



8-2018

# Probabilistic and Extremal Problems in Combinatorics

Sean English

Western Michigan University, sean.j.english@wmich.edu

Follow this and additional works at: <https://scholarworks.wmich.edu/dissertations>



Part of the [Mathematics Commons](#)

## Recommended Citation

English, Sean, "Probabilistic and Extremal Problems in Combinatorics" (2018). *Dissertations*. 3319.  
<https://scholarworks.wmich.edu/dissertations/3319>

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact [maira.bundza@wmich.edu](mailto:maira.bundza@wmich.edu).



# Probabilistic and Extremal Problems in Combinatorics

by

Sean English

A dissertation submitted to the Graduate College  
in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy  
Mathematics  
Western Michigan University  
August 2018

Doctoral Committee:

Andrzej Dudek, Ph.D., Chair  
Patrick Bennett, Ph.D.  
Alan Frieze, Ph.D.  
Allen Schwenk, Ph.D.

Copyright by  
Sean English  
2018

# Probabilistic and Extremal Problems in Combinatorics

Sean English, Ph.D.

Western Michigan University, 2018

Graph theory as a mathematical branch has been studied rigorously for almost three centuries. In the past century, many new branches of graph theory have been proposed. One important branch of graph theory involves the study of extremal graph theory. In 1941, Turán studied one of the first extremal problems, namely trying to maximize the number of edges over all graphs which avoid having certain structures. Since then, a large body of work has been created in the study of similar problems. In this dissertation, a few different extremal problems are studied, but for hypergraphs rather than graphs. In particular we consider the saturation problem for families of Berge hypergraphs.

In addition to the study of extremal hypergraph problems, this dissertation also focuses on problems of a probabilistic nature. Graphs and hypergraphs that are generated through a random process have been popular objects of study since the seminal work on them done by Erdős and Rényi, and independently Gilbert in 1959. Here we study problems involving finding certain Ramsey properties of random hypergraphs, and also finding the probabilistic threshold for specific cycle structures to appear in randomly colored random hypergraphs. We also explore an application of Markov chains to a problem of a topological nature, namely studying Morse functions on the sphere.

# Acknowledgements

This dissertation would not have been possible without the support of many people. Here I will mention only a few of the most important people.

I would like to thank my parents, John and Delores English, for always believing in me and for helping me through every step of my educational journey. Also, I want to thank my siblings, Andrew and Brianna English, for always being there when I needed a friend or favor.

Now, I would like to thank the members of my committee for all of their support. First, Dr. Allen Schwenk, who taught me the basic of mathematical proofs and also the basics of graph theory. Next, Dr. Patrick Bennett, who was a great coauthor to work with and who taught me many advanced topics in combinatorics. Also, Dr. Alan Frieze, who taught me that probability theory is probably not that hard, as long as I get good at estimating. Finally, my advisor, Dr. Andrzej Dudek, for guiding me through the PhD process and helping me develop as a mathematician.

Sean English

# Contents

<b>Acknowledgments</b>	<b>ii</b>
<b>List of Figures</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Large Monochromatic Components and Cycles in Random Hypergraphs</b>	<b>9</b>
2.1 Introduction . . . . .	9
2.1.1 Large Components . . . . .	10
2.1.2 Loose Cycles . . . . .	12
2.2 Notation and Definitions . . . . .	13
2.3 Random Hypergraphs . . . . .	14
2.3.1 Sparse Weak Hypergraph Regularity . . . . .	14
2.3.2 Large Components . . . . .	15
2.3.3 Loose Cycles . . . . .	18
2.4 $k$ -Uniform Hypergraphs Colored with $k$ Colors . . . . .	25
2.5 $k$ -Uniform Hypergraphs Colored with $k + 1$ Colors . . . . .	26
2.6 Technical Lemmas . . . . .	30
2.7 Concluding Remarks . . . . .	35
<b>3 Rainbow Hamilton Cycles in Random Hypergraphs</b>	<b>37</b>
3.1 Introduction . . . . .	37
3.2 $\ell$ -Hamilton Cycles for $\ell \geq 2$ . . . . .	41
3.3 Loose Hamilton Cycles . . . . .	47
3.4 Concluding Remarks . . . . .	49

