

$$\mathrm{ex}_3(n, C_5) \le (1+o(1))\frac{(4-2\alpha_1-\alpha_2)}{12}\sqrt{\frac{5\alpha_1+3\alpha_2+3}{6}}n^{3/2}.$$

The right hand side is maximized when $\alpha_1 = 0$ and $\alpha_2 = 2/3$, so we have

$$\mathrm{ex}_3(n, C_5) \le (1+o(1))\frac{4-2/3}{12}\sqrt{\frac{5}{6}}n^{1.5} < (1+o(1))0.2536n^{3/2}.$$

This finishes the proof.

3.3 C_5-free linear hypergraphs: Proof of the upper bound in Theorem 3.3

Let H be a 3-uniform linear hypergraph on n vertices containing no C_5. Let d and d_{max} denote the average degree and maximum degree of a vertex in H, respectively. We will show that we may assume H has minimum degree at least $d/3$. Indeed, if there is a vertex whose degree less than one-third of the average degree in the hypergraph, we delete it and all the hyperedges incident to it. Notice that this will not decrease the average degree. We repeat this procedure as long as we can and eventually we obtain a (non-empty) hypergraph H' with $n' \le n$ vertices and average degree $d' \ge d$ and minimum degree at least $d/3$. It is easy to see that if $d' \le \sqrt{n'/3 + C}$ then $d \le \sqrt{n/3 + C}$ (for a constant $C > 0$) proving Theorem 3.3. So from now on we will assume H has minimum degree at least $d/3$. Our goal is to upper bound d.

The following claim shows that for any vertex v, the number of hyperedges $h \in E(H)$ with $\left|h \cap N_1^H(v)\right| \ge 2$ is small provided $d(v)$ is small. This is useful for proving Claim 3.23. Using this and the fact that the minimum degree is at least $d/3$, we will show in Claim 3.25 that we may assume the maximum degree in H is small.

Claim 3.22. *Let $v \in V(H)$. Then the number of hyperedges $h \in E(H)$ with $\left|h \cap N_1^H(v)\right| \ge 2$ is at most $6d(v)$.*

Proof of Claim 3.22. We construct an auxiliary graph G_1 whose vertex set is $N_1^H(v)$ in the following way: From each hyperedge $h \in E(H)$ with $\left|h \cap N_1^H(v)\right| \ge 2$ and $v \notin h$, we select exactly one pair $xy \subset h \cap N_1^H(v)$ arbitrarily. We claim that there is no 7-vertex path in G_1. Suppose for the sake of a contradiction that there is a path $v_1v_2v_3v_4v_5v_6v_7$ in G_1. Then, one of the two hyperedges v_1v_4v, v_4v_7v is not in $E(H)$ as the hypergraph is linear. Suppose without loss of generality that $v_1v_4v \notin E(H)$, so there are two different hyperedges h, h' such that $v_1, v \in h$ and $v_4, v \in h'$. These two hyperedges together with the 3 hyperedges containing v_1v_2, v_2v_3, v_3v_4 create a five cycle in H (note that they are different by our construction), a contradiction. So there is no path on seven vertices in G_1 and so by Erdős-Gallai theorem, G_1 contains at most $\frac{7-2}{2}\left|V(G_1)\right| \le 2.5(2d(v)) = 5d(v)$ edges, which implies that the number of hyperedges $h \in E(H)$ with $\left|h \cap N_1^H(v)\right| \ge 2$ is at most $5d(v) + d(v) = 6d(v)$. □

Using the previous claim we will show the following claim.

Claim 3.23. *Let $v \in V(H)$. Then,*

$$\left|N_2^H(v)\right| \geq \sum_{x \in N_1^H(v)} d(x) - 18d(v).$$

Proof of Claim 3.23. First let us count the number of hyperedges $h \in E(H)$ such that $\left|h \cap N_1^H(v)\right| = 1$ and $\left|h \cap N_2^H(v)\right| = 2$. Let $G_2 = (N_2^H(v), E(G_2))$ be an auxiliary graph whose edge set $E(G_2) = \{xy \mid \exists h \in E(H), \left|h \cap N_1^H(v)\right| = 1, \left|h \cap N_2^H(v)\right| = 2 \text{ and } x, y \in h \cap N_2^H(v)\}$. Let $h_1, h_2, \ldots, h_{d(v)}$ be the hyperedges containing v. Now we color an edge $xy \in E(G_2)$ with the color i if $x, y \in h$ and $h \cap h_i \neq \emptyset$. Since the hypergraph is linear this gives a coloring of all the edges of G_2.

Claim 3.24. *If there are three edges $ab, bc, cd \in E(G_2)$ (where a might be the same as d), then the color of ab is the same as the color of cd.*

Proof of Claim 3.24. Suppose that they have different colors i and j respectively. Then, the hyperedges in H containing ab, bc, cd, together with h_i and h_j form a five cycle, a contradiction. □

We claim that G_2 is triangle-free. Suppose for the sake of a contradiction that there is a triangle, say abc, in G_2. Then by Claim 3.24 it is easy to see that all the edges of this triangle must have the same color, say color i. Therefore, at least two of the three hyperedges of H containing ab, bc, ca must contain the same vertex of h_i. This is impossible since H is linear.

We claim that if $v_1v_2v_3 \ldots v_k$ is a cycle of length $k \geq 4$ in G_2, then every vertex in it has degree exactly 2. Suppose without loss of generality that $v_3w \in E(G_2)$ where $w \neq v_2$, $w \neq v_4$. Since G_2 is triangle free, $w \neq v_1$ and $w \neq v_5$ (note that if $k = 4$, then $v_5 = v_1$). By Claim 3.24, the color of v_1v_2 is the same as the colors of v_3v_4 and v_3w. Also, the color of v_4v_5 is the same as the colors of v_3w and v_2v_3. This implies that the edges v_2v_3, v_3w, v_3v_4 must have the same color, which is a contradiction since the hypergraph is linear. Thus, G_2 is a disjoint union of cycles and trees. So $|E(G_2)| \leq |V(G_2)| = \left|N_2^H(v)\right|$.

Since $\sum_{x \in N_1^H(v)} d(x)$ is at most the number of edges in G_2 plus three times the number of hyperedges $h \in E(H)$ with $\left|h \cap N_1^H(v)\right| \geq 2$, applying Claim 3.22 we have

$$\sum_{x \in N_1^H(v)} d(x) \leq \left|N_2^H(v)\right| + 3(6d(v)),$$

completing the proof of the claim. □

Using the above claim we will show Theorem 3.3 holds if $d_{max} > 6d$. We do not optimize the constant multiplying d here.

Claim 3.25. *We may assume $d_{max} \leq 6d$ for large enough n (i.e., whenever $n \geq 34992$).*

Proof. Suppose that $v \in V(H)$ and $d(v) = d_{max} > 6d$. Recall that H has minimum degree at least $\frac{d}{3}$. Then by Claim 3.23,

$$\left|N_2^H(v)\right| \geq \sum_{x \in N_1^H(v)} d(x) - 18d(v) \geq \frac{d}{3}\left|N_1^H(v)\right| - 18d(v) =$$

$$= \frac{d}{3}(2d(v)) - 18d(v) = \left(\frac{2d}{3} - 18\right) \cdot d(v) > \left(\frac{2d}{3} - 18\right) \cdot 6d \geq 3d^2$$

if $d > 108$. That is, if $d > 108$, then $3d^2 \leq \left|N_2^H(v)\right| \leq n$ which implies that

$$|E(H)| = \frac{nd}{3} \leq \frac{1}{3\sqrt{3}} n^{3/2},$$

as required. On the other hand, if $d \leq 108$, then

$$|E(H)| = \frac{nd}{3} \leq 36n \leq \frac{1}{3\sqrt{3}} n^{3/2}$$

for $n \geq 34992$, proving Theorem 3.3. □

In the next definition, for each hyperedge of H we identify a subhypergraph of H corresponding to this hyperedge. (We will later see that this subhypergraph has a negligible fraction of the hyperedges of H.)

Definition 3.26. *For $abc \in E(H)$, the subhypergraph H'_{abc} of H consists of the hyperedges $h = uvw \in E(H)$ such that $h \cap \{a, b, c\} = \emptyset$ and h satisfies at least one of the following properties.*

1. *$\exists x \in \{a, b, c\}$ such that $\left|h \cap N_1^H(x)\right| \geq 2$.*
2. *$h \cap (N_1^H(a) \cap N_1^H(b) \cap N_1^H(c)) \neq \emptyset$.*
3. *$\{x, y, z\} = \{a, b, c\}$ and $u \in N_1^H(x) \cap N_1^H(y)$ and $v \in N_1^H(z)$.*

Definition 3.27. *Let H_{abc} be the subhypergraph of H defined by $V(H_{abc}) = V(H)$ and $E(H_{abc}) = E(H) \setminus E(H'_{abc})$. That is, H_{abc} is the hypergraph obtained after deleting all the hyperedges of H which are in $E(H'_{abc})$.*

The following claim shows that the number of hyperedges in H'_{abc} is small.

Claim 3.28. *Let $abc \in E(H)$. Then*

$$|E(H'_{abc})| \leq 25d_{max}.$$

Proof. By Claim 3.22, the number of hyperedges $h \in E(H)$ satisfying property 1 of Definition 3.26 is at most

$$6d(a) + 6d(b) + 6d(c) \leq 18d_{max}.$$

Now we estimate the number of hyperedges satisfying property 2 of Definition 3.26 . First let us show that $\left|N_1^H(a) \cap N_1^H(b) \cap N_1^H(c)\right| \leq 1$ which implies that the number of hyperedges satisfying property 2 of Definition 3.26 is at most d_{max}. Assume for the sake of a contradiction that $\{u, v\} \subseteq N_1^H(a) \cap N_1^H(b) \cap N_1^H(c)$. Then by linearity of H, it is impossible that $uva, uvb, uvc \in E(H)$. Without loss of generality, assume that $uva \notin E(H)$. Then it is easy to see that the pairs ua, av, vc, cb, bu are contained in distinct hyperedges by linearity of H, creating a C_5 in H, a contradiction.

Now we estimate the number of hyperedges satisfying property 3 of Definition 3.26. Fix x, y, z such that $\{x, y, z\} = \{a, b, c\}$. We will show that for each $v \in N_1^H(z)$, there is at most one hyperedge containing v and a vertex from $N_1^H(x) \cap N_1^H(y)$. Assume for the sake of a contradiction that there are two different hyperedges $u_1vw_1, u_2vw_2 \in E(H)$ such that $u_1, u_2 \in N_1^H(x) \cap N_1^H(y)$ and $v \in N_1^H(z)$. Now it is easy to see that the pairs $u_1x, xy, yu_2, u_2v, vu_1$ are contained in five distinct hyperedges since H is linear and u_1vw_1, u_2vw_2 are disjoint from abc, so there is a C_5 in H, a contradiction. So for each choice of $z \in \{a, b, c\}$ the number of hyperedges satisfying property 3 of Definition 3.26 is at most $\left|N_1^H(z)\right|$. So the total number of hyperedges satisfying property 3 of Definition 3.26 is at most

$$\left|N_1^H(a)\right| + \left|N_1^H(b)\right| + \left|N_1^H(c)\right| \leq 2(d(a) + d(b) + d(c)) \leq 6d_{max}.$$

Adding up these estimates, we get the desired bound in our claim. □

A 3-*link* in H is a set of 3 hyperedges $h_1, h_2, h_3 \in E(H)$ such that $h_1 \cap h_2 \neq \emptyset$, $h_2 \cap h_3 \neq \emptyset$ and $h_1 \cap h_3 = \emptyset$. The hyperedges h_1 and h_3 are called *terminal* hyperedges of this 3-link. (Notice that a given 3-link defines four different Berge paths because each end vertex can be chosen in two ways. Also note that a 3-link is simply the set of hyperedges of a linear path of length three.)

Given a hypergraph H and $abc \in E(H)$, let $p_{abc}(H)$ denote the number of 3-links in H in which abc is a terminal hyperedge and let $p(H)$ denote the total number of 3-links in H. Notice

$$p(H) = \frac{1}{2} \sum_{abc \in E(H)} p_{abc}(H).$$

In Section 3.3.1, we prove an upper bound on $p(H)$ and in Section 3.3.2, we prove a lower bound on $p(H)$ and combine it with the upper bound to obtain the desired bound on d.

3.3.1 Upper bounding $p(H)$

For any given $abc \in E(H)$, the following claim upper bounds the number of 3-links in H in which abc is a terminal hyperedge by a little bit more than $2\,|V(H)|$.

Claim 3.29. *Let $abc \in E(H)$. Then,*

$$p_{abc}(H) \leq 2\,|V(H)| + 273d_{max}.$$

Proof of Claim 3.29. First we show that most of the 3-links of H are in H_{abc}.

Claim 3.30. *We have,*

$$p_{abc}(H) \leq p_{abc}(H_{abc}) + 225d_{max}.$$

Proof. Consider $h \in E(H) \setminus E(H_{abc}) = E(H'_{abc})$. Note that $h \cap \{a,b,c\} = \emptyset$. The number of 3-links containing both abc and h is at most 9 since the number of hyperedges in H that intersect both h and abc is at most 9 as H is linear. Therefore the total number of 3-links in H containing abc and a hyperedge of $E(H) \setminus E(H_{abc})$ is at most $9\,|E(H'_{abc})| \leq 9(25d_{max}) = 225d_{max}$ by Claim 3.28 which implies that $p_{abc}(H) \leq p_{abc}(H_{abc}) + 225d_{max}$, as required. □

For $x \in \{a,b,c\}$, let H_x be a subhypergraph of H_{abc} whose edge set is $E(H_x) = E_1^x \bigcup E_2^x$ where $E_1^x = \{h \in E(H_{abc}) \mid x \in h \text{ and } h \neq abc\}$ and $E_2^x = \{h \in E(H_{abc}) \mid \exists h' \in E_1^x, x \notin h \text{ and } h \cap h' \neq \emptyset\}$ and its vertex set is $V(H_x) = \{v \in V(H_{abc}) \mid \exists h \in E(H_x) \text{ and } v \in h\}$. Note that $|E_1^x| = d^{H_x}(x) = d^H(x) - 1$ and every hyperedge in E_1^x contains exactly two vertices of $N_1^{H_x}(x)$ and every hyperedge in E_2^x contains one vertex of $N_1^{H_x}(x)$ and two vertices of $N_2^{H_x}(x)$ because hyperedges containing more than one vertex of $N_1^{H_x}(x)$ do not belong to H_{abc} (since they are in H'_{abc} by property 1 of Definition 3.26) and thus, do not belong to H_x.

We will show that the number of ordered pairs (x,h) such that $x \in \{a,b,c\}$ and $h \in E_2^x$ is equal to $p_{abc}(H_{abc})$ by showing a bijection between the set of ordered pairs (x,h) such that $x \in \{a,b,c\}$ and $h \in E_2^x$ and the set of 3-links in H_{abc} where abc is a terminal hyperedge. To each 3-link abc, h', h in H_{abc} where $abc \cap h = \emptyset$ and $h' \cap abc = \{x\}$, let us associate the ordered pair (x,h). Clearly $x \in \{a,b,c\}$ and $h \in E_2^x$. Now consider an ordered pair (x,h) where $x \in \{a,b,c\}$ and $h \in E_2^x$. Then h contains exactly one vertex $u \in N_1^{H_x}(x)$, so there is a unique hyperedge $h' \in E(H)$ containing the pair ux. Therefore, there is a unique 3-link in H_{abc} associated to (x,h), namely abc, h', h, establishing the required bijection. So,

$$p_{abc}(H_{abc}) = |\{(x,h) \mid x \in \{a,b,c\}, h \in E_2^x\}| = \sum_{x \in \{a,b,c\}} |E_2^x|\,. \tag{3.13}$$

Now our aim is to upper bound $p_{abc}(H_{abc})$ in terms of $\sum_{x\in\{a,b,c\}} \left|N_2^{H_x}(x)\right|$, which will be upper bounded in Claim 3.31.

Substituting $v = x$ and $H = H_x$ in Claim 3.23, we get, $\left|N_2^{H_x}(x)\right| \geq \sum_{y \in N_1^{H_x}(x)} d^{H_x}(y) - 18d^{H_x}(x)$ for each $x \in \{a,b,c\}$. Now since $\sum_{y\in N_1^{H_x}(x)} d(y) = 2\,|E_1^x| + |E_2^x|$, we have $\left|N_2^{H_x}(x)\right| \geq 2\,|E_1^x| + |E_2^x| - 18d^{H_x}(x)$. So by (3.13),

$$\sum_{x\in\{a,b,c\}} \left|N_2^{H_x}(x)\right| \geq \sum_{x\in\{a,b,c\}} (2\,|E_1^x| + |E_2^x| - 18d^{H_x}(x)) = \sum_{x\in\{a,b,c\}} (2\,|E_1^x| - 18d^{H_x}(x)) + p_{abc}(H_{abc}).$$

Since $|E_1^x| = d^{H_x}(x) = d^H(x) - 1$, we have $2\,|E_1^x| - 18d^{H_x}(x) = -16(d^H(x) - 1)$. So,

$$\sum_{x\in\{a,b,c\}} \left|N_2^{H_x}(x)\right| \geq -16 \sum_{x\in\{a,b,c\}} (d^H(x)-1)+p_{abc}(H_{abc}) \geq -48(d_{max}-1)+p_{abc}(H_{abc}). \quad (3.14)$$

Now we want to upper bound $\sum_{x\in\{a,b,c\}} \left|N_2^{H_x}(x)\right|$ by $2\,|V(H)|$.

Claim 3.31. *Each vertex $v \in V(H)$ belongs to at most two of the sets $N_2^{H_a}(a), N_2^{H_b}(b), N_2^{H_c}(c)$. So*

$$\sum_{x\in\{a,b,c\}} \left|N_2^{H_x}(x)\right| \leq 2\,|V(H)|\,.$$

Proof. Suppose for the sake of a contradiction that there exists a vertex $v \in V(H)$ which is in all three sets $N_2^{H_a}(a), N_2^{H_b}(b), N_2^{H_c}(c)$. Then for each $x \in \{a,b,c\}$, there exists $h_x \in E_2^x$ such that $v \in h_x$.

First let us assume $h_a = h_b = h_c = h$ and let $h_x \cap N_1^{H_x}(x) = \{v_x\}$ for each $x \in \{a,b,c\}$. If $v_a = v_b = v_c$ then $h \cap (N_1^H(a) \cap N_1^H(b) \cap N_1^H(c)) \neq \emptyset$, so by property 2 of Definition 3.26, $h \in E(H'_{abc})$ so $h \notin E(H_{abc}) \supseteq E_2^x$, a contradiction. If $v_x = v_y \neq v_z$ for some $\{x,y,z\} = \{a,b,c\}$ then by property 3 of Definition 3.26, $h \notin E(H_{abc}) \supseteq E_2^x$, a contradiction again. Therefore, v_a, v_b, v_c are distinct. Moreover, for each $x \in \{a,b,c\}$, $v_x \in N_1^{H_x}(x)$ and $v \in N_2^{H_x}(x)$. However, since $N_1^{H_x}(x)$ and $N_2^{H_x}(x)$ are disjoint for each $x \in \{a,b,c\}$ by definition (see the Notation section for the precise definition of first and second neighborhoods), v is different from v_a, v_b and v_c. So $v, v_a, v_b, v_c \in h$, a contradiction since h is a hyperedge of size 3.

So there exist $x, y \in \{a,b,c\}$ such that $h_x \neq h_y$. Also, there exist $h'_x \in E_1^x, h'_y \in E_1^y$ such that $h_x \cap h'_x \neq \emptyset$ and $h_y \cap h'_y \neq \emptyset$. Now it is easy to see that the hyperedges $h_x, h_y, h'_x, h'_y, abc$ form a C_5, a contradiction, proving the claim. □

So by Claim 3.31, $\sum_{x\in\{a,b,c\}} \left|N_2^{H_x}(x)\right| \leq 2\,|V(H)|$. Combining this with (3.14), we get

$$p_{abc}(H_{abc}) - 48(d_{max}-1) \leq \sum_{x\in\{a,b,c\}} \left|N_2^{H_x}(x)\right| \leq 2\,|V(H)|\,. \quad (3.15)$$

Therefore, by Claim 3.30 and the above inequality, we have

$$p_{abc}(H) \leq p_{abc}(H_{abc}) + 225d_{max} \leq 2\,|V(H)| + 48(d_{max}-1) + 225d_{max} \leq 2\,|V(H)| + 273d_{max},$$

completing the proof of Claim 3.29. □

So by Claim 3.29, we have

$$p(H) = \frac{1}{2} \sum_{abc\in E(H)} p_{abc}(H) \leq \frac{1}{2}(2\,|V(H)| + 273d_{max})\,|E(H)|\,. \quad (3.16)$$

By Claim 3.25, we can assume $d_{max} \leq 6d$. Using this in the above inequality we obtain,

$$p(H) \leq \frac{1}{2}(2\,|V(H)| + 1638d)\,|E(H)| = (n + 819d)\frac{nd}{3}. \quad (3.17)$$

3.3.2 Lower bounding $p(H)$

We introduce some definitions that are needed in the rest of our proof where we establish a lower bound on $p(H)$ and combine it with the upper bound in (3.17).

A *walk* of length k in a graph is a sequence $v_0e_0v_1e_1\dots v_{k-1}e_{k-1}v_k$ of vertices and edges such that $e_i = v_iv_{i+1}$ for $0 \leq i < k$. For convenience we simply denote such a walk by $v_0v_1\dots v_{k-1}v_k$. A walk is called unordered if $v_0v_1\dots v_{k-1}v_k$ and $v_kv_{k-1}\dots v_1v_0$ are considered as the same walk. From now on, unless otherwise stated, we only consider unordered walks. A *path* is a walk with no repeated vertices or edges. Blakley and Roy [7] proved a matrix version of Hölder's inequality, which implies that any graph G with average degree d^G has at least as many walks of a given length as a d^G-regular graph on the same number of vertices.

We will now prove a lower bound on $p(H)$. Consider the shadow graph ∂H of H. The number of edges in ∂H is equal to $3|E(H)| = 3 \cdot \frac{nd}{3} = nd$. Then the average degree of a vertex in ∂H is $d^{\partial H} = 2d$, and the maximum degree $\Delta^{\partial H}$ in ∂H is at most $2d_{max} \leq 12d$ by Claim 3.25. Applying the Blakley-Roy inequality [7] to the graph ∂H, we obtain that there are at least $\frac{1}{2}n(d^{\partial H})^3$ (unordered) walks of length 3 in ∂H. Then there are at least

$$\frac{1}{2}n(d^{\partial H})^3 - 3n(\Delta^{\partial H})^2$$

paths of length 3 in ∂H as there are at most $3n(\Delta^{\partial H})^2$ walks that are not paths. Indeed, if $v_1v_2v_3v_4$ is a walk that is not a path, then there exists a repeated vertex v in the walk such that either $v_1 = v_3 = v$ or $v_2 = v_4 = v$ or $v_1 = v_4 = v$. Since v can be chosen in n ways and the other two vertices of the walk are adjacent to v, we can choose them in at most $(\Delta^{\partial H})^2$ different ways.

A path in ∂H is called a rainbow path if the edges of the path are contained in distinct hyperedges of H. If a path $abcd$ is not rainbow then there are two (consecutive) edges in it that are contained in the same hyperedge of H. So there are two hyperedges $h, h' \in E(H)$, $h \cap h' \neq \emptyset$ such the path $abcd$ is contained in the 2-shadow of h, h'. Now we estimate the number of non-rainbow paths.

We can choose these pairs $h, h' \in E(H)$ in $\sum_{v\in V(H)} \binom{d^H(v)}{2}$ ways and for a fixed pair $h, h' \in E(H)$, it is easy to see that the path $abcd$ can chosen in 8 different ways in the 2-shadow of h, h'. Therefore, the number of non-rainbow paths in ∂H is at most

$$\sum_{v\in V(H)} 8\binom{d^H(v)}{2} \leq 4n(d_{max})^2 \leq 4n(6d)^2 = 144nd^2.$$

So the number of rainbow paths in ∂H is at least

$$\frac{1}{2}n(d^{\partial H})^3 - 3n(\Delta^{\partial H})^2 - 144nd^2 = \frac{1}{2}n(2d)^3 - 3n(12d)^2 - 144nd^2 = 4nd^3 - 576nd^2.$$

Since each 3-link in H produces 4 rainbow paths in ∂H, the number of rainbow paths in ∂H is $4p(H)$. So, $4p(H) \geq 4nd^3 - 576nd^2$. That is,

$$p(H) \geq nd^3 - 144nd^2.$$

Combining this with (3.17), we get

$$nd^3 - 144nd^2 \le p(H) \le (n + 819d)\frac{nd}{3}.$$

Simplifying, we get $d^2 - 144d \le (n + 819d)/3$. That is,

$$d \le \sqrt{\frac{n}{3} + \frac{173889}{4}} + \frac{417}{2}.$$

So,

$$|E(H)| = \frac{nd}{3} \le \frac{n}{3} \cdot \left(\sqrt{\frac{n}{3} + \frac{173889}{4}} + \frac{417}{2}\right) = \frac{1}{3\sqrt{3}} n^{3/2} + O(n),$$

completing the proof of Theorem 3.3.

3.4 C_4-free linear hypergraphs: Proof of Theorem 3.4

Let H be a 3-uniform linear hypergraph on n vertices containing no (Berge) C_4. Let d denote the average degree of a vertex in H.

Outline of the proof: Our plan is to first upper bound $\sum_{x\in N_1^H(v)} 2d(x)$ for each fixed $v \in V(H)$, which as the following claim shows, is not much more than n. Then we estimate $\sum_{v\in V(H)} \sum_{x\in N_1^H(v)} 2d(x)$ in two different ways to get the desired bound on d.

Claim 3.32. *For every $v \in V(H)$, we have*

$$\sum_{x\in N_1^H(v)} 2d(x) \le n + 12d(v).$$

Proof. First we show that most of the hyperedges incident to $x \in N_1^H(v)$ contain only one vertex from $N_1^H(v)$.

Claim 3.33. *For any given $x \in N_1^H(v)$, the number of hyperedges $h \in E(H)$ containing x such that $\left|h \cap N_1^H(v)\right| \ge 2$ is at most 3.*

Proof. Suppose for a contradiction that there is a vertex $x \in N_1^H(v)$ which is contained in 4 hyperedges h such that $\left|h \cap N_1^H(v)\right| \ge 2$. One of them is the hyperedge containing x and v. Let h_1, h_2, h_3 be the other 3 hyperedges. Then it is easy to see that two of these hyperedges intersect two different hyperedges incident to v, and these four hyperedges form a C_4 in H, a contradiction. □

For each $x \in N_1^H(v)$, let $E_x = \{h \in E(H) \mid h \cap N_1^H(v) = \{x\}\}$. Note that any hyperedge of E_x does not contain v, so it contains exactly two vertices from $N_2^H(v)$. Let $S_x = \{w \in N_2^H(v) \mid \exists h \in E_x \text{ with } w \in h\}$. Then $|S_x| = 2\,|E_x|$ since H is linear. Notice that $|E_x| \ge d(x) - 3$ by Claim 3.33, so

$$|S_x| \ge 2d(x) - 6. \tag{3.18}$$

The following claim shows that the sets $\{S_x \mid x \in N_1^H(v)\}$ do not overlap too much.

Claim 3.34. *Let $x, y \in N_1^H(v)$ be distinct vertices. If $xyv \notin E(H)$ then $S_x \cap S_y = \emptyset$ and if $xyv \in E(H)$ then $|S_x \cap S_y| \leq 2$.*

Proof. Take $x, y \in N_1^H(v)$ with $x \neq y$. Let $h_x, h_y \in E(H)$ be hyperedges incident to v such that $x \in h_x$ and $y \in h_y$. First suppose $h_x \neq h_y$. Then it is easy to see that $S_x \cap S_y = \emptyset$ because otherwise h_x, h_y and the two hyperedges containing xw, yw for some $w \in S_x \cap S_y$ form a C_4, a contradiction.

Now suppose $h_x = h_y$. We claim that $|S_x \cap S_y| \leq 2$. Suppose for the sake of a contradiction that there are 3 distinct vertices $v_1, v_2, v_3 \in S_x \cap S_y$. Then it is easy to see that there exist $i, j \in \{1, 2, 3\}$ such that neither v_iv_jx nor v_iv_jy is a hyperedge in H. So there are two different hyperedges $h_1, h_2 \in E_x$ such that $xv_i \in h_1$ and $xv_j \in h_2$. Similarly there are two different hyperedges $h_1', h_2' \in E_y$ such that $yv_i \in h_1'$ and $yv_j \in h_2'$. As $E_x \cap E_y = \emptyset$, the hyperedges h_1, h_2, h_1', h_2' are distinct and form a C_4, a contradiction. □

We will upper bound $\sum_{x \in N_1^H(v)} |S_x|$. It follows from Claim 3.34 that each vertex $w \in N_2^H(v)$ belongs to at most two of the sets in $\{S_x \mid x \in N_1^H(v)\}$. Moreover, w belongs to two sets $S_p, S_q \in \{S_x \mid x \in N_1^H(v)\}$ only if there exists a unique pair $p, q \in N_1^H(v)$ such that $pqv \in E(H)$ and for any such pair p, q with $pqv \in E(H)$, there are at most 2 vertices w with $w \in S_p, S_q$. So there are at most $2d(v)$ vertices in $N_2^H(v)$ that are counted twice in the summation $\sum_{x \in N_1^H(v)} |S_x|$. That is,

$$\left|N_2^H(v)\right| \geq \sum_{x \in N_1^H(v)} |S_x| - 2d(v). \tag{3.19}$$

As $N_2^H(v)$ and $N_1^H(v)$ are disjoint, we have $n \geq \left|N_2^H(v)\right| + \left|N_1^H(v)\right|$. So by (3.19),

$$n \geq \sum_{x \in N_1^H(v)} |S_x| - 2d(v) + \left|N_1^H(v)\right| = \sum_{x \in N_1^H(v)} |S_x| - 2d(v) + 2d(v) = \sum_{x \in N_1^H(v)} |S_x|. \tag{3.20}$$

Combining this with (3.18), we get

$$n \geq \sum_{x \in N_1^H(v)} (2d(x) - 6) = \sum_{x \in N_1^H(v)} 2d(x) - 6\left|N_1^H(v)\right| = \sum_{x \in N_1^H(v)} 2d(x) - 12d(v), \tag{3.21}$$

completing the proof of Claim 3.32. □

We now estimate $\sum_{v \in V(H)} \sum_{x \in N_1^H(v)} 2d(x)$ in two different ways. On the one hand, by Claim 3.32

$$\sum_{v \in V(H)} \sum_{x \in N_1^H(v)} 2d(x) \leq \sum_{v \in V(H)} (n + 12d(v)) = n^2 + 12nd. \tag{3.22}$$

On the other hand,

$$\sum_{v \in V(H)} \sum_{x \in N_1^H(v)} 2d(x) = \sum_{v \in V(H)} 2d(v) \cdot 2d(v) = \sum_{v \in V(H)} 4d(v)^2 \geq 4nd^2. \tag{3.23}$$

The last inequality follows from the Cauchy-Schwarz inequality. Finally, combining (3.22) and (3.23), we get $4nd^2 \le n^2 + 12nd$. Dividing by n, we have $4d^2 \le n + 12d$, so $d \le \frac{1}{2}(\sqrt{n+9}+3)$. Therefore,

$$|E(H)| = \frac{nd}{3} \le \frac{1}{6}n\sqrt{n+9} + \frac{n}{2},$$

proving Theorem 3.4.

3.5 Proof of Theorem 3.5: Construction

We prove Theorem 3.5 by constructing a linear hypergraph H below, and then we show that it is $\mathcal{C}_{2k+1}^{\text{lin}}$-free. Finally, we count the number of hyperedges in it.

Construction of H**:** Let $G = (V(G), E(G))$ be a $\mathcal{C}_{2k-2}$-free bipartite graph (i.e., girth at least $2k$) on z vertices. Let the two color classes of G be $L = \{l_1, l_2, \ldots l_{z_1}\}$ and $R = \{r_1, r_2, \ldots, r_{z_2}\}$ where $z = z_1 + z_2$.

Now we construct a hypergraph $H = (V(H), E(H))$ based on G. Let q be an integer. For each $1 \le t \le q$, let $L_t = \{l_1^t, l_2^t, \ldots, l_{z_1}^t\}$ and $R_t = \{r_1^t, r_2^t, \ldots, r_{z_2}^t\}$. Let $B = \{v_{i,j} \mid 1 \le i \le z_1, 1 \le j \le z_2 \text{ and } l_i r_j \in E(G)\}$. (Note that $|B| = |E(G)|$ as we only create a vertex in B if the corresponding edge exists in G.) Now let $V(H) = \bigcup_{i=1}^{q} L_i \cup \bigcup_{i=1}^{q} R_i \cup B$ and $E(H) = \{v_{i,j} l_i^t r_j^t \mid v_{i,j} \in B \text{ and } l_i r_j \in E(G) \text{ and } 1 \le t \le q\}$. Clearly H is a linear hypergraph.

Proof that H **is** $\mathcal{C}_{2k+1}^{\text{lin}}$**-free:** Suppose for the sake of a contradiction that H contains $C_{2k'+1}^{\text{lin}}$, a linear cycle of length $2k'+1$ for some $k' \le k$.

Since the basic cycle of $C_{2k'+1}^{\text{lin}}$ is of odd length it must contain at least one vertex in B. (Note that here we used that the length of the linear cycle is odd.)

First let us assume that the basic cycle of $C_{2k'+1}^{\text{lin}}$ contains exactly one vertex $x \in B$. Then $\bigcup_{i=1}^{q} L_i \cup \bigcup_{i=1}^{q} R_i \cup x$ contains all the basic vertices of $C_{2k'+1}^{\text{lin}}$. For $X \subseteq V(H)$, let $H[X]$ denote the subhypergraph in H induced by X. Notice that x is a cut vertex in the 2-shadow of $H[\bigcup_{i=1}^{q} L_i \cup \bigcup_{i=1}^{q} R_i \cup x]$. Therefore, there exists a t such that the basic vertices of $C_{2k'+1}^{\text{lin}}$ belong to $L_t \cup R_t \cup x$. Let xu and xv be the two edges incident to x in the basic cycle of $C_{2k'+1}^{\text{lin}}$. However, by construction the hyperedge containing xu is the same as the hyperedge containing xv, which is impossible since $C_{2k'+1}^{\text{lin}}$ is a linear cycle. Therefore, there are at least two basic vertices of $C_{2k'+1}^{\text{lin}}$ in B.

Let $c_1, c_2, \ldots, c_s$ be the basic vertices of $C_{2k'+1}^{\text{lin}}$ in B and let us suppose that they are ordered such that the subpaths $P_{i,i+1}$ of the basic cycle of $C_{2k'+1}^{\text{lin}}$ from c_i to c_{i+1}, are pairwise edge-disjoint for $1 \le i \le s$ (addition in the subscript is taken modulo s from now on). Note that $s \ge 2$ by the previous paragraph and $s \le k'$ because for each i, the subpath $P_{i,i+1}$ contains at least two edges. It is easy to see that for each $1 \le i \le s$, there exists a t such that $V(P_{i,i+1}) \subseteq L_t \cup R_t \cup \{c_i, c_{i+1}\}$. Let $P'_{i,i+1}$ be a path in G with the edge set

$\{l_\alpha r_\beta \mid l^t_\alpha r^t_\beta \in E(P_{i,i+1}) \text{ for some } t\}$ for $1 \le i \le s$. Clearly, $\left|E(P'_{i,i+1})\right| = |E(P_{i,i+1})| - 2 \ge 0$. For each c_i, there exists $1 \le \alpha_i \le z_1$, $1 \le \beta_i \le z_2$ such that $c_i = v_{\alpha_i,\beta_i}$. Let $e_i = l_{\alpha_i} r_{\beta_i}$ for each $1 \le i \le s$, and let $e^t_i = l^t_{\alpha_i} r^t_{\beta_i}$ for each $1 \le t \le q$. Notice that $P'_{i,i+1}$ is a path in G and $e_i \in E(G)$. Moreover, $P'_{i,i+1}$ is a path between a vertex of e_i and a vertex of e_{i+1} and if $E(P'_{i,i+1}) = \emptyset$, then $e_i \cap e_{i+1} \ne \emptyset$.

Claim 3.35. *The paths $P'_{i,i+1}$ (for $1 \le i \le s$) cannot contain any of the edges e_j (for $1 \le j \le s$). Moreover, for any $1 \le i \ne j \le s$, the paths $P'_{i,i+1}$ and $P'_{j,j+1}$ are edge-disjoint.*

Proof. Assume for the sake of contradiction a path $P'_{i,i+1}$ (for some $1 \le i \le s$) contains an edge e_j (for some $1 \le j \le s$). This implies there exists t with $1 \le t \le q$, such that e^t_j is contained in $P_{i,i+1}$, so e^t_j is contained the basic cycle of $C^{\text{lin}}_{2k'+1}$. Then the (only) hyperedge containing e^t_j, namely $l^t_{\alpha_j} r^t_{\beta_j} v_{\alpha_j,\beta_j} = l^t_{\alpha_j} r^t_{\beta_j} c_j$ is a hyperedge of the linear cycle $C^{\text{lin}}_{2k'+1}$. However, by definition of a linear cycle, the basic cycle must use exactly two vertices of any hyperedge of its linear cycle, a contradiction. Therefore the paths $P'_{i,i+1}$, $1 \le i \le s$, cannot contain any of the edges e_j (for $1 \le j \le s$).

Now we will show that for any $1 \le i \ne j \le s$, $P'_{i,i+1}$ and $P'_{j,j+1}$ are edge-disjoint. Suppose for a contradiction that $l_\alpha r_\beta \in E(P'_{i,i+1}) \cap E(P'_{j,j+1})$ for some $1 \le \alpha \le z_1$ and $1 \le \beta \le z_2$. Then there exist $t \ne t'$ such that $l^t_\alpha r^t_\beta$ and $l^{t'}_\alpha r^{t'}_\beta$ are two disjoint edges of the basic cycle of $C^{\text{lin}}_{2k'+1}$. However, $l^t_\alpha r^t_\beta v_{\alpha,\beta}, l^{t'}_\alpha r^{t'}_\beta v_{\alpha,\beta} \in E(H)$, which is impossible since the hyperedges containing disjoint edges of the basic cycle of a linear cycle must also be disjoint, by the definition of a linear cycle. □

Recall that by definition, the first vertex of $P_{j,j+1}$ is c_j. So the first edge of $P_{j,j+1}$ is contained in a hyperedge of the form $e^t_j \cup c_j$ for some t (indeed all the hyperedges containing c_j are of this form). This means the second vertex of $P_{j,j+1}$ is contained in e^t_j, so the first vertex of $P'_{j,j+1}$ is contained in e_j. Similarly, the last vertex of $P'_{j-1,j}$ is also contained in e_j. Therefore, the last vertex of $P'_{j-1,j}$ and the first vertex of $P'_{j,j+1}$ are both contained in e_j. If these vertices are different, then we call e_j a *connecting edge*. So using Claim 3.35, the edges of $\cup_i E(P'_{i,i+1})$ together with the connecting edges form a circuit $\mathcal{C}$ in G (i.e., a cycle where vertices may repeat but edges do not repeat).

Now we claim that $\mathcal{C}$ is non-empty and contains at most $2k-1$ edges. Indeed, the number of edges of $\mathcal{C}$ is at least $\sum_{i=1}^s \left|E(P'_{i,i+1})\right|$. Moreover, as the number of connecting edges is at most s, the number of edges in $\mathcal{C}$ is at most $\sum_{i=1}^s \left|E(P'_{i,i+1})\right| + s$. Since $\sum_{i=1}^s \left|E(P'_{i,i+1})\right| = \sum_{i=1}^s |E(P_{i,i+1})| - 2s = 2k' + 1 - 2s$, and $2 \le s \le k'$, it is easily seen that $\mathcal{C}$ is non-empty and contains at most $2k' + 1 - s \le 2k' - 1 \le 2k - 1$ edges, as claimed. (Let us remark that here the fact that the length of the linear cycle $C^{\text{lin}}_{2k'+1}$ is odd played a crucial role in ensuring that the circuit $\mathcal{C}$ is non-empty –indeed, if the length is even, it is possible that $E(P'_{i,i+1})$ is empty for each i.)

Since every non-empty circuit contains a cycle, we obtain a cycle of length at most $2k-1$ in G, a contradiction, as desired.

Bounding $\mathrm{ex}^{\text{lin}}_3(n, \mathcal{C}^{\text{lin}}_{2k+1})$ from below: We assumed $\mathrm{ex}_{\text{bip}}(z, \mathcal{C}_{2k-2}) \ge (1+o(1))c(z/2)^\alpha$

for some $c, \alpha > 0$. So there is a $\mathcal{C}_{2k-2}$-free bipartite graph G on z vertices with

$$|E(G)| = (1+o(1))c\left(\frac{z}{2}\right)^{\alpha}. \tag{3.24}$$

Let H be the $\mathcal{C}^{\text{lin}}_{2k+1}$-free hypergraph constructed based on G (as described in the Construction above). Then the number of hyperedges in H is $|E(G)| \cdot q$. So we have

$$\text{ex}_3^{\text{lin}}(n, \mathcal{C}^{\text{lin}}_{2k+1}) \geq |E(H)| = |E(G)| \cdot q \geq |E(G)| \cdot \left\lfloor \frac{n - |E(G)|}{z} \right\rfloor. \tag{3.25}$$

Substituting (3.24) in (3.25) and choosing $z = (1+o(1))\left(\frac{2^{\alpha}(\alpha-1)}{c(2\alpha-1)}\right)^{\frac{1}{\alpha}} n^{\frac{1}{\alpha}}$, we obtain that

$$\text{ex}_3^{\text{lin}}(n, \mathcal{C}^{\text{lin}}_{2k+1}) \geq (1+o(1))\frac{\alpha c}{4\alpha - 2} \cdot \left(\frac{\alpha - 1}{c(2\alpha - 1)}\right)^{1-\frac{1}{\alpha}} n^{2-\frac{1}{\alpha}},$$

completing the proof of Theorem 3.5.

Chapter 4

Triangles in C_5-free graphs and Hypergraphs of Girth Six

4.1 Introduction

Motivated by a conjecture of Erdős [16] on the maximum possible number of pentagons in a triangle-free graph, Bollobás and Győri [8] initiated the study of the natural converse of this problem. Let $ex(n, K_3, C_5)$ denote the maximum possible number of triangles in a graph on n vertices without containing a cycle of length five as a subgraph. Bollobás and Győri [8] showed that

$$(1 + o(1))\frac{1}{3\sqrt{3}}n^{3/2} \leq ex(n, K_3, C_5) \leq (1 + o(1))\frac{5}{4}n^{3/2}. \tag{4.1}$$

Their lower bound comes from the following example: Take a C_4-free bipartite graph G_0 on $n/3 + n/3$ vertices with about $(n/3)^{3/2}$ edges and double each vertex in one of the color classes and add an edge joining the old and the new copy to produce a graph G. Then, it is easy to check that G contains no C_5 and it has $(n/3)^{3/2}$ triangles.

Recently, Füredi and Özkahya [38] gave a simpler proof showing a slighly weaker upper bound of $\sqrt{3}n^{3/2} + O(n)$. Alon and Shikhelman [4] improved these results by showing that

$$ex(n, K_3, C_5) \leq (1 + o(1))\frac{\sqrt{3}}{2}n^{3/2}. \tag{4.2}$$

E., Győri, Methuku and Salia [33] recently showed that

Theorem 4.1. *(E., Győri, Methuku, Salia [33])*

$$ex(n, K_3, C_5) \leq (1 + o(1))\frac{1}{2\sqrt{2}}n^{3/2}.$$

In this chapter our aim is to introduce a new approach and use it to improve two old results and prove a new one. Our approach consists of carefully counting paths of length 5 (or paths of length 3) by making use of the structure of certain subgraphs. Roughly speaking, we are able to efficiently bound the number of 5-paths if its middle edge lies in a dense subgraph (for e.g., in a K_4). We expect this approach to have further applications.

The main result of this chapter improves the previous estimates (4.1), (4.2) and theorem 4.1, on the maximum possible number of triangles in a C_5-free graph, as follows.

Theorem 4.2 (E., Methuku [26]). *We have,*

$$ex(n, K_3, C_5) < (1 + o(1))\frac{1}{3\sqrt{2}}n^{3/2}.$$

Given a hypergraph H, its *2-shadow* is the graph consisting of the edges $\{ab \mid ab \subset e \in E(H)\}$. Applying our approach to the 2-shadow of a hypergraph of girth 6, we prove the following result.

Theorem 4.3 (E., Methuku [26]). *Let H be an r-uniform hypergraph of girth 6. Then*

$$|E(H)| \leq (1 + o(1))\frac{n^{3/2}}{r^{3/2}(r-1)}.$$

Let us mention a related result of Lazebnik and Verstraëte [59] which states the following. If H is an r-uniform hypergraph of girth 5, then

$$|E(H)| \leq (1 + o(1))\frac{n^{3/2}}{r(r-1)}.$$

Note that Theorem 4.3 shows that if a (Berge) cycle of length 5 is also forbidden, then the above bound can be improved by a factor of $\sqrt{r}$. It would be interesting to determine whether there is a matching construction for the bound in Theorem 4.3, at least when $r = 3$.

In Section 4.3.2, we show a close connection between Theorem 4.2 and Theorem 4.3, and prove that the estimate in Theorem 4.2 can be slightly improved using Theorem 4.3. However, to illustrate the main ideas of the proof of Theorem 4.2, we decided to state Theorem 4.2 in a slightly weaker form.

Loh, Tait, Timmons and Zhou [61] introduced the problem of simultaneously forbidding an induced copy of a graph and a (not necessarily induced) copy of another graph. A graph is called induced-F-free if it does not contain an induced copy of F. They asked the following question: What is the largest size of an induced-C_4-free and C_5-free graph on n vertices? They noted that the example showing the lower bound in (4.1) is in fact induced-C_4-free and C_5-free, thus it gives a lower bound of $(1+o(1))\frac{2}{3\sqrt{3}}n^{3/2}$. (If the "induced-$C_4$-free" condition is replaced by "C_4-free" condition, then Erdős and Simonovits [19] showed that the answer is $(1 + o(1))\frac{1}{2\sqrt{2}}n^{3/2}$.) In [31], Győri and the current authors determined (asymptotically) the maximum size of an induced-$K_{s,t}$-free and C_{2k+1}-free graph on n vertices in all the cases except in the case when $s = t = 2$ and $k = 2$ (i.e., the question stated above), and in this case an upper bound of only $n^{3/2}/2$ was proven [31]. Here we show that using our approach one can slightly improve this upper bound.