<b>4</b>	<b>Saturation of Berge Hypergraphs</b>	<b>50</b>
4.1	Introduction . . . . .	50
4.2	Berge Saturation Numbers for Paths . . . . .	54
4.2.1	A Lower Bound for $\text{sat}_k(n, \text{Berge-}P_m)$ . . . . .	54
4.3	Constructing an Upper Bound . . . . .	57
4.3.1	Constructing the Tree $T_m^{(k)}$ . . . . .	58
4.3.2	The Saturation of $T_m^{(k)}$ . . . . .	62
4.3.3	Using $T_m^{(k)}$ to Construct the Saturated Hypergraph $H_{n,m}^{(k)}$ . . . . .	67
4.4	Lower Bound . . . . .	70
4.4.1	Branching Lemmas . . . . .	72
4.4.2	Counting Lemmas . . . . .	76
4.4.3	Central Structure and Final Count . . . . .	81
4.5	Berge Saturation Numbers for $K_3$ , $C_m$ , $K_{1,m}$ , and $\ell K_2$ . . . . .	95
4.6	Concluding Remarks . . . . .	103
<b>5</b>	<b>A Random Variant of the Game of Plates and Olives</b>	<b>105</b>
5.1	Introduction . . . . .	105
5.2	Bounds on the Expected Number of Olives . . . . .	109
5.2.1	Lower Bound . . . . .	109
5.2.2	Upper Bound . . . . .	111
5.3	Concentration . . . . .	112
5.4	A Related Markov Chain . . . . .	117
5.5	Concluding Remarks . . . . .	121
<b>A</b>	<b>Some Tools and Inequalities</b>	<b>122</b>
	<b>Bibliography</b>	<b>124</b>

# List of Figures

2.1	A 4-uniform loose cycle and a 5-uniform diamond. . . . .	12
2.2	A Berge path on four edges. . . . .	21
3.1	A 2-Hamilton and a 3-Hamilton 5-uniform cycle. . . . .	37
4.1	A hypergraph $H$ that contains a Berge- $P_5$ . . . . .	52
4.2	The saturated tree $T_8^{(3)}$ . . . . .	58
4.3	The saturated tree $T_{11}^{(3)}$ . . . . .	59
4.4	The saturated tree $T_{18}^{(4)}$ . . . . .	60
4.5	The saturated tree $T_8^{(k)}$ . . . . .	62
4.6	The saturated tree $T_{13}^{(k)}$ . . . . .	62
4.7	Two cases in the proof of Lemma 4.3.7. . . . .	65
4.8	The hypergraph $T'$ in Construction 4.3.8. . . . .	68
4.9	The graph $H + e$ as described in Lemma 4.4.4. . . . .	73
4.10	The edges $e_1$ , $e_2$ and $e_3$ as described in Lemma 4.4.10. . . . .	79
4.11	Edges $e_1$ and $e_2$ when $m = 2\ell + 1$ in situation (i) of Lemma 4.4.12. . . . .	82
4.12	Edges $e_1$ and $e_2$ when $m = 2\ell + 1$ in situation (ii) of Lemma 4.4.12. . . . .	82
4.13	Edges $e_1$ - $e_4$ when $m = 2\ell$ in situation (ii) of Theorem 4.4.15. . . . .	92
4.14	Edges $e_1$ - $e_4$ when $m = 2\ell$ in situation (iii) of Theorem 4.4.15. . . . .	92
4.15	Edges $e_1$ - $e_5$ when $m = 2\ell + 1$ in situation (v) of Theorem 4.4.15. . . . .	94
4.16	The Berge- $K_3$ saturated graph, $S_{12}^{(4)}$ . . . . .	96
4.17	Minimal Berge- $K_3$ saturated hypergraphs on 8 vertices with uniformity 3. . . . .	99



# Chapter 1

## Introduction

A  $k$ -uniform hypergraph  $H = (V, E)$  is a collection of vertices  $V = V(H)$  along with a set of edges, or  $k$ -tuples of vertices in  $V$ , i.e.  $E = E(H) \subseteq \binom{V}{k}$ . When  $k = 2$ , we have the familiar object of a simple graph. This dissertation is comprised of four chapters of content, each which explores a distinct body of problems involving or closely related to the study of uniform hypergraphs or graphs.

Extremal graph theory generally involves the study of finding structures that maximize or minimize a certain property over a restricted set of graphs. Possibly the first example of a result in extremal graph theory is due to Mantel.

**Mantel's Theorem** (1907, [62]) *Let  $G$  be a graph on  $n$  vertices with no triangles. Then  $|E(G)| \leq \lfloor n^2/4 \rfloor$ .*

Furthermore, the only graph that attains the maximum in Mantel's theorem is the balanced complete bipartite graph,  $K_{\lfloor n/2 \rfloor, \lfloor n/2 \rfloor}$ . Extremal graph theory did not begin to become popular though until Turán generalized Mantel's Theorem in 1941. Let the *Turán graph*, denoted  $T(n, r)$ , be the balanced (any two partite sets differ by at most one vertex) complete  $(r - 1)$ -partite graph on  $n$  vertices.

**Turán's Theorem** (1941, [77]) *Let  $G$  be a graph on  $n$  vertices that contains no copy of  $K_r$  as a subgraph. Then  $|E(G)| \leq |E(T(n, r))|$ .*

Note that the case of  $r = 3$  coincides with Mantel's Theorem. The work of Turán inspired a lot of attention to similar problems, all trying to maximize the number of edges in

a graph that avoids containing a certain subgraph. Given a forbidden graph  $F$ , the *extremal number* of a graph,  $\text{ex}(n, F)$ , is the maximum number of edges over all graphs on  $n$  vertices that do not have a subgraph isomorphic to the forbidden graph  $F$ , or  $F$ -free graphs. Thus, Turán's Theorem can be rephrased succinctly as follows:

**Turán's Theorem (v2)** (1941, [77])

$$\text{ex}(n, K_r) = |E(T(n, r))|.$$

Extremal numbers have been extensively studied for various classes of graphs. One of the most fundamental results on extremal numbers, and even the whole branch of extremal graph theory relates the growth rate of extremal numbers to the *chromatic number* of the forbidden graph  $F$ ,  $\chi(F)$ , which is the least number of colors necessary to color the vertices of  $F$  such that all adjacent vertices receive distinct colors.

**Erdős-Stone Theorem** (1946, [26])

$$\text{ex}(n, F) = \left( \frac{\chi(F) - 2}{\chi(F) - 1} + o(1) \right) \binom{n}{2}.$$

While the study of extremal numbers formed the basis for the study of extremal graph theory, there are many other interesting topics one can study. Saturation numbers and Ramsey-type problems are well-studied topics, and hypergraph analogues of both these problems will appear extensively in this dissertation, the former in Chapter 4 and the latter in Chapter 2.

We say a graph  $G$  is  *$F$ -saturated* if  $G$  is  $F$ -free and for any edge  $e \notin E(G)$ ,  $G + e$  contains a copy of  $F$ . Any  $F$ -free graph with  $\text{ex}(n, F)$  edges must necessarily be  $F$ -saturated, so the problem of finding extremal numbers can be rephrased as looking for the maximum number of edges in an  $F$ -saturated graph on  $n$  vertices. This phrasing yields an

interesting minimization problem that complements the maximization for extremal numbers. The *saturation number* of a forbidden graph  $F$ , denoted  $\text{sat}(n, F)$  is the least number of edges in an  $F$ -saturated graph on  $n$  vertices. It is worth noting that since saturation numbers are a minimum over the set of saturated graphs and extremal numbers are a maximum over the same set, we always have  $\text{sat}(n, F) \leq \text{ex}(n, F)$ .

Saturation problems have been well-studied. These numbers have been looked at for many common structures including complete graphs, cycles, paths, matchings and stars. One of the foundational theorems in the study of saturation numbers bounds the growth rate of these numbers. If  $f(n)$  and  $g(n)$  are functions such that  $\lim_{n \rightarrow \infty} f(n)/g(n) < \infty$ , then we say  $f(n) = O(g(n))$ .

**Theorem** (Kászonyi and Tuza, 1986, [56]) *For all graphs  $F$ ,*

$$\text{sat}(n, F) = O(n).$$

This is significant, especially when compared to the Erdős-Stone Theorem, which gives that for non-bipartite graphs, extremal numbers grow quadratically in the number of vertices.

The saturation problem has also been extended to  $k$ -uniform hypergraphs. When  $F$  is a  $k$ -uniform forbidden graph, we denote by  $\text{sat}_k(n, F)$ , the  *$k$ -uniform saturation number*, which is the least number of edges in an  $F$ -saturated  $k$ -uniform hypergraph on  $n$  vertices. As with most graph theoretic concepts, hypergraph saturation problems tend to be a great deal harder than their graph counterparts, and so much less is known about them in general. The result by Kászonyi and Tuza has been generalized though.

**Theorem** (Pikhurko, 1999, [68]) *For any  $k$ -uniform hypergraph  $F$ ,*

$$\text{sat}_k(n, F) = O(n^{k-1}).$$

The work in Chapter 4 studies the  $k$ -uniform saturation numbers for particular families of hypergraphs related to graph structures, known as Berge hypergraphs.

Given a  $k$ -uniform hypergraph  $H$  and a graph  $F$ , we say  $H$  is *Berge- $F$*  if  $|E(H)| = |E(F)|$ , and there is a way to remove each hypergraph edge of  $H$  and replace it by a single graph edge contained within the removed hyperedge, such that this process results with a graph isomorphic to  $F$ . For example, it can be seen that the complete 3-uniform hypergraph on 4 vertices,  $K_4^{(3)}$  is a Berge- $C_4$ , but does not contain a Berge- $K_4$  as  $|E(K_4^{(3)})| < |E(K_4)|$ .

If we do not allow for  $H$  to have isolated vertices, there are always finitely many  $k$ -uniform hypergraphs that are Berge- $F$  for any  $F$ . In this way, we say that a hypergraph is Berge- $F$ -saturated if it does not contain any subgraph that is Berge- $F$ , but adding any hyperedge would create a Berge- $F$ . With this, we can define the saturation number for Berge- $F$ ,  $\text{sat}_k(n, \text{Berge-}F)$ . In Chapter 4, we will study these numbers for many common classes of graphs.