Theorem 4.4 (E., Methuku [26]). *If a graph G is C_5-free and induced-C_4-free, then*

$$|E(G)| \leq (1 + o(1)) \frac{n^{3/2}}{2\sqrt[10]{2}}.$$

Structure of the chapter: In Section 4.2, we prove Theorem 4.2. In Section 4.3, we prove Theorem 4.3 and show how it can be used to slightly improve Theorem 4.2. Finally in Section 4.4, we prove Theorem 4.4.

Notation: Given a graph G and a vertex v of G, let $N_1(v)$ and $N_2(v)$ denote the first neighborhood and the second neighborhood of v respectively.

For a vertex v of G, let $d(v)$ be the degree of v. The average degree of a graph G is denoted by $d(G)$, or simply d if it is clear from the context. The maximum degree of a graph G is denoted by $d_{max}(G)$ or simply d_{max}.

A *walk* or *path* usually referes to an unordered one, unless specified otherwise. That is, a walk or path $v_1v_1v_2 \dots v_k$ is considered equivalent to $v_kv_{k-1}v_2 \dots v_1$.

4.2 Number of triangles in a C_5-free graph: Proof of Theorem 4.1 and 4.2

Let G be a C_5-free graph with maximum possible number of triangles. We may assume that each edge of G is contained in a triangle, because otherwise, we can delete it without changing the number of triangles. Two triangles T, T' are said to be in the same *block* if they either share an edge or if there is a sequence of triangles $T, T_1, T_2, \dots, T_s, T'$ where each triangle of this sequence shares an edge with the previous one (except the first one of course). It is easy to see that all the triangles in G are partitioned uniquely into blocks. Notice that any two blocks of G are edge-disjoint. Below we will characterize the blocks of G.

A block of the form $\{abc_1, abc_2, \dots, abc_k\}$ where $k \geq 1$, is called a *crown-block* (i.e., a collection of triangles containing the same edge) and a block consisting of all triangles contained in the complete graph K_4 is called a K_4*-block*. See Figure 4.1.

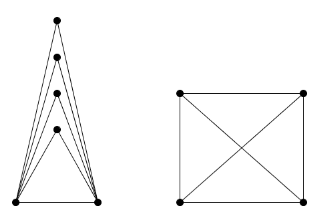

Figure 4.1: An example of a crown-block and a K_4-block

The following claim was proved in [33]. We repeat its proof for completeness.

Claim 4.1. *Every block of G is either a crown-block or a K_4-block.*

Proof. If a block contains only one or two triangles, then it is easy to see that it is a crown-block. So we may assume that a block of G contains at least three triangles and let abc_1, abc_2 be some two triangles in it. We claim that if bc_1x or ac_1x is a triangle in G which is different from abc_1, then $x = c_2$. Indeed, if $x \neq c_2$, then the vertices a, x, c_1, b, c_2 contain a C_5, a contradiction. Similarly, if bc_2x or ac_2x is a triangle in G which is different from abc_2, then $x = c_1$.

Therefore, if ac_i or bc_i (for $i = 1, 2$) is contained in two triangles, then abc_1c_2 forms a K_4. However, then there is no triangle in G which shares an edge with this K_4 and is not contained in it because if there is such a triangle, then it is easy to find a C_5 in G, a contradiction. So in this case, the block is a K_4-block, and we are done.

So we can assume that whenever abc_1, abc_2 are two triangles then the edges ac_1, bc_1, ac_2, bc_2 are each contained in exactly one triangle. Therefore, any other triangle which shares an edge with either abc_1 or abc_2 must contain ab. Let abc_3 be such a triangle. Then applying the same argument as before for the triangles abc_1, abc_3 one can conclude that the edges ac_3, bc_3 are contained in exactly one triangle and so, any other triangle of G which shares an edge with one of the triangles abc_1, abc_2, abc_3 must contain ab again. So by induction, it is easy to see that all of the triangles in this block must contain ab. Therefore, it is a crown-block, as needed. □

4.2.1 Proof of Theorem 4.1

Claim 4.2. *The edges of any C_4 in G are contained in only one block of G.*

Proof. Let $xyzw$ be a 4-cycle in G. Every edge of G is contained in a triangle. So in particular, let xyu be a triangle containing the edge xy. If $u \notin \{x, y, z, w\}$ then $uxwzy$ is a C_5, a contradiction. Therefore, $u = z$ or $u = w$. So either xyz and yzw or xyw and ywz are triangles of G. In both cases, the two triangles share an edge, so they belong to the same block. Hence, all four edges of $xyzw$ lie in the same block. □

We are now ready to prove the theorem using the above claims. We want to select a C_4-free subgraph G_0 of G such that the number of edges in G_0 is the same as the number of triangles in G. By Claim 4.2 the edge set of every C_4 is completely contained in some block of G. So in order to make sure the selected subgraph G_0 is C_4-free, it suffices to make sure the edges selected from each block of G do not contain a C_4, which is done as follows: From each crown-block $\{abc_1, abc_2, \ldots, abc_k\}$, we select the edges $ac_1, ac_2, \ldots, ac_k$ to be in G_0. From each K_4-block $abcd$ we select the edges ab, bc, ac, ad to be in G_0 (since every block is either a crown-block or a K_4-block by Claim 4.1, we have dealt with all the blocks of G). Finally, notice that the number of selected edges in each block is exactly the number of triangles in that block. Moreover, since blocks are edge-disjoint, we never select the same edge twice. Therefore, as every triangle of G is contained in some block, the total number of triangles in G is the same as the number of edges in G_0. On the other hand, as G_0 is C_4-free and also C_5-free (as it is a subgraph of G), we can use the theorem of Erdős and Simonovits [19], which states that the maximum possible number of edges in a graph on n

vertices containing no C_4 or C_5 as a subgraph is at most $\frac{1}{2\sqrt{2}}(1+o(1))n^{3/2}$. So we get that the number of edges in G_0 is at most $\frac{1}{2\sqrt{2}}(1+o(1))n^{3/2}$, completing the proof of Theorem 4.1.

4.2.2 Proof of Theorem 4.2

Edge Decomposition of G: We define a decomposition $\mathcal{D}$ of the edges of G into paths of length 2, triangles and K_4's, as follows: Since each edge of G belongs to a triangle, and all the triangles of G are partitioned into blocks, it follows that the edges of G are partitioned into blocks as well. Moreover, by Claim 4.1, edges of G can be decomposed into crown-blocks and K_4-blocks. We further partition the edges of each crown-block $\{abc_1, abc_2, \ldots, abc_k\}$ (for some $k \geq 1$) into the triangle abc_1 and paths ac_ib where $2 \leq i \leq k$. This gives the desired decomposition $\mathcal{D}$ of $E(G)$.

Claim 4.3. *Let u, v be two non-adjacent vertices of G. Then the number of paths of length 2 between u and v is at most two. Moreover, if uxv and uyv are the paths of length 2 between u and v, then x and y are adjacent.*

Proof. First let us prove the second part of the claim. Since we assumed every edge is contained in a triangle and u and v are not adjacent, there is a vertex $w \neq v$ such that uxw is a triangle. If $w \neq y$, then $uwxvy$ is a C_5, a contradiction. So $w = y$, so x and y are adjacent, as desired.

Now suppose that there are 3 distinct vertices x, y, z such that uxv, uyv, uzv are paths of length 2 between u and v. Then x and y are adjacent by the discussion in the previous paragraph. Therefore $uxyvz$ is a C_5 in G, a contradiction, proving the claim. □

Let $t(v)$ be the number of triangles containing a vertex v and let $t(G) = t = \sum_{v \in V(G)} \frac{t(v)}{n}$. Observe that number of triangles in G is $nt/3$. Our goal is to bound t from above.

First we claim that for any vertex v of G,

$$t(v) \leq d(v) \leq 2t(v). \tag{4.3}$$

Indeed, $d(v) \leq 2t(v)$ simply follows by noting that every edge is in a triangle. Now notice that $t(v)$ is equal to the number of edges contained in the first neighborhood of v (denoted by $N_1(v)$). Moreover, there is no path of length three in the subgraph induced by $N_1(v)$ because otherwise there is a C_5 in G. So by Erdős-Gallai theorem, the number of edges contained in $N_1(v)$ is at most $\frac{3-1}{2}|N_1(v)| = d(v)$. Therefore, $t(v) \leq d(v)$.

Note that by adding up (4.3) for all the vertices $v \in V(G)$ and dividing by n, we get

$$t \leq d \leq 2t. \tag{4.4}$$

Suppose there is a vertex v of G, such that $t(v) < t/3$. Then we may delete v and all the edges incident to v from G to obtain a graph G' such that $t(G') > 3(nt/3 - t/3)/(n-1) = t(G)$. Then it is easy to see that if the theorem holds for G', then it holds for G as well. Repeating this procedure, we may assume that for every vertex v of G, $t(v) \geq t/3$. Therefore, by (4.3), we may assume that the degree of every vertex of G is at least $t/3$.

Claim 4.4. *We may assume that $d_{max}(G) \leq 6\sqrt{3}\sqrt{n}$.*

Proof. Suppose that there is a vertex v such that $d(v) > 6\sqrt{3}\sqrt{n}$. The sum of degrees of the vertices in $N_1(v)$ is at least $\frac{|N_1(v)|t}{3} = \frac{d(v)t}{3}$ as we assumed that the degree of every vertex is at least $t/3$. The number of edges inside $N_1(v)$ is $t(v)$, which is at most $d(v)$ by (4.3). Therefore the number of edges between $N_1(v)$ and $N_2(v)$ is at least $\frac{d(v)t}{3} - 2d(v)$. Now notice that any vertex in $N_2(v)$ is incident to at most two of these edges by Claim 4.3. Therefore, $|N_2(v)| \geq \frac{d(v)t}{6} - d(v)$.

Thus we have,

$$n \geq |N_1(v)| + |N_2(v)| \geq d(v) + \frac{d(v)t}{6} - d(v) = \frac{d(v)t}{6} > \frac{6\sqrt{3}\sqrt{n}t}{6},$$

which implies $t < \sqrt{\frac{n}{3}}$. Therefore, the total number of triangles in G is less than $\frac{n^{3/2}}{3\sqrt{3}}$, proving Theorem 4.2. □

By the Blakley-Roy inequality, the number of (unordered) walks of length five in G is $nd^5/2$. First let us show that most of these walks are paths. Let $v_0v_1v_2v_3v_4v_5$ be a walk that is not a path. Then $v_i = v_j$ for some $i < j$. Fix some $i < j$. Then there are n choices for v_0, and then at most d_{max} choices for every v_k with $k \leq j-1$, then since $v_j = v_i$, there is only choice for v_j and again at most d_{max} choices for every v_k with $k \geq j+1$. So in total the number of walks that are not paths is at most $\binom{6}{2}n(d_{max})^4$ as there are $\binom{6}{2} = 15$ choices for i, j. Thus the number of (unordered) paths of length five in G is at least $nd^5/2 - 15n(d_{max})^4$. From now, we refer to a path of length five as a 5-path.

We say a 5-path $v_0v_1v_2v_3v_4v_5$ is *bad* if there exists an i such that $v_iv_{i+1}v_{i+2}$ is a triangle of G; otherwise it called *good*. Our aim is to show that the number of bad 5-paths is very small. Let $v_0v_1v_2v_3v_4v_5$ be a bad 5-path. Then there is an i so that $v_iv_{i+1}v_{i+2}$ is a triangle. If we fix an i, there are at most $2nt$ choices for $v_iv_{i+1}v_{i+2}$ as each of the $nt/3$ triangles can be ordered in $3! = 6$ ways, and there are at most d_{max} choices for every vertex v_k with $k < i$ or $k > i+2$. There are four choices for i. Therefore, the total number of 5-paths that are bad is at most $8nt(d_{max})^3$. This means that the number of good 5-paths is at least $nd^5/2 - 15n(d_{max})^4 - 8nt(d_{max})^3$. By (4.1), the number of triangles of G is at most $(1+o(1))\frac{5n^{3/2}}{4}$. Since the number of triangles of G is $nt/3$, we have $t \leq \frac{15}{4}(1+o(1))n^{1/2}$. Now using Claim 4.4, it follows that the number of good 5-paths is at least

$$\frac{nd^5}{2} - 15n(6\sqrt{3}\sqrt{n})^4 - 8n\frac{15}{4}n^{1/2}(6\sqrt{3}\sqrt{n})^3 \geq \frac{nd^5}{2} - Cn^3, \tag{4.5}$$

where C is some positive constant.

Now we seek to bound the number of good 5-paths from above. Recall that we defined a decomposition $\mathcal{D}$ of the edges of G into three types of subgraphs: paths of length 2, triangles and K_4's. We distingush three cases depending on which type of subgraph the middle edge of a good 5-path belongs to, and bound the number of good 5-paths in each of those cases separately in the following three claims.

A path of length two (or a 2-path) xyz is called *good* if x and z are not adjacent.

Claim 4.5. *Let abc be a 2-path of the edge-decomposition $\mathcal{D}$. Then the number of good 5-paths in G whose middle edge is either ab or bc is at most n^2.*

Proof. A good 5-path $xypqzw$ whose middle edge is ab or bc contains good 2-paths, xyp, qzw as subpaths (where pq is either ab or bc). Moreover, since $xypqzw$ is a good 5-path and the 2-path abc is contained in the triangle abc (because of the way we defined the decomposition $\mathcal{D}$), it follows that $x, y \notin \{a, b, c\}$ and $z, w \notin \{a, b, c\}$.

Let n_a be the number of good 2-paths in G of the form axy where $x, y \notin \{a, b, c\}$, and let n_b be the number of good 2-paths in G of the form bxy where $x, y \notin \{a, b, c\}$. We define n_c similarly. Then the number of good 5-paths whose middle edge is either ab or bc is at most

$$n_a n_b + n_b n_c = n_b(n_a + n_c) \le \left(\frac{n_a + n_b + n_c}{2}\right)^2.$$

We claim that for any fixed vertex $y \notin \{a, b, c\}$, there are at most two good 2-paths of the form pxy with $p \in \{a, b, c\}$ and $x \notin \{a, b, c\}$. If this claim is true, then $n_a + n_b + n_c \le 2n$, so the right-hand-side of the above inequality is at most n^2, proving Claim 4.5.

It remains to prove this claim. Suppose for a contradiction that there are three such good 2-paths, say, $p_1x_1y, p_2x_2y, p_3x_3y$. Notice that if p_ix_i is disjoint from p_jx_j for some $i, j \in \{1, 2, 3\}$, then $p_ip_jx_jyx_i$ forms a C_5 in G, a contradiction (note that here we used that p_i and p_j are adjacent even when $\{p_i, p_j\} = \{a, c\}$ because of the way we defined $\mathcal{D}$). Thus the edges p_1x_1, p_2x_2, p_3x_3 pair-wise intersect, which implies that either $p_1 = p_2 = p_3 = p$ or $x_1 = x_2 = x_3 = x$ (since $p_1, p_2, p_3 \in \{a, b, c\}$ and $x_1, x_2, x_3 \notin \{a, b, c\}$). The former case is impossible by Claim 4.3 and in the latter case, note that a, b, c, x forms a K_4, but this contradicts the definition of $\mathcal{D}$ since abc was assumed to be a 2-path component of $\mathcal{D}$ and no 2-path of $\mathcal{D}$ comes from a K_4-block of G. □

Claim 4.6. *Let abc be a triangle of the edge-decomposition $\mathcal{D}$. Then the number of good 5-paths in G whose middle edge is either ab, bc, ca is at most $\frac{4n^2}{3}$.*

Proof. The proof is very similar to that of the proof of Claim 4.5. A good 5-path $xypqzw$ whose middle edge is ab, bc, ca contains good 2-paths, xyp, qzw as subpaths. Moreover, since $xypqzw$ is a good 5-path, it follows that $x, y \notin \{a, b, c\}$ and $z, w \notin \{a, b, c\}$.

Let n_a be the number of good 2-paths in G of the form axy where $x, y \notin \{a, b, c\}$, and let n_b, n_c be defined similarly. Then the number of good 5-paths whose middle edge is ab, bc or ca is at most

$$n_a n_b + n_b n_c + n_c n_a \le \frac{(n_a + n_b + n_c)^2}{3}.$$

By the same argument as in the proof of Claim 4.5, it is easy to see that $n_a + n_b + n_c \le 2n$, so the above inequality finishes the proof. □

Claim 4.7. *Let $abcd$ be a K_4 of the edge-decompostion $\mathcal{D}$. Then the number of good 5-paths in G whose middle edge belongs to the K_4 is at most $\frac{3n^2}{2}$.*

Proof. Notice that any good 5-path $xypqzw$ contains good 2-paths, xyp, qzw as subpaths. Suppose the middle edge of $xypqzw$ belongs to the K_4, $abcd$. Then since $xypqzw$ is a good 5-path, it follows that $x, y \notin \{a, b, c, d\}$ and $z, w \notin \{a, b, c, d\}$.

Let n_a be the number of good 2-paths in G of the form axy where $x, y \notin \{a, b, c, d\}$, and let n_b, n_c, n_d be defined similarly. Then the number of good 5-paths whose middle edge belongs to the K_4, $abcd$ is at most

$$\sum_{i,j\in\{a,b,c,d\}} n_i n_j \leq \frac{3}{8}(n_a + n_b + n_c + n_d)^2. \tag{4.6}$$

To see that the above inequality is true one simply needs to expand and rearrange the inequality $\sum_{i,j\in\{a,b,c,d\}}(n_i - n_j)^2 \geq 0$.

Using a similar argument as in the proof of Claim 4.5, it is easy to see that for any fixed vertex $y \notin \{a, b, c, d\}$, there are at most two good 2-paths of the form pxy with $p \in \{a, b, c, d\}$ and $x \notin \{a, b, c, d\}$. This implies that $n_a + n_b + n_c + n_d \leq 2n$, so using (4.6), the proof is complete. □

Now we are ready to bound the number of good 5-paths in G from above. Suppose the number of edges of G is $e(G)$, and let $\alpha_1 e(G)$ and $\alpha_2 e(G)$ be the number of edges of G that are contained in triangles and 2-paths of the edge-decomposition $\mathcal{D}$ of G, respectively. Let $\alpha_1 + \alpha_2 = \alpha$. In other words, $(1-\alpha)e(G)$ edges of G belong to the K_4's in $\mathcal{D}$. Then the number of triangles and 2-paths in $\mathcal{D}$ is at most $\frac{\alpha_1}{3}e(G)$ and $\frac{\alpha_2}{2}e(G)$ respectively and the number of K_4's in $\mathcal{D}$ is at most $\frac{(1-\alpha)}{6}e(G)$. Therefore, using Claim 4.5, Claim 4.6 and Claim 4.7, the total number of good 5-paths in G is at most

$$\frac{\alpha_1}{3}e(G)\frac{4n^2}{3} + \frac{\alpha_2}{2}e(G)n^2 + \frac{(1-\alpha)}{6}e(G)\frac{3n^2}{2} \leq \frac{\alpha}{2}e(G)n^2 + \frac{(1-\alpha)}{4}e(G)n^2 = \frac{(1+\alpha)}{8}n^3 d.$$

Combining this with the fact that the number of good 5-paths is at least $nd^5/2 - Cn^3$ (by (4.5)), we get

$$\frac{nd^5}{2} - Cn^3 \leq \frac{(1+\alpha)}{8}n^3 d,$$

which simplifies to $\frac{d^5}{2} \leq \frac{(1+\alpha)}{8}n^2 d + Cn^2 = (1+o(1))\frac{(1+\alpha)}{8}n^2 d$. Here we used that $d \geq t = \Omega(\sqrt{n})$ (by (4.4)). Therefore,

$$d \leq (1+o(1))\left(\frac{1+\alpha}{4}\right)^{1/4}\sqrt{n}. \tag{4.7}$$

Recall that when defining $\mathcal{D}$ we decomposed the edges of each crown-block into a triangle and 2-paths. This means that the number of triangles of G that belong to crown-blocks of G is at most $\frac{\alpha_1 e(G)}{3} + \frac{\alpha_2 e(G)}{2} \leq \frac{\alpha e(G)}{2}$, and the number of triangles that belong to K_4-blocks of G is at most $\frac{4(1-\alpha)e(G)}{6}$. Therefore, the total number of triangles in G is at most

$$\frac{\alpha e(G)}{2} + \frac{4(1-\alpha)e(G)}{6} = \frac{4-\alpha}{6}e(G) = \frac{(4-\alpha)nd}{12}. \tag{4.8}$$

Now using (4.7), we obtain that the number of triangles in G is at most

$$(1+o(1))\left(\frac{1+\alpha}{4}\right)^{1/4}\frac{(4-\alpha)}{12}n^{3/2}.$$

Now optimizing the coefficient of $n^{3/2}$ over $0 \le \alpha \le 1$, one obtains that it is maximized at $\alpha = 0$, giving the desired upper bound of $(1+o(1))\frac{1}{3\sqrt{2}}n^{3/2}$.

4.3 On hypergraphs of girth 6 and further improvement

In this section we will first study r-uniform hypergraphs of girth 6, and prove Theorem 4.3. Then we use Theorem 4.3 to further (slightly) improve the estimate in Theorem 4.2 on the number of triangles in a C_5-free graph.

4.3.1 Girth 6 hypergraphs: Proof of Theorem 4.3

Let d be the average degree of H. Our aim is to show that $d \le \frac{\sqrt{n}}{\sqrt{r}(r-1)}$. If a vertex has degree less than d/r, then we may delete it and the edges incident to it without decreasing the average degree. So we may assume that the minimum degree of H, $\delta(H) \ge d/r$.

Suppose there is a vertex v of degree $c\sqrt{n}$ for some constant c. Then the first neighborhood $N_1^H(v) := \{x \in V(H)\backslash\{v\} \mid v, x \in h \text{ for some } h \in E(H)\}$ has size more than $c\sqrt{n}(r-1)$ (since H is linear), and the second neighborhood $N_2^H(v) = \{x \in V(H) \setminus (N_1^H(v) \cup \{v\}) \mid \exists h \in E(H) \text{ such that } x \in h \text{ and } h \cap N_1^H(v) \neq \emptyset\}$ has size more than

$$c\sqrt{n}(r-1) \times \delta(H)(r-1) \ge c\sqrt{n}(r-1) \times \frac{d(r-1)}{r} = \frac{c\sqrt{n}(r-1)^2 d}{r}.$$

Note that here we used that H has no cycles of length at most four. On the other hand, since $\left|N_2^H(v)\right| \le n$, we have $\frac{c\sqrt{n}(r-1)^2 d}{r} \le n$, implying that $d \le \frac{r}{(r-1)^2 c}\sqrt{n}$. So if $c > \frac{r^{3/2}}{r-1}$, we have the desired bound on d. Thus, we may assume $c \le \frac{r^{3/2}}{r-1}$, which proves that the maximum degree of H, $d_{max} \le \frac{r^{3/2}}{r-1}\sqrt{n}$.

Let ∂H denote the 2-shadow graph of H. Let $d^{\partial H}$ and $d^{\partial H}_{max}$ denote the average degree and maximum degree of ∂H, respectively. Note that since H is linear, $d^{\partial H} = (r-1)d$ and $d^{\partial H}_{max} = (r-1)d_{max} \le r^{3/2}\sqrt{n}$.