The other main extremal problem that will be discussed in this dissertation involves Ramsey-type problems. Ramsey theory is a broad area study that has non-trivial intersection with almost all branches of combinatorics, but here we will only be concerned with the relation of Ramsey theory to graphs and hypergraphs. A  *$k$ -edge-coloring* of a graph  $G$  is an assignment of labels, or colors, from the set  $[k]$  to the edges of  $G$ . Given a  $k$ -edge coloring, we say a subgraph is *monochromatic* if each edge in the subgraph has been assigned the same color. One general version of Ramsey's Theorem for graphs is as follows:

**Ramsey's Theorem** (1930, [70]) *Given any integer  $k \geq 1$ , and any graphs  $F_1, \dots, F_k$ , there exists sufficiently large  $n$  such that in any  $k$ -coloring of the edges of the complete graph  $K_n$ , there is an  $i \in [k]$  such that there is a copy of  $F_i$ , all whose edges are colored  $i$ .*

The main problem in Ramsey theory involves investigating the smallest  $n$  such that the conditions of Ramsey's Theorem hold. We define the *Ramsey number*  $R(F_1, \dots, F_k)$  to

be this smallest value of  $n$ , and if  $F_i = F$  for all  $i \in [k]$ , we write the *diagonal Ramsey number*,  $R(F; k) = R(F_1, \dots, F_k)$ .

Consider the family of diagonal Ramsey numbers,  $R(K_\ell; k)$ . In order to understand this family, it can be useful to explore a related problem: Given fixed integers  $n, k \geq 1$ , what is the largest integer  $\ell$  such that every  $k$ -edge coloring of  $K_n$  contains a monochromatic  $K_\ell$ ? In essence, this related problem studies the inverse of the Ramsey numbers  $R(K_\ell, k)$ , when thought of as a function of  $\ell$  ( $k$  is fixed).

In Chapter 2, we study problems related to this inverse Ramsey problem, but extended to hypergraphs. Namely, we study problems involving finding the size of the largest monochromatic component or longest monochromatic cycle guaranteed over all  $k$ -edge colorings of a host hypergraph. For our purposes, the host hypergraph we are concerned with is the random  $k$ -uniform hypergraph  $\mathcal{H}^{(k)}(n, p)$ .

The *random  $k$ -uniform hypergraph*  $\mathcal{H}^{(k)}(n, p)$  is a random hypergraph on  $n$  vertices such that each of the  $\binom{n}{k}$  edges are included independently with probability  $p$ . Given a hypergraph property  $P$ , if the probability that a random sample of  $\mathcal{H}^{(k)}(n, p)$  has property  $P$  goes to 1 as  $n$  goes to infinity, then we say  $\mathcal{H}^{(k)}(n, p)$  has property  $P$  *asymptotically almost surely*, or a.a.s. for short. In this way, even though  $\mathcal{H}^{(k)}(n, p)$  is technically a probability space, we often colloquially speak of it as if it is a single graph (when  $p$  is some fixed function of  $n$ ) with certain properties.

The study of random graphs and hypergraphs has been a booming area of mathematics since 1959 when Paul Erdős and Alfred Rényi, and independently Edgar Gilbert, proposed closely related models of random graphs,  $\mathcal{G}(n, p)$ , which is the 2-uniform version of  $\mathcal{H}^{(k)}(n, p)$ , and  $\mathcal{G}(n, M)$ , which uniformly at random chooses a graph on  $n$  vertices with exactly  $M$  edges. Various other models of random graphs have been studied since then, including, but not limited to, models which produce regular graphs, models based on the geometry of some underlying space, and models that involve adding new vertices in a prescribed

manner.

In this dissertation, we will only consider the model  $\mathcal{H}^{(k)}(n, p)$  in Chapter 2, and a randomly colored version of  $\mathcal{H}^{(k)}(n, p)$ , denoted  $\mathcal{H}^{(k)}(n, p, \kappa)$  in Chapter 3, where each edge that is included in our model is colored with a color from  $[\kappa]$ , uniformly at random, and chosen independently of all other colors. The analysis of random graphs and hypergraphs can give insight into “generic” or “common” graphs. While studying random graphs usually does not yield results that work for all graphs, it can lead to much more powerful results about “typical” graphs or hypergraphs.

One theorem that we will find useful in our study of random graph theory is non-probabilistic in nature. Given a graph  $G = (V, E)$  and disjoint subsets  $X, Y \subseteq V$ , the *density*  $d(X, Y)$  of the pair is defined to be the ratio of the number of edges  $e \in E$  with one endpoint in  $X$  and one endpoint in  $Y$ , and  $|X| \cdot |Y|$ . A partition of the vertex set  $V = V_1 \cup \dots \cup V_k$  is called an  $\epsilon$ -regular partition (for some fixed  $\epsilon > 0$ ) if the sets in the partition differ in size by at most one vertex, and all except for at most  $\epsilon k^2$  of the pairs  $(V_i, V_j)$  has the property that for any  $A \subset V_i$  and  $B \subset V_j$  with  $|A| \geq \epsilon|V_i|$  and  $|B| \geq \epsilon|V_j|$ , we have that  $|d(A, B) - d(V_i, V_j)| \leq \epsilon$ . In other words, as long as  $A$  and  $B$  are not too small, the density of edges between them is about the same as the density of edges between  $V_i$  and  $V_j$ . Szemerédi’s regularity lemma says something very strong about all large graphs.

**Szemerédi’s Regularity Lemma** (1978, [74]) *For every  $\epsilon > 0$  and positive integer  $m$ , there exists an integer  $M \geq m$  such that every graph  $G = (V, E)$  of order at least  $m$  admits an  $\epsilon$ -regular partition  $V = V_1 \cup \dots \cup V_k$  with  $m \leq k \leq M$ .*

Much work has been done to strengthen the regularity lemma. When the number of vertices in  $G$  is very large compared to  $M$ , it has been shown that under certain conditions, the edges between an  $\epsilon$ -regular pair exhibit very nice structural properties. These structural properties makes finding certain structures tractable. The regularity lemma does not say much about  $\epsilon$ -regular pairs whose density is much lower than the chosen  $\epsilon$ . This can be

problematic if the graph in question has both large dense areas and large sparse areas. The random graph  $\mathcal{G}(n, p)$  avoids this issue as the density between large subsets of vertices is concentrated around its mean. This makes the regularity lemma particularly useful for studying random graphs and hypergraphs.

The regularity lemma has been expanded and generalized in many ways. Of particular interest for this dissertation is the extension of the regularity lemma to hypergraphs ([13], [72], [75], [44]). Hypergraph regularity has been useful for solving many problems both deterministic and probabilistic, and has been useful in the study of  $\mathcal{H}^{(k)}(n, p)$ . Another important generalization involves sparse regularity, which allows for the study of graphs with non-constant density, and in particular random graphs with probability  $p = o(1)$ . In Chapter 2, we will use the so-called “sparse” hypergraph regularity lemma to study extremal properties of the random graph  $\mathcal{H}^{(k)}(n, p)$ .

A graph property  $P$  is called monotone increasing (decreasing) if it is closed under adding (removing) edges. When considering the random graph  $\mathcal{G}(n, p)$ , if we increase the probability  $p$ , we can view the change in the random graph as similar to adding more edges to the graph. In this sense, one might expect monotone properties to work well with random graphs. We say that the probability  $p^* = p^*(n)$  is a threshold for a monotone increasing property  $P$  if  $\mathcal{G}(n, p)$  has property  $P$  a.a.s. if  $p \gg p^*$  and  $\mathcal{G}(n, p)$  does not have property  $P$  a.a.s. if  $p \ll p^*$ . Thresholds for monotone decreasing properties are defined analogously.

**Theorem** (Bollobás and Thomason, 1985, [11]) *Every monotone property has a threshold.*

Many of the most interesting monotone properties that one can study involve the existence of certain structures. We call a cycle that spans the entire graph a Hamilton cycle. Hamilton cycles have been studied essentially since the advent of graph theory, and are one of the most common and popular structures to look at in graphs. When studying hypergraphs, there is more than one natural way to define a Hamilton cycle. In Chapter

3, we will study thresholds for the existence of many common types of rainbow Hamilton cycles in the randomly colored random graph  $\mathcal{H}^{(k)}(n, p, \kappa)$ .

Random graphs and random hypergraphs are very interesting objects that have been highly studied recently, but they are not the only object of study in probabilistic graph theory. Also very popular is the study of random walks on graphs. For our purposes, we will consider random walks on directed graphs, or graphs where each edge has a direction. A probability assignment to the edges of a directed graph is an assignment of weights (probabilities) between 0 and 1 to each edge in such a way that the sum of the weights of the outgoing edges at every vertex adds up to 1. If we have a directed graph with a probability assignment, a random walk on this directed graph that respects the probabilities of the edges is called a Markov chain.

Markov chains have been common objects of study for over a century. In Chapter 5, we will study a randomized version of the game of plates and olives, which is a combinatorial process designed to model certain problems in Morse theory. The process of randomly playing the game of plates and olives gives rise to an infinite state Markov chain, or a Markov chain on a directed graph that has infinitely many vertices. A natural question to study in infinite state Markov chains involves the long-run behavior of the Markov chain. In Chapter 5, we will study the long-run behavior of the number of olives at time  $t$  in the randomized game of plates and olives.



# Chapter 2

## Large Monochromatic Components and Cycles in Random Hypergraphs

### 2.1 Introduction

The results of this chapter are joint work with Patrick Bennett, Louis DeBiasio, and Andrzej Dudek [8].