We say a 3-path $v_0v_1v_2v_3$ in ∂H is *bad* if either $\{v_0, v_1, v_2\} \subseteq h$ or $\{v_1, v_2, v_3\} \subseteq h$ for some hyperedge $h \in E(H)$; otherwise it is *good*.

By the Blakley-Roy inequality the total number of (ordered) 3-walks in ∂H is at least $n(d^{\partial H})^3$. We claim that at most $3n(d^{\partial H}_{max})^2$ of these 3-walks are not 3-paths. Indeed, suppose $v_0v_1v_2v_3$ is a 3-walk that is not a 3-path. Then then there exists a repeated vertex v in the walk such that either $v_0 = v_2 = v$ or $v_1 = v_3 = v$ or $v_0 = v_3 = v$. Since v can be chosen in n ways and the other two vertices of the walk are adjacent to v, we can choose them in at most $(d^{\partial H}_{max})^2$ different ways. Therefore, the number of (ordered) 3-paths in ∂H is at least $n(d^{\partial H})^3 - 3n(d^{\partial H}_{max})^2 \ge n(d^{\partial H})^3 - 3n(r^{3/2}\sqrt{n})^2 = n(d^{\partial H})^3 - 3r^3n^2$.

We will show that most of these 3-paths are good by bounding the number of bad 3-paths. Suppose $v_0v_1v_2v_3$ is a bad 3-path. Then either $\{v_0, v_1, v_2\}$ or $\{v_1, v_2, v_3\}$ is contained in some hyperedge $h \in E(H)$. In the first case, the number of choices for $v_0v_1v_2$ is $|E(H)| \binom{r}{3} 3!$ as there are $\binom{r}{3}$ ways to choose the vertices v_0, v_1, v_2 from a hyperedge of H and then 3! ways to order them. And there are at most $d^{\partial H}_{max}$ choices for v_3. The second case is similar. Therefore, in total, the number of bad 3-paths in ∂H is at most $2\,|E(H)| \binom{r}{3} 3!\, d^{\partial H}_{max} < 2\frac{nd}{r} r^3 d^{\partial H}_{max} \leq 2nr^2 d_{max} d^{\partial H}_{max} \leq 2\frac{r^5}{r-1} n^2$. So the number of (ordered) good 3-paths in ∂H is at least

$$n(d^{\partial H})^3 - 3r^3n^2 - 2\frac{r^5}{r-1}n^2 = n(d^{\partial H})^3 - c_r n^2 = (r-1)^3 d^3 n - c_r n^2, \tag{4.9}$$

where $c_r = 3r^3 + \frac{2r^5}{r-1}$.

The following claim is useful for upper bounding the number of (ordered) good 3-paths in ∂H.

Claim 4.8. *If C is a cycle of length at most five in ∂H, then its vertex set is contained in some hyperedge of H.*

Proof. Let $v_1, v_2, \ldots, v_k, v_1$ be a cycle of length k in ∂H (for some $k \leq 5$). For each i, let h_i be the hyperedge of H containing v_i, v_{i+1} (addition in the subscripts is taken modulo k). If these k hyperedges are not all the same, there exists j, j' such that $h_j, h_{j+1}, \ldots, h_{j'}$ are all distinct but $h_{j'+1} = h_j$. So these hyperedges form a cycle in H of length at most $k \leq 5$, a contradiction. Therefore, $h_1 = h_2 = \ldots = h_k = h$; then $v_1, v_2, \ldots, v_k \in h$, as desired. □

In order to upper bound the number of (ordered) good 3-paths in ∂H, let us first fix a hyperedge h of H, and bound the number of good 3-paths $v_0v_1v_2v_3$ such that $v_0, v_1 \in h$.

Claim 4.9. *For any vertex $v \notin h$, there are at most $(r-1)$ good 3-paths $v_0v_1v_2v$ such that $v_0, v_1 \in h$.*

Proof. Suppose $v_0v_1v_2v$ and $v'_0v'_1v'_2v$ are good 3-paths with $v_0, v_1, v'_0, v'_1 \in h$. Then $v_2, v'_2 \notin h$ because it would contradict the definition of a good 3-path. We will prove that $v_1 = v'_1$ and $v_2 = v'_2$.

Suppose $v_1 \neq v'_1$. Then depending on whether $v_2 = v'_2$ or not, either $v_1v'_1v'_2vv_2$ forms a five-cycle or $v_1v'_1v'_2$ forms a triangle in ∂H. Then by Claim 4.8, $v_1, v'_1, v'_2 \in h'$ for some hyperedge $h' \in E(H)$. (Note that $h' \neq h$, since $v'_2 \notin h$.) But then h and h' are two different hyperedges of H that share at least two vertices, namely v_1, v'_1, contradicting the fact that H is linear. Thus $v_1 = v'_1$.

Now if $v_2 \neq v'_2$, then $vv_2v_1v'_2$ is a four-cycle in ∂H, so it must be contained in a hyperedge of H, but this means the 3-path $v_0v_1v_2v$ is bad, a contradiction. Thus $v_2 = v'_2$.

In summary, any two good 3-paths $v_0v_1v_2v$ and $v'_0v'_1v'_2v$ with $v_0, v_1, v'_0, v'_1 \in h$ can only differ in their first vertex, of which there are at most $r-1$ choices, proving the claim. □

Claim 4.9 implies that for any fixed hyperedge $h \in E(H)$, there are at most $(r-1)n$ good 3-paths $v_0v_1v_2v_3$ with $v_0, v_1 \in h$. Therefore, the total number of good 3-paths in H is at most $|E(H)|\,(r-1)n = \frac{(r-1)dn^2}{r}$.

Combining this with (4.9), we obtain $(r-1)^3d^3n - c_rn^2 \leq \frac{(r-1)dn^2}{r}$. Dividing through by d and using that $d = \Omega(\sqrt{n})$, we get $(r-1)^3d^2n \leq (1+o(1))\frac{(r-1)n^2}{r}$ and upon simplification and rearranging, we get

$$d \leq (1+o(1))\frac{\sqrt{n}}{\sqrt{r}(r-1)},$$

so using $|E(H)| = nd/r$, completes the proof.

4.3.2 Further improving the estimate on $ex(n, K_3, C_5)$

Here we slightly improve Theorem 4.2, by establishing a connection to girth 6 hypergraphs and using Theorem 4.3.

Recall that in the proof of Theorem 4.2, G denotes a C_5-free graph, and $(1-\alpha)e(G)$ edges of G belong to the K_4's in the edge-decomposition $\mathcal{D}$ of G. Let us note that the vertex sets of two different K_4's of G do not share more than one vertex, since G is C_5-free. Consider the 4-uniform hypergraph H formed by taking the vertex sets of all the K_4's of G. Then notice that H is linear and if H contains a (Berge) cycle of length at most 5, then G contains a C_5. Therefore, H is of girth 6. Therefore, by Theorem 4.3, H contains at most $n^{3/2}/24$ hyperedges. Thus at most $n^{3/2}/24 \times \binom{4}{2} = n^{3/2}/4$ edges of G belong to the K_4's in the edge-decomposition $\mathcal{D}$. Therefore, $(1-\alpha)e(G) \leq \frac{n^{3/2}}{4}$, which implies $d \leq \frac{\sqrt{n}}{2(1-\alpha)}$. Combining this with (4.7), we get

$$d \leq (1+o(1)) \min\left\{\frac{1}{2(1-\alpha)}, \left(\frac{1+\alpha}{4}\right)^{1/4}\right\}\sqrt{n},$$

so using (4.8), we obtain that the number of triangles in G is at most

$$(1+o(1))\frac{(4-\alpha)}{12}\min\left\{\frac{1}{2(1-\alpha)}, \left(\frac{1+\alpha}{4}\right)^{1/4}\right\}n^{3/2}.$$

The above function is maximized at $\alpha = 0.343171$, proving that $ex(n, K_3, C_5) \leq 0.231975n^{3/2}$.

4.4 C_5-free and induced-C_4-free graphs: Proof of Theorem 4.4

Let G be a C_5-free graph on n vertices having no induced copies of C_4. Let G_Δ be the subgraph of G consisting of the edges that are contained in triangles of G, and let G_S be the subgraph of G consisting of the remaining edges of G. Since G_Δ is C_5-free and every edge of it is contained in a triangle, by the same argument of the proof of Theorem 4.2, the triangles of G_Δ can be partitioned into crown-blocks and K_4-blocks. So there is a decomposition $\mathcal{D}$ of the edges of G_Δ into paths of length 2, triangles and K_4's. First let us note that Claim 4.3 in the proof of Theorem 4.2 still holds for G (not just for G_Δ), as shown below.

Claim 4.10. *Let u, v be two non-adjacent vertices of G. Then the number of paths of length 2 between u and v is at most two. Moreover, if uxv and uyv are the paths of length 2 between u and v, then x and y are adjacent.*

Proof. The second part of the claim is trivial since G does not contain an induced copy of C_4. To see the first part of the claim, suppose uxv, uyv, uzv are three distinct paths of length 2 in G. Then x and y are adjacent, so $uxyvz$ is a C_5 in G, a contradiction. □

Our goal is to bound the average degree d of G. If a vertex has degree less than $d/2$, then it may be deleted without decreasing the average degree of G, so we may assume that G has minimum degree at least $d/2$. Now using this fact and Claim 4.10, one can show that the maximum degree of G is at most $10\sqrt{n}$ by repeating the same argument as in the proof of Claim 4.4.

We say a 5-path $v_0v_1v_2v_3v_4v_5$ is *bad* if there exists an i such that $v_iv_{i+1}v_{i+2}$ is a triangle of G; otherwise it called *good*. Similarly, a 2-path abc is *good* if a and c are not adjacent. By the same argument as in the proof of Theorem 4.2, the number of (unordered) good 5-paths in G is at least

$$\frac{nd^5}{2} - Cn^3 \tag{4.10}$$

for some constant $C > 0$. Now we bound the number of good 5-paths in G from above. Let $|E(G_\Delta)| = \alpha\,|E(G)|$ for some $\alpha \geq 0$, so $|E(G_S)| = (1-\alpha)\,|E(G)|$.

Claim 4.11. *The number of good 5-paths in G whose middle edge is contained in G_S is at most $|E(G_S)|\,n^2$.*

Proof. The proof is very similar to that of the proof of Claim 4.5. A good 5-path $xyabzw$ whose middle edge ab is in G_S contains good 2-paths, xya, bzw as subpaths.

Let n_a be the number of good 2-paths in G of the form axy where $x, y \neq b$, and let n_b be the number of good 2-paths in G of the form bxy where $x, y \neq a$. Then the number of good 5-paths whose middle edge is ab is at most $n_an_b \leq (n_a+n_b)^2/4$. By the same argument as in the proof of Claim 4.5, it is easy to see that $n_a + n_b \leq 2n$, so the number of good 5-paths whose middle edge is $ab \in E(G_S)$ at most n^2. Adding these estimates for all the edges $ab \in E(G_S)$ finishes the proof of the claim. □

Let us further assume that the number of edges of G_Δ that belong to paths of length 2, triangles and K_4's in its edge-decomposition $\mathcal{D}$ be $\alpha_1\,|E(G)|\,, \alpha_2\,|E(G)|\,, \alpha_3\,|E(G)|$, respectively. (Of course, $\alpha_1 + \alpha_2 + \alpha_3 = \alpha$.) Since Claim 4.10 holds, one can easily check that the proofs of Claim 4.5, Claim 4.6 and Claim 4.7 are still valid, so these claims hold in the current setting too. These claims, together with Claim 4.11, imply that the number of good 5-paths in G is at most

$$\frac{\alpha_1\,|E(G)|}{2}n^2 + \frac{\alpha_2\,|E(G)|}{3}\frac{4n^2}{3} + \frac{\alpha_3\,|E(G)|}{6}\frac{3n^2}{2} + |E(G_S)|\,n^2 \leq \frac{\alpha\,|E(G)|}{2}n^2 + (1-\alpha)\,|E(G)|\,n^2.$$

We will now bound the right-hand-side of the above inequality by carefully selecting a C_5-free, and C_4-free subgraph G' of G, as follows: We select all the edges of G_S and the following edges from G_Δ: From each crown-block $\{abc_1, abc_2, \dots, abc_k\}$ of G_Δ, we select the edges $ac_1, ac_2, \dots, ac_k$ to be in G'. From each K_4-block $abcd$ we select the edges ab, bc, ac, ad to be in G'.

By Claim 4.2, the edge set of every C_4 is completely contained in some block of G_Δ, and it is easy to check that the selected edges in each block of G_Δ form a C_4-free graph. Therefore, G' is C_4-free. Since it is a subgraph of G, it is also C_5-free. Therefore, by a theorem of Erdős and Simonovits [19], $|E(G')| \leq \frac{1}{2\sqrt{2}} n^{3/2}$. On the other hand, since all the edges of G_S and at least half the edges of G_Δ are selected, we have $|E(G')| \geq |E(G_S)| + \frac{|E(G_\Delta)|}{2} = (1-\alpha)\,|E(G)| + \frac{\alpha |E(G)|}{2}$. Therefore,

$$\frac{\alpha\,|E(G)|}{2} + (1-\alpha)\,|E(G)| \leq \frac{1}{2\sqrt{2}} n^{3/2}.$$

Therefore, by the discussion above, the number of good 5-paths in G is at most $\frac{1}{2\sqrt{2}} n^{3/2} \times n^2 = \frac{1}{2\sqrt{2}} n^{7/2}$. Combining this with (4.10), we get

$$\frac{nd^5}{2} - Cn^3 \leq \frac{1}{2\sqrt{2}} n^{7/2},$$

so $\frac{nd^5}{2} \leq (1+o(1)) \frac{1}{2\sqrt{2}} n^{7/2}$, implying that $d \leq \frac{\sqrt{n}}{\sqrt[10]{2}}$, finishing the proof.

Chapter 5

On a hypergraph bipartite Turán problem

5.1 Introduction

An *r-graph* is an r-uniform hypergraph. Determining the asymptotic order of $\mathrm{ex}(n,\mathcal{F})$ is generally very difficult. For an excellent survey on the study of hypergraph Turán numbers, see [54]. In this chapter, we study a hypergraph Turán problem that is motivated by the study of Turán numbers of complete bipartite graphs as well as by a question of Erdős.

Definition 5.1. *Let $r \geq 3$ be an integer. Let G be a bipartite graph with an ordered bipartition (X,Y). Suppose that $Y=\{y_1,\dots,y_m\}$. Let $Y_1,\dots,Y_m$ be disjoint sets of size $r-2$ that are disjoint from $X\cup Y$. Let $G^{(r)}_{X,Y}$ denote the r-graph with vertex set $(X\cup Y)\cup(\bigcup_{i=1}^m Y_i)$ and edge set $\bigcup_{i=1}^m\{e\cup Y_i : e\in E(G), y_i\in e\}$.*

Let $s,t\geq 2$ be positive integers. If G is the complete bipartite graph with an ordered bipartition (X,Y) where $|X|=s, |Y|=t$, then let $G^{(r)}_{X,Y}$ be denoted by $K^{(r)}_{s,t}$.

As mentioned in Chapter 1, Mubayi and Verstraëte [64] initiated the study of the general problem of determining $\mathrm{ex}(n,K^{(r)}_{2,t})$ for any $t\geq 2$. They showed that for any $t\geq 2$ and $n\geq 2t$

$$\mathrm{ex}(n,K^{(3)}_{2,t}) < t^4\binom{n}{2},$$

and that for infinitely many n, $\mathrm{ex}(n,K^{(3)}_{2,t}) \geq \frac{2t-1}{3}\binom{n}{2}$, where the lower bound is obtained by replacing each hyperedge in $S(n,2t+1,2)$ with all its 3-element subsets, where $S(n,2t+1,2)$ is an (n,r,t)-Steiner system.

Mubayi and Verstraëte noted that $g(t) := \lim_{n\to\infty}\mathrm{ex}(n,K^{(3)}_{2,t})/\binom{n}{2}$ exists and raised the question of determining the growth rate of $g(t)$. It follows from their results that

$$\frac{2t-1}{3} \leq g(t) \leq t^4. \tag{5.1}$$

In this chapter, we prove that as $t \to \infty$,

$$g(t) = \Theta(t^{1+o(1)}), \tag{5.2}$$

showing that their lower bound is close to the truth. More precisely, we prove the following.

Theorem 5.1 (E., Jiang, Methuku [27]). *For any $t \geq 2$, we have*

$$\mathrm{ex}(n, K_{2,t}^{(3)}) \leq (15t \log t + 40t)\, n^2.$$

Notation. Given a hypergraph (or a graph) H, throughout the Cha,tpre we also denote the set of its edges by H. For example $|H|$ denotes the number of edges of H. Given two vertices x, y in a graph H, let $N_H(x, y)$ denote the common neighborhood of x and y in H. We drop the subscript H when the context is clear.

5.2 Proof of Theorem 5.1: $K_{2,t}^{(3)}$-free hypergraphs

We will use the a special case of a well-known result of Erdős and Kleitman [25].

Lemma 5.2. *Let H be a 3-graph on $3n$ vertices. Then H contains a 3-partite 3-graph, with all parts of size n, and with at least $\frac{2}{9}|H|$ hyperedges.*

Let us define the sets $A = \{a_1, a_2, \ldots, a_n\}$, $B = \{b_1, b_2, \ldots, b_n\}$ and $C = \{c_1, c_2, \ldots, c_n\}$. Throughout the proof we define various 3-partite 3-graphs whose parts are A, B and C.

Suppose H is a $K_{2,t}^{(3)}$-free 3-partite 3-graph on $3n$ vertices with parts A, B and C. First let us show that it suffices to prove the following inequality.

$$|H| \leq (30t \log t + 80t)n^2. \tag{5.3}$$

It is easy to see that inequuity (5.3) and Lemma 5.2 together imply that any $K_{2,t}^{(3)}$-free 3-graph on $3n$ vertices contains at most $\frac{9}{2}(30t \log t + 80t)n^2$ hyperedges, from which Theorem 5.1 would follow after replacing $3n$ by n.

In the remainder of the section, we will prove (5.3). Let us introduce the following notion of sparsity.

Definition 5.3 (q-sparse and q-dense pairs). *Let q be a positive integer. Let G be a bipartite graph with parts X, Y. Let x, y be two different vertices such that $x, y \in X$ or $x, y \in Y$. Then we call $\{x, y\}$ a q-*dense *pair of G if $|N(x, y)| \geq q$. We call $\{x, y\}$ a q-*sparse *pair of G if $|N(x, y)| < q$ but x, y are still contained in a copy of $K_{2,q}$ in G. Note that it is possible that $\{x, y\}$ is neither q-sparse nor q-dense.*

The following Procedure $\mathcal{P}(q)$ about making a bipartite graph $K_{2,q}$-free lies at the heart of the proof. (We think of q as the parameter of the Procedure $\mathcal{P}(q)$, that is changed throughout the proof.)

Procedure $\mathcal{P}(q)$: Making a graph $K_{2,q}$-free

Input: A bipartite graph G with parts A and B.

$\mathcal{G} \leftarrow G$, $\psi \leftarrow 1$.

$F(x,y) \leftarrow \emptyset$, $D(x,y) \leftarrow \emptyset$ and $S(x,y) \leftarrow \emptyset$ for every $x, y \in A$ and $x, y \in B$.

while $\psi = 1$ **do**

$\psi \leftarrow 0$.

Step 1:

For each q-sparse pair $\{x, y\}$ of $\mathcal{G}$ such that $F(x,y) = \emptyset$, let $S(x,y)$ be the set of vertices spanned by the q-dense pairs of $\mathcal{G}$ that are contained in $N_{\mathcal{G}}(x,y)$.
Let $F(x,y) \leftarrow \{ab \in \mathcal{G} \mid a \in \{x,y\} \text{ and } b \in S(x,y)\}$, and let $D(x,y)$ be a spanning forest of the graph formed by the dense pairs of $\mathcal{G}$ that are contained in $S(x,y)$.

If there exists an edge $ab \in \mathcal{G}$ such that ab is contained in $F(x,y)$ for at least $q/2$ different pairs $\{x,y\}$, where $x, y \in A$ or $x, y \in B$,

then $\mathcal{G} \leftarrow \mathcal{G} \setminus \{ab\}$ and $\psi \leftarrow 1$.

Step 2:

If there exists a set M of edges in $\mathcal{G}$ such that removing all of the edges of M from $\mathcal{G}$ decreases the number of q-dense pairs by at least $|M|/2$,

then $\mathcal{G} \leftarrow \mathcal{G} \setminus M$ and $\psi \leftarrow 1$.

end while

$G' \leftarrow \mathcal{G}$

$F'(x,y) \leftarrow F(x,y)$ for every $x, y \in A$ and $x, y \in B$.

$D'(x,y) \leftarrow D(x,y)$ for every $x, y \in A$ and $x, y \in B$.

$S'(x,y) \leftarrow S(x,y)$ for every $x, y \in A$ and $x, y \in B$.

Output: The graph G' and the sets $F'(x,y), D'(x,y), S'(x,y)$ for all $x, y \in A$ and $x, y \in B$.

In the procedure $\mathcal{P}(q)$, initially for all the pairs $\{x, y\}$ (with $x, y \in A$ and $x, y \in B$) the sets $F(x,y), D(x,y), S(x,y)$ are set to be empty. Then as the edges are being deleted during the procedure, possibly, new q-sparse pairs $\{x,y\}$ are being created. When this happens, Step 1 redefines the sets $S(x,y), F(x,y), D(x,y)$ and gives them some non-empty values. (They get non-empty values due to the fact that $\{x,y\}$ is q-sparse, which implies that $\{x,y\}$ is contained in a copy of $K_{2,q}$, so there is at least one q-dense pair in the common neighborhood of x, y.) Therefore, these values stay unchanged throughout the rest of the procedure.

Notice that at the point $S(x,y)$ was redefined, the pair $\{x,y\}$ was q-sparse, so number of common neighbors is less than q. Therefore, as $S(x,y)$ is a subset of the common neighborhood of x and y, we also have $|S(x,y)| < q$. Moreover, since $D(x,y)$ is defined as a spanning forest with the vertex set $S(x,y)$, we have $|D(x,y)| \leq |S(x,y)|$. Also, it easily follows from the definition of $F(x,y)$ that $|F(x,y)| = 2|S(x,y)|$. Finally, notice that $D(x,y)$ does not contain any isolated vertices, because its vertex set $S(x,y)$ spans all of its edges, by definition. Therefore, $|D(x,y)| \geq |S(x,y)|/2$. At the end of the procedure, the sets $F(x,y), D(x,y), S(x,y)$ are renamed as $F'(x,y), D'(x,y), S'(x,y)$. Note also that if a pair $\{x,y\}$ never becomes q-sparse in the process then $S'(x,y) = D'(x,y) = F'(x,y) = \emptyset$.

Observation 5.4. For every $x, y \in A$ and $x, y \in B$, we have

(1) $|S'(x,y)| < q$.

(2) $|D'(x,y)| \leq |S'(x,y)|$.

(3) $|F'(x,y)| = 2|S'(x,y)|$.

(4) $|D'(x,y)| \geq |S'(x,y)|/2$.

For convenience, throughout the chapter we (informally) say that the sets $F'(x,y)$, $D'(x,y)$, $S'(x,y)$ are defined by applying Procedure $\mathcal{P}(q)$ to a graph G to obtain the graph G', instead of saying that the input to Procedure $\mathcal{P}(q)$ is G and the output is the graph G' and the sets $F'(x,y)$, $D'(x,y)$, $S'(x,y)$.