It is known, due to Gyárfás [45], that for any  $r$ -coloring of the edges of  $K_n$ , there is a monochromatic component of order at least  $n/(r-1)$  and this is tight when  $r-1$  is a prime power and  $n$  is divisible by  $(r-1)^2$ . Later, Füredi [35] introduced a more general method which implies the result of Gyárfás. More recently, Mubayi [63] and independently, Liu, Morris, and Prince [61], gave a simple proof of a stronger result which says that in any  $r$ -coloring of the edges of  $K_n$ , there is either a monochromatic component on  $n$  vertices or a monochromatic double star of order at least  $n/(r-1)$ . Recently, Bal and DeBiasio [4] and independently, Dudek and Prałat [23], showed that the Erdős-Rényi random graph behaves very similarly with respect to the size of the largest monochromatic component.

Recall that the *random graph*  $\mathcal{G}(n, p)$  is the random graph  $G$  with vertex set  $[n]$  in which every pair  $\{i, j\} \in \binom{[n]}{2}$  appears independently as an edge in  $G$  with probability  $p$ . An event in a probability space holds *asymptotically almost surely* (or *a.a.s.*) if the probability that it holds tends to 1 as  $n$  goes to infinity. More precisely, it was shown in [4] and [23] that that a.a.s. for any  $r$ -coloring of the edges of  $\mathcal{G}(n, p)$ , there is a monochromatic component of

order  $(\frac{1}{r-1} - o(1))n$ , provided that  $pn \rightarrow \infty$  (that means the average degree tends to infinity). As before, this result is clearly best possible.

In this chapter, we study a generalization of these results to  $k$ -uniform hypergraphs (each edge has order  $k$ ). As in the  $k = 2$  case, our results hold even for very sparse random hypergraphs; that is, we only assume that the average degree,  $pn^{k-1}$ , tends to infinity together with  $n$ .

### 2.1.1 Large Components

We say that a hypergraph  $H$  is *connected* if the *shadow graph* of  $H$  (that is, the graph with vertex set  $V(H)$  and edge set  $\{\{x, y\} : \{x, y\} \subseteq E(H)\}$ ) is connected. A *component* of a hypergraph is a maximal connected subgraph. Let  $r$  be a positive integer and  $H$  be a hypergraph. Let  $\chi_r : E(H) \rightarrow [r]$  be a coloring of the edges of  $H$ . Denote by  $\text{mc}(H, \chi_r)$  the order of the largest monochromatic component under  $\chi$  and let

$$\text{mc}_r(H) = \min_{\chi_r} \text{mc}(H, \chi_r).$$

For hypergraphs much less is known; however, Gyárfás [45] (see also Füredi and Gyárfás [36]) proved the following result. Let  $K_n^{(k)}$  denote the complete  $k$ -uniform hypergraph of order  $n$ .

**Theorem 2.1.1** (Gyárfás, 1977, [45]) *For all  $n \geq k \geq 1$ ,*

1.  $\text{mc}_k(K_n^{(k)}) = n$ , and
2.  $\text{mc}_{k+1}(K_n^{(k)}) \geq \frac{k}{k+1}n$ . *Furthermore, this is optimal when  $n$  is divisible by  $k + 1$ .*

The optimality statement is easy to see. Indeed, split the vertex set into  $k + 1$  parts  $V_1, \dots, V_{k+1}$  each of size  $n/(k + 1)$  and color the edges so that color  $i$  is not used on any edge which intersects  $V_i$ . Clearly each monochromatic component has size  $kn/(k + 1)$ .

Füredi and Gyárfás [36] and Gyárfás and Haxell [46] proved a number of other results regarding the value of  $\text{mc}_r(K_n^{(k)})$ ; however, in general, the value of  $\text{mc}_r(K_n^{(k)})$  is not known (see Section 2.7 for more details).

Our main theorem shows that in order to prove a random analog of any such result about  $\text{mc}_r(K_n^{(k)})$  it suffices to prove a nearly complete (or large minimum degree) analog of such a result. Let the *random hypergraph*  $\mathcal{H}^{(k)}(n, p)$  be the  $k$ -uniform hypergraph  $H$  with vertex set  $[n]$  in which every  $k$ -element set from  $\binom{[n]}{k}$  appears independently as an edge in  $H$  with probability  $p$ .

**Theorem 2.1.2** *Suppose a function  $\varphi = \varphi(r, k)$  satisfies the following condition: for all  $k \geq 2$ ,  $r \geq 1$ , and  $\alpha^* > 0$  there exists  $\epsilon^* > 0$  and  $t_0 > 0$  such that if  $G$  is a  $k$ -uniform hypergraph on  $t \geq t_0$  vertices with  $e(G) \geq (1 - \epsilon^*)\binom{t}{k}$ , then  $\text{mc}_r(G) \geq (\varphi - \alpha^*)t$ .*

*Then for any  $k \geq 2$ ,  $r \geq 1$ ,  $\alpha > 0$ , and  $p = p(n)$  such that  $pn^{k-1} \rightarrow \infty$  we have that a.a.s.  $\text{mc}_r(\mathcal{H}^{(k)}(n, p)) \geq (\varphi - \alpha)n$ .*

As an application, we prove a version of Theorem 2.1.1 for nearly complete hypergraphs (Sections 2.4 and 2.5) and then obtain a version for random hypergraphs as an immediate corollary (Section 2.3.2).

**Theorem 2.1.3** *For all  $\alpha > 0$  and  $k \geq 3$  there exists  $\epsilon > 0$  and  $n_0$  such that if  $G$  is a  $k$ -uniform hypergraph on  $n \geq n_0$  vertices with  $e(G) > (1 - \epsilon)\binom{n}{k}$ , then*

1.  $\text{mc}_k(G) \geq (1 - \alpha)n$ , and
2.  $\text{mc}_k(G) \geq \left(\frac{k}{k+1} - \alpha\right)n$ .

(We give an explicit bound on  $\alpha$  in terms of  $\epsilon$  in the proof of Theorem 2.4.1 and Corollary 2.5.2.)

**Corollary 2.1.4** *Let  $k \geq 3$ , let  $\alpha > 0$ , and let  $p = p(n)$  be such that  $pn^{k-1} \rightarrow \infty$ . Then a.a.s.*

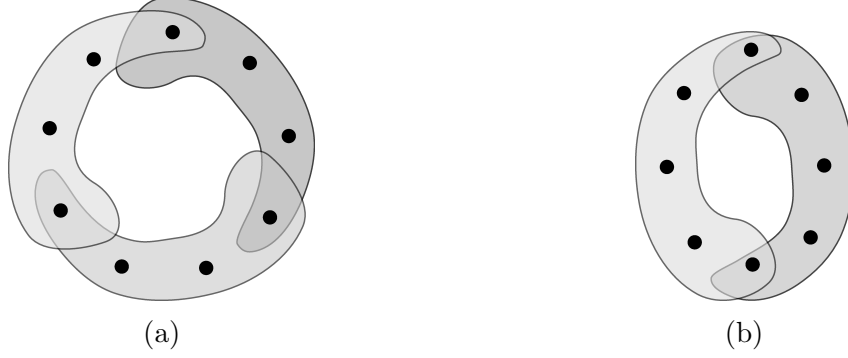


Figure 2.1: A 4-uniform loose cycle (a) and a 5-uniform diamond (b).

1.  $\text{mc}_k(\mathcal{H}^{(k)}(n, p)) \geq (1 - \alpha)n$ , and
2.  $\text{mc}_{k+1}(\mathcal{H}^{(k)}(n, p)) \geq \left(\frac{k}{k+1} - \alpha\right)n$ .

### 2.1.2 Loose Cycles

We say that a  $k$ -uniform hypergraph  $(V, E)$  is a *loose cycle* if there exists a cyclic ordering of the vertices  $V$  such that every edge consists of  $k$  consecutive vertices and every pair of consecutive edges intersects in a single vertex (see Figure 2.1a). Consequently,  $|E| = |V|/(k - 1)$ .

Let  $G$  be a  $k$ -uniform hypergraph. A *connected loose cycle packing on  $t$  vertices* is a connected sub-hypergraph  $H \subseteq G$  and a vertex disjoint collection of loose cycles  $C_1, \dots, C_\ell \subseteq H$  such that  $\sum_{i=1}^{\ell} |V(C_i)| = t$ . A *connected diamond matching on  $t$  vertices* is a connected loose cycle packing on  $t$  vertices such that every cycle consists of 2 edges and so has exactly  $2k - 2$  vertices (see Figure 2.1b). When  $k = 2$  we consider an edge to be a cycle with 2 vertices.

Gyárfás, Sárközy and Szemerédi [47] proved that for all  $k \geq 3$  and  $\eta > 0$ , there exists  $n_0$  such that if  $n \geq n_0$ , then every 2-edge-coloring of  $K_n^{(k)}$  contains a monochromatic cycle of length  $(1 - \eta)\frac{2k-2}{2k-1}n$ . (The  $k = 3$  case was previously proved by Haxell, Łuczak, Peng, Rödl, Ruciński, Simonovits, and Skokan [53].) Their proof follows from a more basic result

which says that for all  $k \geq 3$  and  $\eta > 0$ , there exists  $\epsilon > 0$  and  $n_0$  such that if  $G$  is a  $k$ -uniform hypergraph on  $n \geq n_0$  vertices with  $e(G) \geq (1 - \epsilon) \binom{n}{k}$ , then every 2-edge-coloring of  $G$  contains a monochromatic connected diamond matching on  $(1 - \eta) \frac{2k-2}{2k-1} n$  vertices.