Claim 5.5. *Let the sets $F'(x,y), D'(x,y), S'(x,y)$ (for $x, y \in A$ and $x, y \in B$) be defined by applying Procedure $\mathcal{P}(q)$ to a bipartite graph G to obtain G'. Let $N(x,y)$ denote the number of common neighbors of vertices x, y in the graph G. Then*

$$\frac{|F'(x,y)|}{4} \leq |D'(x,y)| < q.$$

Moreover $|F'(x,y)| \leq 2|N(x,y)|$.

Proof. Combining the parts (3) and (4) of Observation 5.4, we have $|F'(x,y)|/4 \leq |D'(x,y)|$. Combining the parts (1) and (2) of Observation 5.4, we obtain $|D'(x,y)| < q$, proving the first part of the claim.

To prove the second part, notice that $S'(x,y)$ is a common neighborhood of x, y in some subgraph $\mathcal{G}$ of G, we have $|S'(x,y)| \leq |N(x,y)|$. Combining this with part (3) of Observation 5.4, we obtain $|F'(x,y)| \leq 2|N(x,y)|$, as required. □

Finally, let us note the following properties of the graph obtained after applying the procedure.

Observation 5.6. Let the sets $F'(x,y), D'(x,y), S'(x,y)$ (for $x, y \in A$ and $x, y \in B$) be defined by applying Procedure $\mathcal{P}(q)$ to a bipartite graph G to obtain G'. Then

1. Every edge ab in G' is contained in at most $q/2$ members of $\{F'(x,y) : x, y \in A\}$ and in at most $q/2$ members of $\{F'(x,y) : x, y \in B\}$.

2. For any set M of edges in G', removing the edges of M from G' decreases the number of q-dense pairs by less than $|M|/2$.

Definition 5.7. *Let H be a 3-partite 3-graph with parts A, B and C.*

For each $1 \leq i \leq n$, let $G_i[H](A, B)$ be the bipartite graph with parts A and B, whose edge set is $\{ab \mid a \in A, b \in B, abc_i \in E(H)\}$. The graphs $G_i[H](B, C)$ and $G_i[H](A, C)$ are defined similarly.

Definition 5.8 (Applying Procedure $\mathcal{P}(q)$ to a hypergraph)**.** *Let H be a 3-partite 3-graph with parts A, B and C. We define the hypergraph H' as follows:*

For each $1 \leq i \leq n$, let $G'_i[H](A, B)$, $G'_i[H](B, C)$, $G'_i[H](A, C)$ be the graphs obtained by applying the procedure $\mathcal{P}(q)$ to the graphs $G_i[H](A, B)$, $G_i[H](B, C)$, $G_i[H](A, C)$ respectively.

For each edge ab which was removed from $G_i[H](A, B)$ by the procedure $\mathcal{P}(q)$ (i.e. $ab \in G_i[H](A, B) \setminus G'_i[H](A, B)$) we remove the hyperedge abc_i from $\mathcal{H}$ (it may have been removed already). Similarly for each edge bc (resp. ac) which was removed from $G_i[H](B, C)$ (resp. $G_i[H](A, C)$) by the procedure $\mathcal{P}(q)$ we remove the hyperedge a_ibc (resp. ab_ic) from $\mathcal{H}$. Let the resulting hypergraph be H'. More precisely,

$$H' = \{a_ib_jc_k \in H \mid a_ib_j \in G'_k[H](A, B),\ b_jc_k \in G'_i[H](B, C),\ a_ic_k \in G'_j[H](A, C)\}.$$

We say H' is obtained from H by applying the Procedure $\mathcal{P}(q)$.

Remark 5.9. Let H' be obtained by applying the Procedure $\mathcal{P}(q)$ to the hypergraph H. Then,

$$|H| - |H'| \leq \sum_{1 \leq i \leq n} \left(|G_i[H](A, B)| - |G'_i[H](A, B)|\right) + \sum_{1 \leq i \leq n} \left(|G_i[H](B, C)| - |G'_i[H](B, C)|\right) + \sum_{1 \leq i \leq n} \left(|G_i[H](A, C)| - |G'_i[H](A, C)|\right).$$

Indeed, if $a_ib_jc_k \in H \setminus H'$ then it is easy to see that $a_ib_j \in G_k[H](A, B) \setminus G'_k[H](A, B)$ or $b_jc_k \in G_i[H](B, C) \setminus G'_i[H](B, C)$ or $a_ic_k \in G_j[H](A, C) \setminus G'_j[H](A, C)$.

Lemma 5.10. *Let $q \geq 2$ be an even integer and G be a bipartite graph with parts A and B. Suppose G' is the graph obtained by applying Procedure $\mathcal{P}(q)$ to G. Then G' is $K_{2,q}$-free.*

Proof. Let us define a *q-broom* of size k to be a set of q-sparse pairs $\{x_0, x_j\}$ (with $1 \leq j \leq k$), and a q-dense pair $\{y, z\}$ such that $\{y, z\}$ is contained in the common neighborhood of x_0, x_j for every $1 \leq j \leq k$. Note that either $\{x_0, x_1, \ldots, x_k\} \subseteq A$ and $\{y, z\} \subseteq B$ or $\{x_0, x_1, \ldots, x_k\} \subseteq B$ and $\{y, z\} \subseteq A$.

Claim 5.11. *There is no q-broom of size $q/2$ in G'.*

Proof. Suppose by contradiction that there is a set of q-sparse pairs $\{x_0, x_j\}$ (with $1 \leq j \leq q/2$), and a q-dense pair $\{y, z\}$ such that $\{y, z\}$ is contained in the common neighborhood of x_0 and x_j for every $1 \leq j \leq q/2$. Then the edge x_0y is contained in the sets $F'(x_0, x_j)$ for every $1 \leq j \leq q/2$, which contradicts Observation 5.6. □

Let us suppose for a contradiction (to Lemma 5.10) that G' contains a copy of $K_{2,q}$. Then G' contains at least one q-dense pair. Without loss of generality we may assume there is a q-dense pair $\{a, a_1\}$ in A. Suppose $\{a, a_j\}$ (for $1 \leq j \leq p$) are all the q-dense pairs of G' containing the vertex a. For each $1 \leq j \leq p$, let $B_j \subseteq B$ be the common neighborhood of a and a_j in G'. By definition, $|B_j| \geq q$ for $1 \leq j \leq p$.

Claim 5.12. *For any $J \subseteq \{1, 2, \ldots, p\}$, we have $|\bigcup_{j \in J} B_j| > 2\,|J|$.*

Proof. Let us assume for contradiction that there exists a $J \subseteq \{1, 2, \ldots, p\}$ such that $|\bigcup_{j \in J} B_j| \leq 2\,|J|$. Let G^* be obtained from G' by deleting all the edges from a to $\bigcup_{j \in J} B_j$. For each $j \in J$, the pair $\{a, a_j\}$ has no common neighbor in G^* since we have removed all the edges from a to B_j. Thus the pair $\{a, a_j\}$ is not q-dense in G^*. So in forming G^* from G' the number of q-dense pairs decreases by at least $|J|$, while the number of edges decreases by $|\bigcup_{j \in J} B_j| \leq 2|J|$ edges, contradicting Observation 5.6. □

Let $B' = \bigcup_{1 \leq j \leq p} B_j$. For each vertex $v \in B'$ and let

$$J(v) := \{j \mid v \in B_j\},$$

$$D(v) := \{\{v, u\} \mid \{v, u\} \text{ is } q\text{-dense in } G' \text{ and } \{v, u\} \subseteq B_j \text{ for some } j \in J(v)\}.$$

In the next two claims, we will prove two useful inequalities concerning $|J(v)|$ and $|D(v)|$.

Claim 5.13. *For each $v \in B'$, $|J(v)| > 2\,|D(v)|$.*

Proof. Suppose for contradiction that there is a vertex $v \in B'$ such that $|J(v)| \leq 2\,|D(v)|$. Let us delete all the edges of the form va_j, $j \in J(v)$, from G' and let the resulting graph be G^*. Since we deleted $|J(v)|$ edges, by Observation 5.6, the number of q-dense pairs decreases by less than $|J(v)|/2 \leq |D(v)|$. So there exists $\{v, u\} \in D(v)$ such that $\{v, u\}$ is (still) q-dense in G^*. That is, $|N^*(v, u)| \geq q$, where $N^*(v, u)$ denotes the common neighborhood of v and u in G^*. Clearly each pair of vertices in $N^*(v, u)$ is contained in a copy of $K_{2,q}$ in G^* (and hence in G').

For each pair of vertices in $N^*(v, u)$, since it is contained in a copy of $K_{2,q}$ in G', it is either q-sparse or q-dense in G'. Note that $a \in N^*(v, u)$. If all the pairs $\{a, x\}$ with $x \in N^*(v, u) \setminus \{a\}$ are q-sparse in G' then the set of these pairs together with $\{v, u\}$ is a q-broom of size at least $q - 1 \geq q/2$ in G', which contradicts Claim 5.11. So there exists a vertex $x \in N^*(v, u) \setminus \{a\}$ such that $\{a, x\}$ is q-dense in G'. Since v is adjacent to both a and x, by the definition of $J(v)$, $x = a_j$ for some $j \in J(v)$. However, by definition, in forming G^* we have removed vx from G'. This contradicts $x \in N^*(v, u)$ and completes the proof. □

Claim 5.14.

$$\sum_{v \in B'} |D(v)| \geq \frac{1}{2} \sum_{1 \leq j \leq p} |B_j|.$$

Proof. Fix any j with $1 \leq j \leq p$. Since $\{a, a_j\}$ is q-dense in G', every pair $\{x, y\} \subseteq B_j$ is contained in some copy of $K_{2,q}$ and hence is either q-dense or q-sparse in G'. Let v be any vertex in B_j and let $S(v) = \{y \in B_j \mid \{v, y\}$ is q-sparse in $G'\}$. By definition, the set $\{\{v, y\} \mid y \in S(v)\}$ together with $\{a, a_j\}$ is a q-broom of size $|S(v)|$. By Claim 5.11, $|S(v)| \leq q/2 - 1 \leq |B_j|/2 - 1$. Since $|D(v)| + |S(v)| \geq |B_j| - 1$, we have

$$|D(v)| \geq \frac{1}{2}|B_j| \tag{5.4}$$

Note that (5.4) holds for every $j = 1, \ldots, p$ and every $v \in B_j$.

Let us define an auxiliary bipartite graph G_{aux} with a bipartition $(\{1, 2, \ldots p\}, B')$ in which a vertex $j \in \{1, \ldots, p\}$ is joined to a vertex $y \in B'$ if and only if $y \in B_j$. Let J be an arbitrary subset of $\{1, 2, \ldots, p\}$. The neighborhood of J in G_{aux} is precisely $\bigcup_{j \in J} B_j$. By Claim 5.12, $|\bigcup_{j \in J} B_j| > 2|J| \geq |J|$. Since this holds for every $J \subseteq \{1, \ldots, p\}$, by Hall's theorem [50] there exist distinct vertices $w_j \in B_j$, for $j = 1, \ldots, p$. By (5.4), for every $j \in \{1, \ldots, p\}$, $|D(w_j)| \geq \frac{1}{2}|B_j|$. Hence

$$\sum_{v \in B'} |D(v)| \geq \sum_{1 \leq j \leq p} |D(w_j)| \geq \frac{1}{2} \sum_{1 \leq j \leq p} |B_j|.$$

□

If we view $\{B_1, \ldots, B_p\}$ as a hypergraph on the vertex set B', then the degree of a vertex $v \in B'$ in it is precisely $|J(v)|$ and the degree sum formula yields

$$\sum_{v \in B'} |J(v)| = \sum_{1 \leq j \leq p} |B_j|. \tag{5.5}$$

Using Claim 5.13 and Claim 5.14 we have

$$\sum_{v \in B'} |J(v)| > \sum_{v \in B'} 2|D(v)| \geq 2 \sum_{1 \leq j \leq p} \frac{1}{2}|B_j| = \sum_{1 \leq j \leq p} |B_j|,$$

which contradicts (5.5). This completes proof of Lemma 5.10. □

In the next subsection we will prove a general lemma about making an arbitrary hypergraph $K_{1,2,q}$-free (for any given value of q). This lemma is used several times in the following subsections.

5.2.1 Applying Procedure $\mathcal{P}(q)$ to an arbitrary hypergraph H

Let q be an even integer and let $q \geq t$. Let H be an arbitrary $K_{2,q}^{(3)}$-free 3-partite 3-graph with parts A, B and C. In this subsection we will prove the following lemma that estimates the number of edges removed from the graphs $G_i = G_i[H](A, B)$ for $1 \leq i \leq n$, when the Procedure $\mathcal{P}(q)$ is applied to them. This lemma together with Remark 5.9 will allow us to estimate the number of edges removed from H when the Procedure $\mathcal{P}(q)$ is applied to it.

Throughout this subsection, $N_i(x, y)$ denotes the set of common neighbors of the vertices x, y in the graph G_i.

Lemma 5.15. *Let $q \geq t$ be an even integer. Let H be an arbitrary $K_{2,q}^{(3)}$-free 3-partite 3-graph with parts A, B and C. Let $G_i = G_i[H](A,B)$ for $1 \leq i \leq n$. For each $1 \leq i \leq n$ and any $x, y \in A$ or $x, y \in B$, let $F_i'(x,y)$ be defined by applying the procedure $\mathcal{P}(q)$ to G_i and let the resulting graph be G_i'. Then,*

$$\sum_{1\leq i\leq n} |G_i \setminus G_i'| < \frac{2}{q}\left(\sum_{u,v\in A}\sum_{1\leq i\leq n}|F_i'(u,v)| + \sum_{u,v\in B}\sum_{1\leq i\leq n}|F_i'(u,v)|\right) + 2tn^2.$$

Proof of Lemma 5.15. First let us prove the following claim.

Claim 5.16. *Let $u, v \in A$ or $u, v \in B$. Then $\{u, v\}$ is q-dense in less than t of the graphs G_i, $1 \leq i \leq n$.*

Proof. Without loss of generality, suppose that $u, v \in A$. Suppose for contradiction that $\{u, v\}$ is q-dense in t of the graphs G_i, $1 \leq i \leq n$. Without loss of generality suppose $\{u, v\}$ is q-dense in $G_1, \ldots, G_t$. Then $|N_i(u,v)| \geq q \geq t$ for $i = 1, \ldots, t$. Therefore, we can greedily choose t distinct vertices $y_1, \ldots, y_t$ such that for each $i \in [t], y_i \in N_i(u,v)$. For each $i \in [t]$, since $y_i \in N_i(u,v)$ we have $uy_ic_i, vy_ic_i \in E(H)$. However, the set of hyperedges $\{uy_ic_i, vy_ic_i \in E(H) \mid 1 \leq i \leq t\}$ forms a copy of $K_{2,t}^{(3)}$ in H, a contradiction. □

Note that when procedure $\mathcal{P}(q)$ is applied to G_i (to obtain G_i'), Step 1 and Step 2 may be applied several times (and each time one of these steps is applied it may delete an edge of G_i).

For each $i \in [n]$, let m_i denote the number of q-dense pairs of G_i. By Claim 5.16, we know that each pair $\{u, v\}$ with $u, v \in A$ or $u, v \in B$, is q-dense in less than t different graphs G_i (for $1 \leq i \leq n$). Therefore,

$$\sum_{1\leq i\leq n} m_i \leq \sum_{u,v\in A}(t-1) + \sum_{u,v\in B}(t-1) = 2\binom{n}{2}(t-1). \tag{5.6}$$

For each $i \in [n]$, let α_i denote the total number of edges that were removed by Step 1 when procedure $\mathcal{P}(q)$ is applied to G_i and β_i be the number of edges removed by Step 2 when procedure $\mathcal{P}(q)$ is applied to G_i. Then $\alpha_i + \beta_i = |G_i \setminus G_i'|$, so $\sum_{i=1}^n \alpha_i + \sum_{i=1}^n \beta_i = \sum_{i=1}^n |G_i \setminus G_i'|$.

First, we bound $\sum_{i=1}^n \beta_i$. Let $i \in [n]$. Observe that whenever a set M of edges were removed by Step 2 of Procedure $\mathcal{P}(q)$ applied to G_i, the number of q-dense pairs decreased by at least $|M|/2$. Hence $\beta_i \leq 2m_i$. So summing up over all $1 \leq i \leq n$, and using (5.6), we get

$$\sum_{1\leq i\leq n} \beta_i \leq 2\sum_{1\leq i\leq n} m_i \leq 2n(n-1)(t-1) < 2tn^2. \tag{5.7}$$

Next, we bound $\sum_{i=1}^n \alpha_i$. Let $i \in [n]$. If an edge xy was removed from G_i by Step 1 of the procedure $\mathcal{P}(q)$ then there are vertices $z_1, z_2, \ldots, z_{q/2}$ such that $xy \in F_i'(x, z_j)$ for every $j \in \{1, 2, \ldots, q/2\}$ or $xy \in F_i'(y, z_j)$ for every $j \in \{1, 2, \ldots, q/2\}$. So

$$\alpha_i \le \frac{1}{q/2}\left(\sum_{u,v\in A}|F_i'(u,v)| + \sum_{u,v\in B}|F_i'(u,v)|\right).$$

Therefore,

$$\sum_{1\le i\le n}\alpha_i \le \frac{2}{q}\left(\sum_{1\le i\le n}\sum_{u,v\in A}|F_i'(u,v)| + \sum_{1\le i\le n}\sum_{u,v\in B}|F_i'(u,v)|\right).$$

This is equivalent to the following.

$$\sum_{1\le i\le n}\alpha_i \le \frac{2}{q}\left(\sum_{u,v\in A}\sum_{1\le i\le n}|F_i'(u,v)| + \sum_{u,v\in B}\sum_{1\le i\le n}|F_i'(u,v)|\right). \tag{5.8}$$

Combining this inequality with (5.7) completes the proof of Lemma 5.15. □

5.2.2 The overall plan

Let us define the sequence $q_0, q_1, \dots, q_k$ as follows. Let $q_0 = 2^l$ where l is an integer such that $q_0 = 2^l \le t^2 < 2^{l+1} = 2q_0$. For each $1 \le j \le k$, let $q_j = \frac{q_{j-1}}{2}$ and $q_k \ge t > \frac{q_k}{2}$. Clearly $\frac{q_0}{q_k} = 2^k$, moreover

$$2^k = \frac{q_0}{q_k} \le \frac{t^2}{t} = t.$$

So we have

$$k \le \log t. \tag{5.9}$$

Now we apply the procedure $\mathcal{P}(q_0)$ to the hypergraph H (recall Definition 5.8) to obtain a $K_{1,2,q_0}$-free hypergraph H_0. For each $0 \le j < k$ we obtain $K_{1,2,q_{j+1}}$-free hypergraph H_{j+1} by applying the procedure $\mathcal{P}(q_{j+1})$ to the hypergraph H_j.

This way, in the end we will get a $K_{1,2,q_k}$-free hypergraph H_k. In the following section, we will upper bound $|H| - |H_0|$. Then in the next section, using the information that H_j is $K_{1,2,q_j}$-free, we will upper bound $|H_{j+1}| - |H_j|$ for each $0 \le j < k$. Then we sum up these bounds to upper bound the total number of deleted edges (i.e., $|H| - |H_k|$) from H to obtain H_k. Finally, we bound the size of H_k, which will provide us the desired bound on the size of H.

5.2.3 Making H $K_{1,2,q_0}$-free

First, we are going to prove an auxiliary lemma that is similar to Lemma A.4 of [64]. In an edge-colored multigraph G, an *s-frame* is a collection of s edges all of different colors such that it is possible to pick one endpoint from each edge with all the selected endpoints being distinct.

Lemma 5.17. *Let G be an edge-colored multigraph with e edges such that each edge has multiplicity at most p and each color class has size at most q. If G contains no t-frame then $|G| \le \binom{t-1}{2}p + tq$.*

Proof. Consider a maximum frame S, say with edges $e_1, \ldots, e_s$ such that for every $i \in \{1, 2, \ldots, s\}$, e_i has color i and that there exist $x_1 \in e_1, x_2 \in e_2, \ldots, x_s \in e_s$ with $x_1, \ldots, x_s$ being distinct. By our assumption, $s \leq t-1$. Let f be any edge with a color not in $[s]$. Then both vertices of f must be in $\{x_1, \ldots, x_s\}$, otherwise $e_1, \ldots, e_s, f$ give a larger frame, a contradiction. On the other hand, each edge with both of its vertices in $\{x_1, \ldots, x_s\}$ has multiplicity at most p. Hence there are at most $\binom{s}{2}p$ edges with colors not in $\{1, 2, \ldots, s\}$. The number of edges with color in $\{1, 2, \ldots, s\}$ is at most sq by our assumption. So $|G| \leq \binom{s}{2}p + sq \leq \binom{t-1}{2}p + tq$. □

Let us recall that H is 3 partite $K_{2,t}^{(3)}$-free hypergraph with A, B, C. For convenience we denote $G_i = G_i[H](A, B)$ where $1 \leq i \leq n$. For each $1 \leq i \leq n$ and any $x, y \in A$ or $x, y \in B$, let $F_i'(x, y)$, $D_i'(x, y)$ and $S_i'(x, y)$ be defined by applying the procedure $\mathcal{P}(q_0)$ on G_i and let the obtained graph be G_i'.

First, observe that $t^2/2 < q_0 \leq t^2$ according to our definition.

Claim 5.18. *Let $u, v \in A$ or $u, v \in B$. Then $\sum_{1 \leq i \leq n} |F_i'(u, v)| \leq 6t^3$.*

Proof. Let D^* be an edge-colored multigraph in which a pair of vertices e is an edge of color $i \in [n]$ whenever e is an edge of $D_i'(u, v)$. The number of edges of color i in D^* is $|D_i'(u, v)|$. By Claim 5.5 we have $|D_i'(u, v)| < q_0$. Hence the number of edges in each color class of D^* is less than q_0.

Let xy be an arbitrary edge of D^* and let $I = \{i \in [n] \mid xy \in D_i'(u, v)\}$. For each $i \in I$, the pair $\{x, y\}$ is q_0-dense in G_i by the definition of $D_i'(u, v)$. Therefore, by Claim 5.16, we have $|I| < t$. So xy has multiplicity less than t in D^*. Since xy is arbitrary, the multiplicity of each edge of D^* is less than t.

Next, observe that D^* contains no t-frame. Indeed, otherwise without loss of generality we may assume that D^* contains t edges $x_1y_1, \ldots, x_ty_t$, where x_iy_i has color i for each $i \in [t]$ and $y_1, \ldots, y_t$ are distinct. For each $i \in [t]$ since $x_iy_i \in D_i'(u, v)$, in particular $y_i \in N_i(u, v)$ (where $N_i(u, v)$ denotes the common neighborhood of u and v in G_i), which means that $uy_ic_i, vy_ic_i \in H$. But now, $\{uy_ic_i, vy_ic_i \mid i \in [t]\}$ forms a copy of $K_{2,t}^{(3)}$, contradicting H being $K_{2,t}^{(3)}$-free.

Therefore, applying Lemma 5.17, we have $|D^*| \leq \binom{t-1}{2}t + tq_0$. By Claim 5.5, we have

$$\frac{|F_i'(u, v)|}{4} \leq |D_i'(u, v)| .$$

So

$$\sum_{1 \leq i \leq n} \frac{|F_i'(u, v)|}{4} \leq \sum_{1 \leq i \leq n} |D_i'(u, v)| = |D^*| \leq \binom{t-1}{2}t + tq_0 < \frac{3}{2}t^3,$$

which proves the claim. □

By Lemma 5.15 we have

$$\sum_{1 \leq i \leq n} |G_i \setminus G_i'| < \frac{2}{q_0} \left(\sum_{u,v \in A} \sum_{1 \leq i \leq n} |F_i'(u, v)| + \sum_{u,v \in B} \sum_{1 \leq i \leq n} |F_i'(u, v)| \right) + 2tn^2.$$

Combining it with Claim 5.18 we get

$$\sum_{1\le i\le n} |G_i \setminus G'_i| < \frac{2}{q_0}\left(\sum_{u,v\in A} 6t^3 + \sum_{u,v\in B} 6t^3\right) + 2tn^2.$$

Therefore, as $q_0 > t^2/2$, we have

$$\sum_{1\le i\le n} |G_i \setminus G'_i| < \frac{4}{t^2}\left(12t^3\binom{n}{2}\right) + 2tn^2 < 26tn^2.$$

So,

$$\sum_{1\le i\le n} |G_i \setminus G'_i| = \sum_{1\le i\le n} |G_i[H](A,B) \setminus G'_i[H](A,B)| < 26tn^2.$$

By symmetry, using the same arguments, we have

$$\sum_{1\le i\le n} |G_i[H](B,C) \setminus G'_i[H](B,C)| < 26tn^2,$$

and

$$\sum_{1\le i\le n} |G_i[H](A,C) \setminus G'_i[H](A,C)| < 26tn^2.$$

Therefore, by Remark 5.9, we have

$$|H| - |H_0| < 78tn^2. \tag{5.10}$$

5.2.4 Making a $K_{1,2,q_j}$-free hypergraph $K_{1,2,q_{j+1}}$-free

In this subsection, we fix a j with $0 \le j < k$. Recall that H_j is $K_{1,2,q_j}$-free, and H_{j+1} is obtained by applying the $\mathcal{P}(q_{j+1})$ to H_j. Our goal in this subsection is to estimate $|H_j|-|H_{j+1}|$. The key difference between arguments in this subsection and in the previous subsection is that now in addition to H_j being $K_{2,t}^{(3)}$-free we can also utilize the fact that H_j is $K_{1,2,q_j}$-free. In particular, this extra condition leads to Claim 5.19, which improves upon Claim 5.18.

For convenience of notation, in this subsection, let $G_i = G_i[H_j](A,B)$ for each $1 \le i \le n$. For every $1 \le i \le n$ and every $u,v \in A$ or $u,v \in B$ let the sets $F'_i(u,v)$ and $D'_i(u,v)$ be defined by applying the procedure $\mathcal{P}(q_{j+1})$ to the graph G_i, to obtain the graph G'_i.

Claim 5.19. *Let $u,v \in A$ or $u,v \in B$. Then $\sum_{1\le i\le n} |F'_i(u,v)| < 2q_j t$.*

Proof. For each $i \in [n]$ we denote the set of common neighbors of u,v in G_i as $N_i(x,y)$. For each $i \in [n]$, since H_j is $K_{1,2,q_j}$-free, G_i is K_{2,q_j}-free and so $|N_i(u,v)| < q_j$.

Without loss of generality let us assume $u,v \in A$. For each vertex w of B, let $I_w = \{i \in \{1,2,\dots,n\} \mid w \in N_i(u,v)\}$. We claim that $|I_w| < q_j$. Indeed, for each $i \in I_w$, we have $uwc_i, vwc_i \in H_j$. So the set of hyperedges $\{uwc_i, vwc_i \mid i \in I_w\}$ form a copy of $K_{1,2,|I_w|}$ in H_j. Thus if $|I_w| \ge q_j$, then H_j contains a copy of $K_{1,2,q_j}$, a contradiction. Therefore, $|I_w| < q_j$, as desired.

Consider an auxiliary bipartite graph G_{AUX} with parts B and $[n]$ where the vertex $i \in [n]$ is adjacent to $b \in B$ in G_{AUX} if and only if $b \in N_i(u,v)$. Then by the discussion in the previous paragraph, each vertex $w \in B$ has degree $|I_w| < q_j$, and each vertex $i \in [n]$ has degree $|N_i(u,v)| < q_j$. In other words, the maximum degree in G_{AUX} is less than q_j.

We claim that G_{AUX} does not contain a matching of size t. Indeed, suppose for a contradiction that the edges $i_1 b_{i_1}, i_2 b_{i_2}, \ldots, i_t b_{i_t}$ (i.e., $b_{i_l} \in N_{i_l}(u,v)$ for $1 \leq l \leq t$) form a matching of size t in G_{AUX}. Then the set of hyperedges $ub_{i_l}c_{i_l}, vb_{i_l}c_{i_l}$, $1 \leq l \leq t$, form a copy of $K_{2,t}^{(3)}$ in H_j, a contradiction, as desired.

Since G_{AUX} does not contain a matching of size t, by the König-Egerváry theorem it has a vertex cover of size less than t. This fact combined with the fact that the maximum degree of G_{AUX} is less than q_j, implies that the number of edges of G_{AUX} is less than $q_j t$. On the other hand, the number of edges in G_{AUX} is $\sum_{i \in [n]} |N_i(u,v)|$. Therefore, $\sum_{i \in [n]} |N_i(u,v)| < q_j t$. This, combined with the fact that for each $i \in [n]$, $|N_i(u,v)| \geq |F'_i(u,v)|/2$ (see Claim 5.5), completes the proof of the lemma. □

By Lemma 5.15, we have

$$\sum_{1 \leq i \leq n} |G_i \setminus G'_i| \leq \frac{2}{q_{j+1}} \left(\sum_{u,v,\in A} \sum_{1 \leq i \leq n} |F'_i(u,v)| + \sum_{u,v,\in B} \sum_{1 \leq i \leq n} |F'_i(u,v)| \right) + 2tn^2.$$

Now using Claim 5.19, we have

$$\sum_{1 \leq i \leq n} |G_i \setminus G'_i| \leq \frac{8q_j t}{q_{j+1}} \binom{n}{2} + 2tn^2 < \frac{4tq_j}{q_{j+1}} n^2 + 2tn^2.$$

Since $q_{j+1} = q_j/2$, we have

$$\sum_{1 \leq i \leq n} |G_i \setminus G'_i| < 8tn^2 + 2tn^2 = 10tn^2.$$

So,

$$\sum_{1 \leq i \leq n} |G_i \setminus G'_i| = \sum_{1 \leq i \leq n} |G_i[H_j](A,B) \setminus G'_i[H_j](A,B)| < 10tn^2.$$

By symmetry, using the same arguments, we have

$$\sum_{1 \leq i \leq n} |G_i[H_j](B,C) \setminus G'_i[H_j](B,C)| < 10tn^2,$$

and

$$\sum_{1 \leq i \leq n} |G_i[H_j](A,C) \setminus G'_i[H_j](A,C)| < 10tn^2.$$

Therefore, by Remark 5.9, we have

$$|H_j| - |H_{j+1}| < 30tn^2. \tag{5.11}$$

5.2.5 Putting it all together

By (5.10) and (5.11) we have

$$|H| - |H_k| = |H| - |H_0| + \sum_{0 \le j < k} (|H_j| - |H_{j+1}|) < 78tn^2 + k(30tn^2).$$

By (5.9) we have $k \le \log t$, so we obtain,

$$|H| - |H_k| < 78tn^2 + 30t \log tn^2. \tag{5.12}$$

Notice that H_k is $K_{1,2,q_k}$-free and $q_k < 2t$. Therefore H_k is $K_{1,2,2t}$-free. Moreover, we know that the hypergraph H_k is 3-partite and $K_{2,t}^{(3)}$-free with parts A, B, C (as it is a subhypergraph of H). Now we bound the size of H_k.

Claim 5.20. *We have $|H_k| \le 2tn^2$.*

Proof. Suppose for a contradiction that $|H_k| > 2tn^2$. For any pair $\{a, b\}$ of vertices with $a \in A$ and $b \in B$, let $\text{codeg}(a, b)$ denote the number of hyperedges of H_k containing the pair $\{a, b\}$. Then the number of copies of $K_{2,1,1}$ in H_k of the form $\{abc, a'bc\}$ where $a, a' \in A$, $b \in B$, $c \in C$ is

$$\sum_{\substack{b,c \\ b \in B, c \in C}} \binom{\text{codeg}(b, c)}{2}.$$

As the average codegree (over all the pairs $b \in B, c \in C$) is more than $2t$, by convexity, this expression is more than

$$\binom{2t}{2} n^2 > (2t-1)^2 \binom{n}{2}.$$

This means there exist a pair $a, a' \in A$ and a set of $(2t-1)^2 + 1 > (t-1)(2t-1) + 1$ pairs $S := \{bc \mid b \in B, c \in C\}$ such that $abc, a'bc \in E(H_k)$ whenever $bc \in S$. Let G_{AUX} be a bipartite graph whose edges are elements of S. Since G_{AUX} has $|S| \ge (t-1)(2t-1) + 1$ edges, it either contains a matching M with t edges or a vertex v of degree $2t$ (see Lemma A.3 in [64] or the last paragraph of our proof of Claim 5.19 for a proof). In the former case, the set of all hyperedges of the form $abc, a'bc$ with $bc \in M$, form a copy of $K_{2,t}^{(3)}$ in H_k, a contradiction. In the latter case, let $u_1, u_2, \ldots, u_{2t}$ be the neighbors of v in G_{AUX}. Then the set of hyperedges $\{avu_i, a'vu_i \mid 1 \le i \le 2t\}$ form a copy of $K_{1,2,2t}$ in H_k, a contradiction again. This completes the proof of the claim. □

Combining (5.12) with Claim 5.20, we have $|H| \le 80tn^2 + 30t \log tn^2$, thus proving (5.3), which implies Theorem 5.1, as desired.

5.3 Remarks

Recall that given a bipartite graph G with an ordered bipartition (X, Y), where $Y = \{y_1, \ldots, y_m\}$, $G^{(r)}_{X,Y}$ is the r-graph with vertex set $(X \cup Y) \cup (\bigcup_{i=1}^m Y_i)$ and edge set $\bigcup_{i=1}^m \{e \cup Y_i : e \in E(G), y_i \in e\}$, where $Y_1, \ldots, Y_m$ are disjoint $(r-2)$-sets that are disjoint from $X \cup Y$. A standard reduction argument such as the one used in the proof of Theorem 1.4 in [64] can be used to show the following.

Proposition 5.21. *Let $n, r \geq 3$ be integers and G a bipartite graph with an ordered bipartition (X, Y). There exists a constant c_r depending only on r such that*

$$\mathrm{ex}(n, G^{(r)}_{X,Y}) \leq c_r n^{r-3} \cdot \mathrm{ex}(n, G^{(3)}_{X,Y}).$$

Thus, by Theorem 5.1 and Proposition 5.21, for all $r \geq 4$, we have $\mathrm{ex}(n, K^{(r)}_{2,t}) \leq c_r t \log t \binom{n}{r-1}$ for some constant c_r, depending only on r. On the other hand, taking the family of all r-element subsets of $[n]$ containing a fixed element shows that $\mathrm{ex}(n, K^{(r)}_{2,t}) \geq \binom{n-1}{r-1}$. Recall that in the $r = 3$ case, a better lower bound of $\Omega(t\binom{n}{2})$ was shown by Mubayi and Verstraëte [64]. For $r = 4$, we are able to improve the lower bound to $\Omega(t\binom{n}{3})$ as follows.

Proposition 5.22. *We have*

$$\mathrm{ex}(n, K^{(4)}_{2,t}) \geq (1 + o(1))\frac{t-1}{8}n^3.$$

Proof. (Sketch.) Consider a $K_{2,t}$-free graph G with $(1 + o(1))\frac{\sqrt{t-1}}{2}n^{3/2}$ edges where each vertex has degree $(1 + o(1))\sqrt{(t-1)}\sqrt{n}$. (Such a graph exists by a construction of Füredi [38].) Let us a define a 4-graph $H = \{abcd \mid ab, cd \in G \text{ and } ac, ad, bc, bd \notin G\}$. In other words, let the edges of H be the vertex sets of induced 2-matchings in G. Via standard counting, it is easy to show that $|H| = (1 + o(1))\frac{t-1}{8}n^3$. It remains to show H is $K^{(4)}_{2,t}$-free.

Claim 5.23. *If $axyz, bxyz \in H$, then there is a vertex $c \in \{x, y, z\}$ such that $ac, bc \in G$.*

Proof. By our assumption, $\{a, x, y, z\}$ and $\{b, x, y, z\}$ both induce a 2-matching in G. Without loss of generality, suppose $ax, yz \in G$. If $bx \in G$ then we are done. Otherwise, we have $by, xz \in G$ or $bz, xy \in G$, both contradicting $\{ax, yz\}$ being an induced matching in G. □

Suppose for contradiction that H has a copy of $K^{(4)}_{2,t}$ with edge set $\{ax_iy_iz_i, bx_iy_iz_i \mid 1 \leq i \leq t\}$. By Claim 5.23, for each $1 \leq i \leq t$, there exists a vertex $w_i \in \{x_i, y_i, z_i\}$ such that $aw_i, bw_i \in G$. This yields a copy of $K_{2,t}$ in G, a contradiction. □

For $r \geq 5$, we do not yet have a lower bound that is asymptotically larger than $\binom{n-1}{r-1}$. It would be interesting to narrow the gap between the lower and upper bounds on $\mathrm{ex}(n, K^{(r)}_{2,t})$.

It will be interesting to have a systematic study of the function $\mathrm{ex}(n, G^{(r)}_{X,Y})$. Mubayi and Verstraëte [64] showed that $\mathrm{ex}(n, K^{(3)}_{s,t}) = O(n^{3-1/s})$ and that if $t > (s-1)! > 0$ then

$\mathrm{ex}(n, K^{(3)}_{s,t}) = \Omega(n^{3-2/s})$ and speculated that $n^{3-2/s}$ is the correct order of magnitude. The case when G is a tree is studied in [37], where the problem considered there is slightly more general. The case when G is an even cycle has also been studied. Let $C^{(r)}_{2t}$ denote $G^{(r)}_{X,Y}$ where G is the even cycle C_{2t} of length $2t$. It was shown by Jiang and Liu [52] that $c_1 t\binom{n}{r-1} \le \mathrm{ex}(n, C^{(r)}_{2m}) \le c_2 t^5 \binom{n}{r-1}$, for some positive constants depending c_1, c_2 on r. Using results in this chapter and new ideas, we are able to narrow the gap to $c_1 t\binom{n}{r-1} \le \mathrm{ex}(n, C^{(r)}_{2m}) \le c_2 t^2 \log t\binom{n}{r-1}$, for some positive constants c_1, c_2 depending on r.

Chapter 6

On the Rainbow Turán number of paths

6.1 Introduction

For an integer k, let P_k denote a path of length k, where the length of a path is defined as the number of edges in it. Erdős and Gallai [20] proved that $\mathrm{ex}(n, P_{k+1}) \leq \frac{k}{2}n$; moreover, they showed that if $k+1$ divides n, then the unique extremal graph is the vertex-disjoint union of $\frac{n}{k+1}$ copies of K_{k+1}.

On the other hand, Keevash, Mubayi, Sudakov and Verstraëte [55] showed that in some cases, the rainbow Turán number of P_k can be strictly larger than the usual Turán number of P_k: Maamoun and Meyniel [62] gave an example of a proper coloring of K_{2^k} containing no rainbow path with $2^k - 1$ edges. By taking a vertex-disjoint union of such K_{2^k}'s, Keevash et. al. showed that $\mathrm{ex}^*(n, P_{2^k-1}) \geq \binom{2^k}{2}\left\lfloor \frac{n}{2^k} \right\rfloor = (1+o(1))\frac{2^k-1}{2^k-2}\,\mathrm{ex}(n, P_{2^k-1})$– so $\mathrm{ex}^*(n, P_{2^k-1})$ is not asymptotically equal to $\mathrm{ex}(n, P_{2^k-1})$. They also mentioned that determining the asymptotic behavior of $\mathrm{ex}^*(n, P_{k+1})$ is an interesting open problem, and stated the natural conjecture that the optimal construction is a disjoint union of cliques of size $c(k)$, where $c(k)$ is chosen as large as possible so that the cliques can be properly colored with no rainbow P_{k+1}. For P_4, this conjecture was disproved by Johnston, Palmer and Sarkar [53]: Since any properly edge-colored K_5 contains a rainbow P_4, and K_4 does not contain a P_4, the conjecture for P_4 would be that $\mathrm{ex}^*(n, P_4) \sim \frac{3n}{2}$. But they show that in fact, $\mathrm{ex}^*(n, P_4) \sim 2n$ by showing a proper edge-coloring of $K_{4,4}$ without rainbow P_4, and then taking $\frac{n}{8}$ vertex-disjoint copies of $K_{4,4}$. For general k, they proved the following:

Theorem 6.1 (Johnston, Palmer and Sarkar [53]). *For any positive integer k, we have*

$$\frac{k}{2}n \leq \mathrm{ex}^*(n, P_{k+1}) \leq \left\lceil \frac{3k+1}{2} \right\rceil n.$$

We improve the above bound by showing the following:

Theorem 6.2 (E., Győri, Methuku [34]). *For any positive integer k, we have*

$$\mathrm{ex}^*(n, P_{k+1}) < \left(\frac{9k}{7} + 2\right) n.$$

Let us remark that using the ideas introduced in this chapter, it is conceivable that the upper bound can be further improved (at the cost of making the proof very involved). However, it would be very interesting (and seems to be difficult) to prove an upper bound less than kn or construct an example with kn edges.

We give a construction which shows that $\mathrm{ex}^*(n, P_{2^k}) > \mathrm{ex}(n, P_{2^k})$ for any $k \geq 2$.

Construction. Let us first show a proper edge-coloring of $K_{2^k,2^k}$ (a complete bipartite graph with parts A and B, each of size 2^k) with no rainbow P_{2^k}. The vertices of A and B are both identified with the vectors $\mathbb{F}_2^k$. Each edge uv with $u \in A$ and $v \in B$ is assigned the color $c(uv) := u - v$. Clearly this gives a proper edge-coloring of $K_{2^k,2^k}$. Moreover, if it contains a rainbow path $v_0 v_1 \ldots v_{2^k}$ then such a path must use all of the colors from $\mathbb{F}_2^k$. Therefore $\sum_{i=0}^{2^k-1} c(v_i v_{i+1}) = 0$. On the other hand, $\sum_{i=0}^{2^k-1} c(v_i v_{i+1}) = \sum_{i=0}^{2^k-1}(v_i - v_{i+1}) = v_0 - v_{2^k}$. Thus, $v_0 - v_{2^k} = 0$. But notice that since the length of the path $v_0 v_1 \ldots v_{2^k}$ is even, its terminal vertices v_0 and v_{2^k} are either both in A or they are both in B. So they could not have been identified with the same vector in $\mathbb{F}_2^k$, a contradiction. Taking a vertex-disjoint union of such $K_{2^k,2^k}$'s we obtain that $\mathrm{ex}^*(n, P_{2^k}) \geq (2^k)^2 \left\lfloor n/2^{k+1} \right\rfloor = (1 + o(1)) \frac{2^k}{2^k-1} \mathrm{ex}(n, P_{2^k})$.

Remark. This construction provides a counterexample to the above mentioned conjecture of Keevash, Mubayi, Sudakov and Verstraëte [55] whenever the largest clique that can be properly colored without a rainbow P_{2^k} has size 2^k. This is the case for $k = 2$, as noted before. The question of determining whether this is the case for any $k \geq 3$ remains an interesting open question (see [2] for results in this direction).

Overview of the proof and organization. Let G be a graph which has a proper edge-coloring with no rainbow P_{k+1}. By induction on the length of the path, we assume there is a rainbow path $v_0 v_1 \ldots v_k$ in G. Roughly speaking, we will show that the sum of degrees of the terminal vertices of the path, v_0 and v_k is small. Our strategy is to find a set of distinct vertices $M := \{a_1, b_1, a_2, b_2, \ldots, a_m, b_m\} \subseteq \{v_0, v_1, \ldots, v_k\}$ (whose size is as large as possible) such that for each $1 \leq i \leq m$, there is a rainbow path P of length k with a_i and b_i as terminal vertices and $V(P) = \{v_0, v_1, \ldots, v_k\}$; then we show that there are not many edges of G incident to the vertices of M, which will allow us to delete the vertices of M from G and apply induction. To this end, we define the set $T \subseteq \{v_0, v_1, \ldots, v_k\}$ as the set of all vertices $v \in \{v_0, v_1, \ldots, v_k\}$ where v is a terminal vertex of some rainbow path P with $V(P) = \{v_0, v_1, \ldots, v_k\}$; we call T the set of *terminal vertices*. We will then find M as a subset of T; moreover, it will turn out that if the size of T is large, then the size of M is also large – therefore, the heart of the proof lies in showing that T is large.

In Section 6.2.1, we introduce the notation and prove some basic claims. Using these claims, in Section 6.2.2, we will show that T is large (i.e., that there are many terminal vertices). Then in Section 6.2.3 we will find the desired subset M of T (which has few edges incident to it).

6.2 Proof of Theorem 6.2

Let G be a graph on n vertices, and suppose it has a proper edge-coloring $c : E(G) \to \mathbb{N}$ without a rainbow path of length $k+1$. Consider a longest rainbow path P^* in G. We may suppose it is of length k, otherwise we are done by induction on k. For the base case $k = 1$, notice that any path of length 2, has to be a rainbow path. Thus G can contain at most $\frac{n}{2} < (\frac{9}{7} + 2)n$ edges, so we are done.

6.2.1 Basic claims and Notation

In the rest of the chapter, the degree of a vertex $v \in V(G)$ be denoted by $d(v)$.

Definition 6.1. *Let $P^* = v_0v_1 \dots v_k$. Suppose the color of the edge $v_{i-1}v_i$ is $c(v_{i-1}v_i) = c_i$ for each $1 \le i \le k$. Let L and R denote the sets of colors of edges incident to v_0 and v_k respectively. (Notice that since the edges of G are colored properly, we have $|L| = d(v_0)$ and $|R| = d(v_k)$.)*

We define the following subsets of L, R and $\{c_1, c_2, \dots, c_k\}$ corresponding to P^*.

- *Let L_{out} (respectively R_{out}) be the set of colors of the edges incident to v_0 (respectively v_k) and to a vertex outside P^*.*

  Note that $L_{out} \subseteq \{c_1, c_2, \dots, c_k\}$ and $R_{out} \subseteq \{c_1, c_2, \dots, c_k\}$, otherwise we can extend P^* to a rainbow path longer than k in G.

- *Let $L_{in} = L \setminus L_{out}$ and $R_{in} = R \setminus R_{out}$.*

- *Let $L_{old} = L \cap \{c_1, c_2, \dots, c_k\}$ and $L_{new} = L \setminus \{c_1, c_2, \dots, c_k\}$. Similarly, let $R_{old} = R \cap \{c_1, c_2, \dots, c_k\}$, $R_{new} = R \setminus \{c_1, c_2, \dots, c_k\}$.*

- *Let $S_L = \{c(v_{j-1}v_j) = c_j \mid v_0v_j \in E(G)$ and $c(v_0v_j) \in L_{new}$ and $2 \le j \le k\}$ and $S_R = \{c(v_jv_{j+1}) = c_{j+1} \mid v_kv_j \in E(G)$ and $c(v_kv_j) \in R_{new}$ and $0 \le j \le k-2\}$.*

  Notice that $|S_L| = |L_{new}|$ and $|S_R| = |R_{new}|$.

- *Let $L_{nice} = L \cap S_R$ and let $R_{nice} = R \cap S_L$. (Note that since $L_{nice} \subseteq \{c_1, c_2, \dots, c_k\}$, we have $L_{nice} \cap L_{new} = \emptyset$. Similarly $R_{nice} \cap R_{new} = \emptyset$.)*

- *Let $L_{res} = L_{in} \setminus (L_{new} \cup L_{nice}) = L_{old} \setminus (L_{nice} \cup L_{out})$, and $R_{res} = R_{in} \setminus (R_{new} \cup R_{nice}) = R_{old} \setminus (R_{nice} \cup R_{out})$.*

Notation 6.2. For convenience, we let $|L| = l$ and $|R| = r$. Moreover, let $|L_{\text{out}}| = l_{\text{out}}, |L_{\text{old}}| = l_{\text{old}}, |L_{\text{nice}}| = l_{\text{nice}}, |L_{\text{new}}| = l_{\text{new}}$ and $|R_{\text{out}}| = r_{\text{out}}, |R_{\text{old}}| = r_{\text{old}}, |R_{\text{nice}}| = r_{\text{nice}}, |R_{\text{new}}| = r_{\text{new}}$.

Note that

$$d(v_0) = l_{\text{in}} + l_{\text{out}} = l_{\text{new}} + l_{\text{old}} = l$$

and

$$d(v_k) = r_{\text{in}} + r_{\text{out}} = r_{\text{new}} + r_{\text{old}} = r.$$

Now we prove some inequalities connecting the quantities defined in Definition 6.1 for the path P^*.

Claim 6.3. *$L_{out} \cap S_R = \emptyset = R_{out} \cap S_L$. This implies that $L_{out} \cap L_{nice} = \emptyset = R_{out} \cap R_{nice}$ (since $L_{nice} \subset S_R$ and $R_{nice} \subset S_L$).*

Proof of Claim. Suppose for a contradiction that $L_{\text{out}} \cap S_R \neq \emptyset$. So there exists a vertex $w \notin \{v_0, v_1, \ldots, v_k\}$ such that $c(v_k v_j) \in R_{\text{new}}$ and $c(wv_0) = c(v_j v_{j+1})$ for some $0 \leq j \leq k-2$. Consider the path $v_{j+1}v_{j+2}\ldots v_k v_j v_{j-1}\ldots v_0 w$. The set of colors of the edges in this path is $\{c_1, c_2, \ldots, c_k\} \setminus \{c(v_j v_{j+1})\} \cup \{c(wv_0), c(v_k v_j)\} = \{c_1, c_2, \ldots, c_k\} \cup \{c(v_k v_j)\}$, so it is a rainbow path of length $k+1$ in G, a contradiction.

Similarly, by a symmetric argument, we have $R_{\text{out}} \cap S_L = \emptyset$. □

Claim 6.4. *$l_{out} \leq k - r_{new}$ and $r_{out} \leq k - l_{new}$.*

Proof of Claim. By Claim 6.3, $L_{\text{out}} \cap S_R = \emptyset$. Since both L_{out} and S_R are subsets of $\{c_1, c_2, \ldots, c_k\}$, this implies, $|L_{\text{out}}| = l_{\text{out}} \leq k - |S_R| = k - r_{\text{new}}$, as desired. Similarly, $r_{\text{out}} \leq k - l_{\text{new}}$. □

We will prove Theorem 6.2 by induction on the number of vertices n. For the base cases, note that for all $n \leq k$, the number of edges is trivially at most

$$\binom{n}{2} \leq \frac{kn}{2} < \left(\frac{9k}{7} + 2\right) n,$$

so the statement of the theorem holds. If $d(v) < \frac{9k}{7} + 2$ for some vertex v of G, then we delete v from G to obtain a graph G' on $n-1$ vertices. By induction hypothesis, the number of edges in G' is less than $(\frac{9k}{7} + 2)(n-1)$. So the total number of edges in G is less than $(\frac{9k}{7} + 2)n$, as desired.

Therefore, from now on, we assume that for all $v \in V(G)$,

$$d(v) \geq \frac{9k}{7} + 2.$$

Since $d(v_0) = l = l_{\text{old}} + l_{\text{new}}$ and $l_{\text{old}} \leq k$, we have that

$$l_{\text{new}} \geq \frac{2k}{7} + 2. \tag{6.1}$$

Similarly,

$$r_{\text{new}} \geq \frac{2k}{7} + 2. \tag{6.2}$$

Claim 6.5. *We have*

$$l_{nice} + r_{nice} \geq \frac{4k}{7} + 4.$$

Proof of Claim. First notice that $L_{\text{res}} \cap S_R = \emptyset$. Indeed, by definition, $L_{\text{res}} \cap S_R = (L_{\text{res}} \cap L) \cap S_R = L_{\text{res}} \cap (L \cap S_R) = L_{\text{res}} \cap L_{\text{nice}} = \emptyset$. Moreover, by Claim 6.3, $L_{\text{out}} \cap S_R = \emptyset$. Therefore, we have $(L_{\text{res}} \cup L_{\text{out}}) \cap S_R = \emptyset$. Moreover, $(L_{\text{res}} \cup L_{\text{out}}) \cup S_R \subseteq \{c_1, c_2, \ldots, c_k\}$. Therefore, $l_{\text{res}} + l_{\text{out}} \leq k - |S_R| = k - r_{\text{new}}$. On the other hand, by definition, $l_{\text{res}} + l_{\text{out}} \geq (l_{\text{in}} - l_{\text{new}} - l_{\text{nice}}) + l_{\text{out}} = l - l_{\text{new}} - l_{\text{nice}}$. So we have,

$$l - l_{\text{new}} - l_{\text{nice}} \leq k - r_{\text{new}}.$$

By a symmetric argument, we get

$$r - r_{\text{new}} - r_{\text{nice}} \leq k - l_{\text{new}}.$$

Adding the above two inequalities and rearranging, we get $l + r - l_{\text{nice}} - r_{\text{nice}} \leq 2k$, so

$$l_{\text{nice}} + r_{\text{nice}} \geq l + r - 2k = d(v_0) + d(v_k) - 2k \geq \frac{4k}{7} + 4,$$

as required. □

6.2.2 Finding many terminal vertices

Definition 6.6 (Set of terminal vertices)**.** *Let T be the set of all vertices $v \in \{v_0, v_1, v_2, \ldots, v_k\}$ such that v is a terminal vertex of some rainbow path P with $V(P) = \{v_0, v_1, v_2, \ldots, v_k\}$.*

For convenience, we will denote the size of T by t.

The next lemma yields a lower bound on the number of terminal vertices and is crucial to the proof of Theorem 6.2.

Lemma 6.7. *We have*

$$|T| = t \geq \frac{3k}{7} + 2.$$

The rest of this subsection is devoted to the proof of Lemma 6.7.

Proof of Lemma 6.7

Recall that $P^* = v_0 v_1 \ldots v_k$ and $c(v_j v_{j+1}) = c_j$. First we prove a simple claim.

Claim 6.8. *We may assume $c(v_0 v_1) \notin L_{nice}$ and $c(v_k v_{k-1}) \notin R_{nice}$. Moreover, if $v_0 v_k$ is an edge of G, we can assume $c(v_0 v_k) \notin L_{new} \cup R_{new}$.*

Proof of Claim. First consider the case when $v_0 v_k$ is an edge of G. If $c(v_0 v_k) \in L_{\text{new}} \cup R_{\text{new}}$, then every vertex $v_i \in T$. Indeed, the path $v_i v_{i-1} v_{i-2} \ldots v_0 v_k v_{k-1} \ldots v_{i+1}$ is a rainbow path with v_i as a terminal vertex. Thus $|T| = k + 1 \geq \frac{3k}{7} + 2$, and we are done. So we can assume $c(v_0 v_k) \notin L_{\text{new}} \cup R_{\text{new}}$. This implies that $c(v_0 v_1) \notin L_{\text{nice}}$ and $c(v_k v_{k-1}) \notin R_{\text{nice}}$, because $c(v_0 v_1) \notin S_R$ and $c(v_k v_{k-1}) \notin S_L$.

Now if $v_0 v_k$ is not an edge of G, then again $c(v_0 v_1) \notin S_R$ and $c(v_k v_{k-1}) \notin S_L$, so the claim follows. □

Claim 6.9. *If v_0v_i is an edge such that $c(v_0v_i) \in L_{new}$ then $v_{i-1} \in T$.*

Proof of Claim. Consider the path $v_{i-1}v_{i-2}\dots v_0v_iv_{i+1}\dots v_k$. Clearly it is a rainbow path of length k in which v_{i-1} is a terminal vertex. □

Suppose v_0v_i is an edge such that $c(v_0v_i) \in L_{\text{nice}}$. Since $c(v_0v_k) \notin R_{\text{new}}$, by the definition of L_{nice}, there exists an integer j (with $1 \leq j \leq k-2$) such that $c(v_kv_j) \in R_{\text{new}}$ and $c(v_0v_i) = c(v_jv_{j+1}) = c_j$.

Claim 6.10. *If $c(v_0v_i) \in L_{nice}$ then $v_{i-1} \in T$ or $v_{i+1} \in T$.*

Moreover, let j be an integer (with $1 \leq j \leq k-2$) such that $c(v_kv_j) \in R_{new}$ and $c(v_0v_i) = c(v_jv_{j+1}) = c_j$.

If $j \geq i$, then $v_{i-1} \in T$, and if $j < i$ then $v_{i+1} \in T$.

Proof of Claim. Observe that since $c(v_0v_i) \in L_{\text{nice}} \subset S_R$, we have that $c(v_kv_j) \in R_{\text{new}}$ (by definition of S_R).

First let $j \geq i$. In this case consider the path $v_{i-1}v_{i-2}\dots v_0v_iv_{i+1}\dots v_jv_kv_{k-1}\dots v_{j+1}$. It is easy to see that the set of colors of the edges in this path is $\{c_1, c_2, \dots, c_k\} \setminus \{c_i\} \cup \{c(v_jv_k)\}$. As $c(v_jv_k) \in R_{\text{new}}$, the path is rainbow with v_{i-1} as a terminal vertex. So $v_{i-1} \in T$.

If $j < i$, then consider the path $v_{j+1}v_{j+2}\dots v_iv_0v_1\dots v_jv_kv_{k-1}\dots v_{i+1}$. It is easy to see that the set of colors of the edges in this path is $\{c_1, c_2, \dots, c_k\} \setminus \{c_{i+1}\} \cup \{c(v_jv_k)\}$, so the path is rainbow again, with v_{i+1} as a terminal vertex. So $v_{i+1} \in T$. □

By symmetry, one can see that the same arguments used in the proofs of Claim 6.9 and Claim 6.10, imply the following two statements.

Observation 6.11. If v_kv_i is an edge such that $c(v_kv_i) \in R_{\text{new}}$ then $v_{i+1} \in T$.

If $c(v_kv_i) \in R_{\text{nice}}$ then $v_{i-1} \in T$ or $v_{i+1} \in T$.

Definition 6.12. *Let $b' > b$ be the largest two integers such that $c(v_0v_b) \in L_{new}$ and $c(v_0v_{b'}) \in L_{new}$. Similarly, let $a' < a$ be the smallest two integers such that $c(v_kv_{a'}) \in R_{new}$ and $c(v_kv_a) \in R_{new}$.*

Notation 6.13. For any integers, $0 \leq x \leq y \leq k$, let

$$T^{x,y} = \{v_i \in T \mid x \leq i \leq y\},$$

and $|T^{x,y}| = t^{x,y}$.

Notice that $t = t^{0,k} = 2 + t^{1,k-1}$, as v_0 and v_k are both terminal vertices.

Now we will show that if $a > b$, then Lemma 6.7 holds. Suppose $a > b$. Then by the definition of a and b, we have

$$|\{i \mid 2 \leq i \leq b \text{ and } c(v_0v_i) \in L_{\text{new}}\}| = |L_{\text{new}}| - 1 = l_{\text{new}} - 1.$$

By Claim 6.9, we know that whenever $c(v_0v_i) \in L_{\text{new}}$, we have $v_{i-1} \in T$. This shows that $t^{1,b-1} \geq l_{\text{new}} - 1$. Similarly, by a symmetric argument (using Observation 6.11), we get $t^{a+1,k-1} \geq r_{\text{new}} - 1$. Therefore,

$$t = 2 + t^{1,k-1} = 2 + t^{1,b-1} + t^{b,a} + t^{a+1,k-1} \geq 2 + (l_{\text{new}} - 1) + (r_{\text{new}} - 1) = l_{\text{new}} + r_{\text{new}}.$$

Now using (6.1) and (6.2), we have

$$t = l_{\text{new}} + r_{\text{new}} \geq \frac{2k}{7} + 2 + \frac{2k}{7} + 2 = \frac{4k}{7} + 4,$$

proving Lemma 6.7. Therefore, from now on, we always assume $a \leq b$.

Claim 6.14. *If $c(v_0v_i) \in L_{new}$ or $c(v_kv_i) \in R_{new}$, and $a \leq i \leq b$, then $v_{i-1} \in T$ and $v_{i+1} \in T$.*

Proof of Claim. First suppose $c(v_0v_i) \in L_{\text{new}}$. Then by Claim 6.9, $v_{i-1} \in T$. We want to show that $v_{i+1} \in T$.

Observe that if $i = a$, then by Claim 6.9 again, we have $v_{i+1} \in T$ because $v_kv_i \in R_{\text{new}}$. So let us assume $a < i$ and show that $v_{i+1} \in T$. Notice that there exists $a^* \in \{a, a'\}$ (see Definition 6.12 for the definition of a and a') such that $c(v_0v_i) \neq c(v_{a^*}v_k)$. Now consider the path $v_{a^*+1}v_{a^*+2} \dots v_iv_0v_1 \dots v_{a^*}v_kv_{k-1} \dots v_{i+1}$. The set of colors of the edges in this path are $\{c_1, c_2, \dots, c_k\} \setminus \{c_{a^*+1}, c_{i+1}\} \cup \{c(v_0v_i), c(v_{a^*}v_k)\}$, and it is easy to check that all the colors are different, so the path is rainbow with v_{i+1} as a terminal vertex.

Now suppose $c(v_kv_i) \in R_{\text{new}}$. Then a similar argument (using Observation 6.11) shows that $v_{i-1} \in T$ and $v_{i+1} \in T$ again, completing the proof of the claim. □

Now we introduce some helpful notation.

Notation 6.15. For any integers, $0 \leq x \leq y \leq k$, let

$$L^{x,y}_{\text{nice}} = \{c(v_0v_i) \in L_{\text{nice}} \mid x \leq i \leq y\},$$

$$R^{x,y}_{\text{nice}} = \{c(v_kv_i) \in R_{\text{nice}} \mid x \leq i \leq y\},$$

$$L^{x,y}_{\text{new}} = \{c(v_0v_i) \in L_{\text{new}} \mid x \leq i \leq y\},$$

$$R^{x,y}_{\text{new}} = \{c(v_kv_i) \in R_{\text{new}} \mid x \leq i \leq y\},$$

Moreover, let $|L^{x,y}_{\text{nice}}| = l^{x,y}_{\text{nice}}$, $|R^{x,y}_{\text{nice}}| = r^{x,y}_{\text{nice}}$, $|L^{x,y}_{\text{new}}| = l^{x,y}_{\text{new}}$, $|R^{x,y}_{\text{new}}| = r^{x,y}_{\text{new}}$.

Note that by definition of a and b, $l_{\text{new}} = l^{0,a-1}_{\text{new}} + l^{a,b}_{\text{new}} + 1$ and $r_{\text{new}} = 1 + r^{a,b}_{\text{new}} + r^{b+1,k}_{\text{new}}$. Using Claim 6.8, for any integer z, we have the following:

$$L^{0,z}_{\text{nice}} = L^{2,z}_{\text{nice}} \text{ and } R^{z,k}_{\text{nice}} = R^{z,k-2}_{\text{nice}}. \tag{6.3}$$

Moreover, by definition of L_{new} and R_{new}, we have

$$L^{0,z}_{\text{new}} = L^{2,z}_{\text{new}} \text{ and } R^{z,k}_{\text{new}} = R^{z,k-2}_{\text{new}}. \tag{6.4}$$

Informally speaking, Claim 6.10 and Claim 6.14 assert that each edge $e = v_0v_i$ such that $c(v_0v_i) \in L_{\text{new}} \cup L_{\text{nice}}$ “creates" a terminal vertex $x = v_{i-1} \in T$ or $x = v_{i+1} \in T$ (or sometimes both). Similarly, (using Observation 6.11) each edge $e = v_kv_i$ such that $c(v_kv_i) \in R_{\text{new}} \cup R_{\text{nice}}$ “creates" a terminal vertex $x = v_{i-1} \in T$ or $x = v_{i+1} \in T$ (or both). In the next two claims, by double counting the total number of such pairs (e, x), we prove lower bounds on the number of terminal vertices in different ranges (i.e., $t^{0,a-1}, t^{b+1,k}$ and $t^{a,b}$), in terms of $l_{\text{new}}, r_{\text{new}}, l_{\text{nice}}$ and r_{nice}.

Claim 6.16. *We have,*

$$t^{0,a-1} \geq \frac{1}{2}\left(l^{0,a}_{nice} + l^{0,a}_{new} + \frac{r^{0,a}_{nice}}{2}\right),$$

and

$$t^{b+1,k} \geq \frac{1}{2}\left(r^{b,k}_{nice} + r^{b,k}_{new} + \frac{l^{b,k}_{nice}}{2}\right).$$

Proof of Claim. By Claim 6.10, and by the fact that there is only one j such that $c(v_k v_j) \in R^{0,a-1}_{\text{new}}$, it is easy to see that for all but at most one i, we have the following: if $c(v_0 v_i) \in L^{0,a}_{\text{nice}} = L^{2,a}_{\text{nice}}$ (equality here follows from (6.3)), then $v_{i-1} \in T^{1,a-1}$. So there are at least $l^{2,a}_{\text{nice}} - 1$ pairs $(v_0 v_i, x)$ such that $c(v_0 v_i) \in L^{2,a}_{\text{nice}}$ and $x = v_{i-1} \in T^{1,a-1}$.

If $c(v_0 v_i) \in L^{0,a}_{\text{new}} = L^{2,a}_{\text{new}}$ (equality here follows from (6.4)), then by Claim 6.9, $v_{i-1} \in T^{1,a-1}$. So there are $l^{2,a}_{\text{new}}$ pairs $(v_0 v_i, x)$ such that $c(v_0 v_i) \in L^{2,a}_{\text{new}}$ and $x = v_{i-1} \in T^{1,a-1}$.

Adding the previous two bounds, the total number of pairs $(v_0 v_i, x)$ such that $c(v_0 v_i) \in L^{0,a}_{\text{nice}} \cup L^{0,a}_{\text{new}} = L^{2,a}_{\text{nice}} \cup L^{2,a}_{\text{new}}$ and $x = v_{i-1} \in T^{1,a-1}$, is at least $l^{2,a}_{\text{nice}} - 1 + l^{2,a}_{\text{new}}$. This implies $t^{1,a-1} \geq l^{2,a}_{\text{nice}} - 1 + l^{2,a}_{\text{new}}$. Therefore, using that v_0 is also a terminal vertex, we have

$$t^{0,a-1} \geq l^{2,a}_{\text{nice}} + l^{2,a}_{\text{new}}. \tag{6.5}$$

If $c(v_k v_i) \in R^{0,a-1}_{\text{nice}}$, then by Observation 6.11, there is a vertex $x \in \{v_{i-1}, v_{i+1}\}$ such that $x \in T$. So the number of pairs $(v_k v_i, x)$ such that $c(v_k v_i) \in R^{0,a-1}_{\text{nice}}$, $x \in \{v_{i-1}, v_{i+1}\}$ and $x \in T$, is at least $r^{0,a-1}_{\text{nice}}$. By the pigeonhole principle, either the number of pairs $(v_k v_i, v_{i-1})$ with $c(v_k v_i) \in R^{0,a-1}_{\text{nice}}$, $v_{i-1} \in T$, or the number of pairs $(v_k v_i, v_{i+1})$ with $c(v_k v_i) \in R^{0,a-1}_{\text{nice}}$,$v_{i+1} \in T$, is at least $r^{0,a-1}_{\text{nice}}/2$. In the first case, we get $t^{0,a-2} \geq r^{0,a-1}_{\text{nice}}/2$ and in the second case, we get $t^{1,a} \geq r^{0,a-1}_{\text{nice}}/2$. As $t^{0,a-1} \geq t^{0,a-2}$ and $t^{0,a-1} \geq t^{1,a}$, in both cases we have,

$$t^{0,a-1} \geq \frac{r^{0,a-1}_{\text{nice}}}{2}. \tag{6.6}$$

Therefore, adding up (6.5) and (6.6), we get

$$2t^{0,a-1} \geq l^{2,a}_{\text{nice}} + l^{2,a}_{\text{new}} + \frac{r^{0,a-1}_{\text{nice}}}{2} = l^{0,a}_{\text{nice}} + l^{0,a}_{\text{new}} + \frac{r^{0,a}_{\text{nice}}}{2}.$$

Note that the equality follows from (6.3), (6.4) and the fact that $r^{0,a-1}_{\text{nice}} = r^{0,a}_{\text{nice}}$ because $c(v_k v_a) \in R_{\text{new}}$. By a symmetric argument, we have

$$2t^{b+1,k} \geq r^{b,k-2}_{\text{nice}} + r^{b,k-2}_{\text{new}} + \frac{l^{b+1,k}_{\text{nice}}}{2} = r^{b,k}_{\text{nice}} + r^{b,k}_{\text{new}} + \frac{l^{b,k}_{\text{nice}}}{2}.$$

This finishes the proof of the claim. □

Now we prove a lower bound on $t^{a,b}$.

Claim 6.17.

$$t^{a,b} \geq \frac{1}{4}\left(l^{a+1,b-1}_{nice} + r^{a+1,b-1}_{nice} + 2(l^{a+1,b}_{new} + r^{a,b-1}_{new})\right).$$

Proof of Claim. Let us construct a set S of pairs (e, x) such that $e \in L_{\text{in}} \cup R_{\text{in}}$ and $x \in T$ with certain properties.

For every edge e such that $c(e) \in L_{\text{nice}}^{a+1,b-1} \cup R_{\text{nice}}^{a+1,b-1}$, Claim 6.10 (and Observation 6.11) ensures that there is a vertex $x \in \{v_{i-1}, v_{i+1}\}$ such that $x \in T$ (in particular, $x \in T^{a,b}$). Add all such pairs (e, x) to S. Therefore, the number of pairs (e, x) added to S so far, is $l_{\text{nice}}^{a+1,b-1} + r_{\text{nice}}^{a+1,b-1}$.

For every edge e such that $c(e) \in L_{\text{new}}^{a+1,b} \cup R_{\text{new}}^{a,b-1}$, we have both $v_{i-1}, v_{i+1} \in T$ by Claim 6.14; we add both the pairs (e, v_{i-1}) and (e, v_{i+1}) to S. Therefore the number of pairs (e, x) added to S in this step is $2(l_{\text{new}}^{a+1,b} + r_{\text{new}}^{a,b-1})$. Thus,

$$|S| = l_{\text{nice}}^{a+1,b-1} + r_{\text{nice}}^{a+1,b-1} + 2(l_{\text{new}}^{a+1,b} + r_{\text{new}}^{a,b-1}).$$

Note that all the pairs (e, x) in S are such that $x \in T^{a,b}$. Moreover, for each $x \in T^{a,b}$, there are at most four pairs (e, x) in S. Therefore, we have

$$4t^{a,b} \geq |S| \geq l_{\text{nice}}^{a+1,b-1} + r_{\text{nice}}^{a+1,b-1} + 2(l_{\text{new}}^{a+1,b} + r_{\text{new}}^{a,b-1}),$$

finishing the proof of the claim. □

By Claim 6.16 and Claim 6.17, we have

$$2(2t^{0,a-1} + 2t^{b+1,k}) + 4t^{a,b} \geq 2\left(l_{\text{nice}}^{0,a} + l_{\text{new}}^{0,a} + \frac{r_{\text{nice}}^{0,a}}{2} + r_{\text{nice}}^{b,k} + r_{\text{new}}^{b,k} + \frac{l_{\text{nice}}^{b,k}}{2}\right)$$

$$+ l_{\text{nice}}^{a+1,b-1} + r_{\text{nice}}^{a+1,b-1} + 2(l_{\text{new}}^{a+1,b} + r_{\text{new}}^{a,b-1}).$$

This implies,

$$4t \geq l_{\text{nice}} + r_{\text{nice}} + 2l_{\text{new}}^{0,b} + 2r_{\text{new}}^{a,k} + l_{\text{nice}}^{0,a} + r_{\text{nice}}^{b,k}.$$

By the definition of a and b, $l_{\text{new}}^{0,b} = l_{\text{new}} - 1$ and $r_{\text{new}}^{a,k} = r_{\text{new}} - 1$. So, we get

$$4t \geq l_{\text{nice}} + r_{\text{nice}} + 2l_{\text{new}} + 2r_{\text{new}} + l_{\text{nice}}^{0,a} + r_{\text{nice}}^{b,k} - 4$$

$$\geq l_{\text{nice}} + r_{\text{nice}} + 2(l_{\text{new}} + r_{\text{new}}) - 4.$$

Now by Claim 6.5 and inequalities (6.1) and (6.2), we get that

$$4t \geq \frac{4k}{7} + 4 + 2\left(\frac{2k}{7} + 2 + \frac{2k}{7} + 2\right) - 4 = \frac{12k}{7} + 8.$$

Therefore,

$$t \geq \frac{3k}{7} + 2,$$

completing the proof of Lemma 6.7.

6.2.3 Finding a large subset of vertices with few incident edges

Now we define an auxiliary graph H with the vertex set $V(H) = T$ and edge set $E(H)$ such that $ab \in E(H)$ if and only if there is a rainbow path P in G with a and b as its terminal vertices and $V(P) = V(P^*) = \{v_0, v_1, \ldots, v_k\}$.

Claim 6.18. *The degree of every vertex u in H is at least $2k/7 + 2$.*

Proof of Claim. As $u \in V(H) = T$, u is a terminal vertex. So there is a rainbow path $P = u_0u_1 \ldots u_k$ in G such that $u_0 = u$ and $\{u_0, u_1, \ldots, u_k\} = \{v_0, v_1, \ldots, v_k\}$. We define the sets $L, R, L_{\text{new}}, R_{\text{new}}$ corresponding to P in the same way as we did for P^* (in Definition 6.1). Moreover, since P^* was defined as an arbitrary rainbow path of length k, (6.2) holds for P as well – i.e., $|R_{\text{new}}| = r_{\text{new}} \geq 2k/7 + 2$. We claim that if u_ku_j is an edge in G such that $c(u_ku_j) \in R_{\text{new}}$, then $uu_{j+1} \in E(H)$. Indeed, consider the path $u_0u_1 \ldots u_ju_ku_{k-1} \ldots u_{j+1}$. This is clearly a rainbow path with terminal vertices $u = u_0$ and u_{j+1}. So u and u_{j+1} are adjacent in H, as required. This shows that degree of u in H is at least $r_{\text{new}} \geq 2k/7 + 2$, as desired. □

Size of a matching is defined as the number of edges in it. The following proposition is folklore.

Proposition 6.19. *Any graph G with minimum degree $\delta(G)$ has a matching of size*

$$\min\left\{\delta(G), \left\lfloor \frac{V(G)}{2} \right\rfloor\right\}.$$

We know that $\delta(H) \geq \frac{2k}{7} + 2$ by Claim 6.18. Moreover $|V(H)| = |T| = t$. So applying Proposition 6.19 to the graph H and using Lemma 6.7, we obtain that the graph H contains a matching M of size

$$m := \min\left\{\frac{2k}{7} + 2, \left\lfloor \frac{t}{2} \right\rfloor\right\} \geq \frac{3k}{14}. \tag{6.7}$$

Let the edges of M be $a_1b_1, a_2b_2, \ldots, a_mb_m$. Moreover, let

$$n_i = |\{xy \mid xy \notin E(G), x \in \{a_i, b_i\} \text{ and } y \in \{v_0, v_1, v_2, \ldots, v_k\} \setminus \{a_i, b_i\}\}|.$$

Claim 6.20. *The number of edges in the subgraph of G induced by M is*

$$|E(G[M])| \geq \binom{2m}{2} - \left(\sum_{i=1}^{m} \frac{n_i}{2} + m\right) = 2m^2 - 2m - \sum_{i=1}^{m} \frac{n_i}{2}.$$

Proof of Claim. Note that the sum $\sum_i n_i$ counts each pair $xy \notin E(G)$ with $x, y \in V(M)$ exactly twice unless $xy = a_ib_i$ for some i. Therefore, the number of pairs $xy \notin E(G)$ in the subgraph of G induced by M is at most $\sum_i \frac{n_i}{2} + m$. Thus the number of edges of G in the subgraph induced by M is at least $\binom{2m}{2} - (\sum_i \frac{n_i}{2} + m)$, which implies the desired claim. □

Claim 6.21. *The sum of degrees of a_i and b_i in G is at most $3k - \frac{n_i}{2}$.*

Proof of Claim. Since a_ib_i is an edge in the auxiliary graph H, there is a rainbow path $P = u_0u_1 \dots u_k$ in G such that $u_0 = a_i$, $u_k = b_i$ and $\{u_0, u_1, \dots, u_k\} = \{v_0, v_1, \dots, v_k\}$. We define the sets $L, R, L_{\text{in}}, R_{\text{in}}, L_{\text{out}}, R_{\text{out}}, L_{\text{new}}, R_{\text{new}}$ and the numbers $l, r, l_{\text{in}}, r_{\text{in}}, l_{\text{out}}, r_{\text{out}}, l_{\text{new}}, r_{\text{new}}$ corresponding to P in the same way as we did for P^* (in Definition 6.1). Therefore, degree of a_i is $l \leq l_{\text{new}} + k$. Similarly, degree of b_i is at most $r_{\text{new}} + k$. So the sum of degrees of a_i and b_i in G is at most

$$2k + l_{\text{new}} + r_{\text{new}}. \tag{6.8}$$

On the other hand, the sum of degrees of a_i and b_i in G is $l + r = l_{\text{in}} + l_{\text{out}} + r_{\text{in}} + r_{\text{out}}$. By Claim 6.4, this is at most $(l_{\text{in}} + r_{\text{in}}) + k - r_{\text{new}} + k - l_{\text{new}} = (l_{\text{in}} + r_{\text{in}}) + 2k - l_{\text{new}} - r_{\text{new}}$. Moreover, it is easy to see that $l_{\text{in}} + r_{\text{in}} \leq 2k - n_i$ by the definition of n_i. Therefore, the sum of degrees of a_i and b_i in G is at most

$$2k - n_i + 2k - l_{\text{new}} - r_{\text{new}}. \tag{6.9}$$

Adding up (6.8) and (6.9) and dividing by 2, we get that the sum of degrees of a_i and b_i in G is at most

$$\frac{(2k + 2k - n_i + 2k)}{2} = \frac{(6k - n_i)}{2} = 3k - \frac{n_i}{2},$$

as desired. □

The sum $\sum_{i=1}^{m}(d(a_i)+d(b_i))$ counts each edge in the subgraph of G induced by M exactly twice (note that here $d(v)$ denotes the degree of the vertex v in G). Therefore, the number of edges of G incident to the vertices of M is at most $\sum_{i=1}^{m}(d(a_i) + d(b_i)) - |E(G[M])|$. Now using Claim 6.20 and Claim 6.21, the number of edges of G incident to the vertices of M is at most

$$\sum_{i=1}^{m}\left(3k - \frac{n_i}{2}\right) - \left(2m^2 - 2m - \sum_{i=1}^{m}\frac{n_i}{2}\right) = 3km - 2m^2 + 2m = (3k + 2 - 2m)m.$$

Now by (6.7), this is at most

$$(3k + 2 - 2m)m \leq \left(3k + 2 - 2\left(\frac{3k}{14}\right)\right)m = \left(\frac{9k}{7} + 1\right)2m < \left(\frac{9k}{7} + 2\right)2m.$$

We may delete the vertices of M from G to obtain a graph G' on $n - 2m$ vertices. By induction hypothesis, G' contains less than $(\frac{9k}{7} + 2)(n - 2m)$ edges. Therefore, G contains less than

$$\left(\frac{9k}{7} + 2\right)2m + \left(\frac{9k}{7} + 2\right)(n - 2m) = \left(\frac{9k}{7} + 2\right)n$$

edges, as desired. This completes the proof of Theorem 6.2.

Bibliography

[1] P. Allen, P. Keevash, B. Sudakov and J. Verstraëte. "Turán numbers of bipartite graphs plus an odd cycle." *Journal of Combinatorial Theory, Series B* 106 (2014): 134–162. 3.1

[2] N. Alon, A. Pokrovskiy and B. Sudakov. Random subgraphs of properly edge-coloured complete graphs and long rainbow cycles. *Israel Journal of Mathematics* (2017) **222.1**, 317–331. 6.1

[3] N. Alon, T. Jiang, Z. Miller and D. Pritikin, Properly colored subgraphs and rainbow subgraphs in edge-colorings with local constraints, *Random Structures Algorithms* **23** (2003), 409âĂŞ433. 1.4

[4] N. Alon and C. Shikhelman, Many T copies in H-free graphs. *Journal of Combinatorial Theory, Series B* **121** (2016) 146–172. 1.1, 4.1

[5] N. Alon, L. Rónyai, and T. Szabó: Norm-graphs: variations and applications, J. Combin. Theory Ser. B 76 (1999), 280-290. 1

[6] C. T. Benson. "Minimal regular graphs of girths eight and twelve." *Canad. J. Math.* 18 (1966), 1091–1094. 3.1

[7] R. G. Blakley and P. Roy. "A Hölder type inequality for symmetric matrices with nonnegative entries." *Proceedings of the American Mathematical Society* 16.6 (1965): 1244-1245. 3.1, 3.3.2

[8] B. Bollobás and E. Győri, Pentagons vs. triangles. *Discrete Mathematics* **308.19** (2008) 4332–4336. 1.1, 1.6, 1.2, 4.1

[9] B. Bollobás and E. Győri. "Pentagons vs. triangles." *Discrete Mathematics*, 308 (19) (2008), 4332–4336. 3.1, 3.1

[10] A. Bondy and M. Simonovits. "Cycles of even length in graphs." *Journal of Combinatorial Theory, Series B* 16.2 (1974): 97–105. 1.4, 3.1

[11] R.C. Bose and S. Chowla. Theorems in the additive theory of numbers. *Comment. Math. Helv.* (1962/1963) **37**, 141-147. 1.4

[12] W.G. Brown, P. Erdős and V. Sós. "On the existence of triangulated spheres in 3-graphs and related problems." *Periodica Mathematica Hungaria* 3 (1973), 221–228.

[13] B. Bukh and Z. Jiang. "A bound on the number of edges in graphs without an even cycle." *Combinatorics, Probability and Computing* (2014): 1–15. 3.1

[14] C. Collier-Cartaino, N. Graber and T. Jiang. "Linear Turán numbers of r-uniform linear cycles and related Ramsey numbers." *Combinatorics, Probability and Computing* 27.3 (2018): 358–386. (document), 3.1

[15] S. Das, C. Lee and B. Sudakov. Rainbow Turán problem for even cycles. *European Journal of Combinatorics* (2013) **34**, 905–915. 1.4

[16] P. Erdős, On some problems in graph theory, combinatorial analysis and combinatorial number theory. B. Bollobás (Ed.), Graph Theory and Combinatorics (Cambridge, 1983), Academic Press, London (1984) 1–17. 4.1

[17] P. Erdős, Problems and results in combinatorial analysis, Proceedings of the Eighth Southeastern Conference on Combinatorics, Graph Theory and Computing, Louisana State University, Baton Rouge, LA, 1977, *Congressus Numerantium*, Vol XIX, Utilitas Mathematics, Winnipeg, Manchester, 1977, 3–12. 1.2

[18] P. Erdős. "Some recent progress on extremal problems in graph theory." *Congr. Numer.* 14 (1975), 3–14. 3.1

[19] P. Erdős and M. Simonovits. "Compactness results in extremal graph theory." *Combinatorica* 2 (1982), no. 3, 275–288. 3.1, 4.1, 4.2.1, 4.4

[20] P. Erdős and T. Gallai. On maximal paths and circuits of graphs. *Acta Mathematica Academiae Scientiarum Hungaricae* (1959) **10**, 337–356. 6.1

[21] P. Erdős and M. Simonovits. A limit theorem in graph theory. *Studia Scientiarum Mathematicarum Hungarica* 1 (1965), 51–57. 1, 1.2

[22] P. Erdős and R. Rado, A combinatorial theorem, *J. London Math.* Soc.25(1950), 249-255. 1.4

[23] P. Erdős, A. H. Stone. On the structure of linear graphs. *Bulletin of the American Mathematical Society* 52 (1946), 1087–1091. 1, 1.2

[24] P. Erdős, On some problems in graph theory, combinatorical analysis and combinatorial number theory, in: B. Bollobás (Ed.), *Graph Theory Combin.* pp. 1-17. 1.1

[25] P. Erdős and D. Kleitman, On coloring graphs to maximize the proportion of multicolored k-edges. *J. Combinatorial Theory* **5** (1968) 164–169. 5.2

[26] B. Ergemlidze, A. Methuku. "Triangles in C_5-free graphs and hypergraphs of girth six." arXiv preprint arXiv:1811.11873 (2019). 1.1, 4.2, 4.3, 4.4

[27] B. Ergemlidze, T. Jiang, A. Methuku. "New bounds for a hypergraph Bipartite Turán problem." arXiv preprint arXiv:1902.10258 (2019) 1.2, 5.1

[28] B. Ergemlidze, E. Győri, A. Methuku. "3-Uniform Hypergraphs and Linear Cycles" *SIAM Journal on Discrete Mathematics*, 32(2), 933âĂŞ950. (2018) 1.3, 2.2, 2.4

[29] B. Ergemlidze, E. Győri, A. Methuku. "3-uniform hypergraphs without a cycle of length five." arXiv preprint arXiv:1902.06257 (2019) 1.2, 3.2

[30] B. Ergemlidze, E. Győri, A. Methuku. "Asymptotics for Turán numbers of cycles in 3-uniform linear hypergraphs." *Journal of Combinatorial Theory, Series A* 163 (2019): 163-181. 1.2, 3.3, 3.4, 3.5

[31] B. Ergemlidze, E. Győri and A. Methuku. "Turán number of an induced complete bipartite graph plus an odd cycle." *Combinatorics, Probability and Computing*, (2018) 1–12. 4.1

[32] B. Ergemlidze, E. Győri, and A. Methuku "A note on the Linear Cycle Cover Conjecture of Gyárfás and Sárközy" *The Electronic Journal of Combinatorics,* 25.2, (2018): Paper P2.29. 1.3, 2.5

[33] B. Ergemlidze, E. Győri, A. Methuku and N. Salia. "A note on the maximum number of triangles in a C_5-free graph." *Journal of Graph Theory*, (2018) 1–4, In print. 1.1, 4.1, 4.1, 4.2

[34] B. Ergemlidze, E. Győri and A. Methuku. On the Rainbow Turán number of paths. the electronic journal of combinatorics 26(1), P1.17, (2019). 1.4, 6.2

[35] B. Ergemlidze, E. Győri, A. Methuku, C. Tompkins, N. Salia, O. Zamora. Avoiding long Berge cycles, the missing cases $k = r+1$ and $k = r+2$. arXiv preprint arXiv:1808.07687. 1.2

[36] Z. Füredi, Hypergraphs in which all disjoint pairs have distinct unions. *Combinatorica* 4(2-3) (1984) 161–168. 1.2, 1.12

[37] Z. Füredi, T. Jiang, D. Mubayi, A. Kostochka, J. Verstraëte, The extremal number for (a, b)-paths and other hypergraph trees, manuscript, 19 pp., April 12, 2018. 5.3

[38] Z. Füredi and L. Özkahya. On 3-uniform hypergraphs without a cycle of a given length. *Discrete Applied Mathematics*, **216**, (2017) 582–588. 4.1, 5.3

[39] Z. Füredi, A. Kostochka, R. Luo. Avoiding long Berge cycles. *arXiv preprint* arXiv:1805.04195, (2018). 1.2, 1.1, 1.2

[40] Z. Füredi, and L. Özkahya. "On 3-uniform hypergraphs without a cycle of a given length." *Discrete Applied Mathematics*, 216 (2017): 582–588. 3.1

[41] D. Gerbner and C. Palmer. "Extremal results for Berge-hypergraphs." *SIAM Journal on Discrete Mathematics*, 31.4 (2017): 2314–2327.

[42] D. Gerbner, A. Methuku and M. Vizer. "Asymptotics for the Turán number of Berge-$K_{2,t}$." arXiv preprint arXiv:1705.04134 (2017).

[43] A. Grzesik. On the maximum number of five-cycles in a triangle-free graph. *Journal of Combinatorial Theory, (Series B)*, 102(5) (2012), 1061–1066. 1.1

[44] E. Győri, N. Lemons. Hypergraphs with no cycle of a given length. *Combinatorics, Probability and Computing*, **21**(1-2), 193–201, 2012. 1.10

[45] E. Győri and H. Li, The maximum number of triangles in C2k+1-free graphs, textit-Combinatorics, Probability and Computing **21** (1-2), 187âĂŞ191, 2012. 1.1, 1.7

[46] E. Győri and N. Lemons. "3-uniform hypergraphs avoiding a given odd cycle." *Combinatorica* 32.2 (2012): 187–203. 1.10, 3.1

[47] E. Győri. Triangle-Free Hypergraphs. *Combinatorics, Probability and Computing*, 15 (1-2) (2006),185–191 . doi:10.1017/S0963548305007108. 1.2, 3.1

[48] A. Gyárfás, E. Győri and M. Simonovits. "On 3-uniform hypergraphs without linear cycles." Journal of Combinatorics 7.1 (2016): 205–216. 1.2, 1.3, 2.1, 2.1, 2.1, 2.3

[49] A. Gyárfás and G. Sárközy "Monochromatic loose-cycle partitions in hypergraphs." The Electronic Journal of Combinatorics 21.2 (2014): 2–36. 1.3, 1.14, 1.3, 2.1

[50] P. Hall, On Representatives of Subsets. *J. London Math. Soc* **10**, (1935) 26–30. 5.2

[51] H. Hatami, J. Hladký, D. Král, S. Norine, A. Razborov. On the number of pentagons in triangle-free graphs. *Journal of Combinatorial Theory, (Series A)*, 120 (3) (2013), 722–732. 1.1

[52] T. Jiang, X. Liu, Turán numbers of enlarged cycles, manuscript 5.3

[53] D. Johnston, C. Palmer and A. Sarkar. Rainbow Turán Problems for Paths and Forests of Stars. *The Electronic Journal of Combinatorics* (2017) **24(1)**, 1–34. (document), 1.4, 6.1, 6.1

[54] P. Keevash, Hypergraph Turán problems, Surveys in Combinatorics, Cambridge University Press, 2011, 83–140. 5.1

[55] P. Keevash, D. Mubayi, B. Sudakov and J. Verstraëte. Rainbow Turán problems. *Combinatorics, Probability and Computing* (2007) **16**, 109–126. 1.4, 6.1, 6.1

[56] J. Kollár, L. Rónyai, T. Szabó. Norm-graphs and Bipartite Turán numbers. *Combinatorica*16 (1996), 399–406. 1

[57] A. Kostochka, R. Luo, On r-uniform hypergraphs with circumference less than r, arXiv preprint arXiv:1807.04683 (2018). 1.2

[58] T. Kővári, V. Sós, P. Turán. On a problem of K. Zarankiewicz. In *Colloquium Mathematicae*, 3(1), (1954), 50–57. 1.3

[59] F. Lazebnik and J. Verstraëte. On hypergraphs of girth five. *Electron. J. Combin* 10 (2003) R25. 3.1, 3.1, 3.1, 4.1

[60] F. Lazebnik, V. A. Ustimenko and A. J. Woldar. "A new series of dense graphs of high girth." *Bull. Amer. Math. Soc.* 32 (1995), no. 1, 73–79. 3.1

[61] P. Loh, M. Tait, C. Timmons and R.M. Zhou. Induced Turán numbers. *Combinatorics, Probability and Computing*, **27(2)** (2017) 274–288. 4.1

[62] M. Maamoun and H. Meyniel. On a problem of G. Hahn about coloured Hamiltonian paths in K_{2^t}. *Discrete Mathematics* (1984) **51**, 213–214. 6.1

[63] W. Mantel. Problem 28. *Wiskundige Opgaven* 10 (1907), 60–61. 1

[64] D. Mubayi and J. Verstraëte, A hypergraph extension of the bipartite Turán problem. *J. Combinatorial Theory, Series A* 106.2 (2004): 237–253. 1.2, 5.1, 5.2.3, 5.2.5, 5.3, 5.3, 5.3

[65] O. Pikhurko, and J. Verstraëte, "The maximum size of hypergraphs without generalized 4-cycles." *J. Combinatorial Theory, Series A* 116.3 (2009): 637-649. 1.2

[66] O. Pikhurko. "A note on the Turán function of even cycles." *Proceedings of the American Mathematical Society* 140.11 (2012): 3687–3692. 3.1

[67] L. Pósa "On the circuits of finite graphs." Magyar Tud. Akad. Mat. Kutató Int. Közl 8 (1963): 355–361. 1.3, 2.1

[68] I. Ruzsa and E. Szemerédi. "Triple systems with no six points carrying three triangles." in *Combinatorics, Keszthely, Colloq. Math. Soc. J. Bolyai* 18, Vol II (1976): 939–945. 1.9, 3.1

[69] R. R. Singleton. "On minimal graphs of maximum even girth." *J. Combinatorial Theory* 1 (1966), 306–332. 3.1

[70] C. Timmons. "On r-uniform linear hypergraphs with no Berge-$K_{2,t}$." arXiv preprint arXiv:1609.03401 (2016).

[71] P. Turán. On an extremal problem in graph theory. *Matematikai ès Fizikai Lapok (in Hungarian),* 48 (1941), 436–452. 1.1

[72] J. Verstraëte. "On arithmetic progressions of cycle lengths in graphs." *Combinatorics, Probability and Computing* 9.04 (2000): 369–373. 3.1