We provide a result which reduces the problem of finding long monochromatic loose cycles in random hypergraphs to the problem of finding large monochromatic connected loose cycle packing in nearly complete hypergraphs (Section 2.3.3). Then applying the result of [47], we get an asymptotically tight result for 2-colored random  $k$ -uniform hypergraphs.

**Theorem 2.1.5** *Suppose a function  $\psi = \psi(r, k)$  satisfies the following condition: for all  $k \geq 2$ ,  $r \geq 1$ , and  $\alpha^* > 0$  there exists  $\epsilon^* > 0$  and  $t_0 > 0$  such that if  $G$  is a  $k$ -uniform hypergraph on  $t \geq t_0$  vertices with  $e(G) \geq (1 - \epsilon^*) \binom{t}{k}$ , then every  $r$ -coloring of the edges of  $G$  contains a monochromatic connected loose cycle packing on at least  $(\psi - \alpha^*)t$  vertices.*

*Then for any  $k \geq 2$ ,  $r \geq 1$ ,  $\alpha > 0$ , and  $p = p(n)$  such that  $pn^{k-1} \rightarrow \infty$  we have that a.a.s. every  $r$ -edge-coloring of  $\mathcal{H}^{(k)}(n, p)$  contains a monochromatic loose cycle on at least  $(\psi - \alpha)n$  vertices.*

Using a result of Gyárfás, Sárközy, and Szemerédi [47] about large monochromatic connected diamond matchings in nearly complete hypergraphs, we get the following corollary.

**Corollary 2.1.6** *Let  $k \geq 2$  and  $\alpha > 0$ . Choose  $p = p(n)$  such that  $pn^{k-1} \rightarrow \infty$ . Then a.a.s. there exists a monochromatic loose cycle on at least  $(\frac{2k-2}{2k-1} - \alpha)n$  vertices in any 2-coloring of the edges of  $\mathcal{H}^{(k)}(n, p)$ .*

## 2.2 Notation and Definitions

Let  $G$  be a  $k$ -uniform hypergraph. Let  $v \in V(G)$  and  $U \subseteq V(G)$ . Define

$$d(v, U) = \left| \left\{ S \in \binom{U}{k-1} : S \cup \{v\} \in E(G) \right\} \right|.$$

Furthermore, let  $\delta(G) = d(v, V(G))$ , which is just the *minimum degree of  $G$* . Suppose now that  $v \notin U$  and define the *restricted link graph of  $v$* , denoted by  $L(v, U)$ , with vertex set  $U$  and edge set  $\{S \in \binom{U}{k-1} : S \cup \{v\} \in E(G)\}$ . We call  $L(v) = L(v, V(G) \setminus \{v\})$  the *link graph of  $v$* , which clearly is a  $(k-1)$ -uniform hypergraph on  $n-1$  vertices.

Also recall that the *1-core* of  $G$  is the largest induced subgraph of  $G$  with no isolated vertices.

Now suppose that the edges of  $G$  are  $r$ -colored. We follow the convention that any edge  $e$  of the link graph of a vertex  $v$  inherits the color of the edge  $e \cup \{v\}$  in  $G$ .

For expressions such as  $n/t$  (for example the size of a cluster in the regularity lemma) that are supposed to be an integer, we always assume that  $n/t \in \mathbb{Z}$  by rounding appropriately without affecting the argument.

## 2.3 Random Hypergraphs

### 2.3.1 Sparse Weak Hypergraph Regularity

First, we will provide a few definitions.

Let  $H = (V, E)$  be a  $k$ -uniform hypergraph. Given pairwise-disjoint sets  $U_1, \dots, U_k \subseteq V$ , and  $p > 0$ , we define  $e_H(U_1, \dots, U_k)$  to be the total number of edges  $e = v_1 \dots v_k$  in  $H$  such that  $v_i \in U_i$  for all  $1 \leq i \leq k$ . Also, we define

$$d_{H,p}(U_1, \dots, U_k) = \frac{e_H(U_1, \dots, U_k)}{p|U_1| \cdots |U_k|}.$$

When the host graph  $H$  is clear from context, we may refer to  $e_H(U_1, \dots, U_k)$  as  $e(U_1, \dots, U_k)$  and  $d_{H,p}(U_1, \dots, U_k)$  as just  $d_p(U_1, \dots, U_k)$ .

For  $\epsilon > 0$ , we say the  $k$ -tuple  $(U_1, \dots, U_k)$  of pairwise-disjoint subsets of  $V$  is  $(\epsilon, p)$ -

regular if

$$|d_p(W_1, \dots, W_k) - d_p(U_1, \dots, U_k)| \leq \epsilon$$

for all  $k$ -tuples of subsets  $W_1 \subseteq U_1, \dots, W_k \subseteq U_k$  satisfying  $|W_1| \cdot \dots \cdot |W_k| \geq \epsilon |U_1| \cdot \dots \cdot |U_k|$ .

We say that  $H$  is a  $(\eta, p, D)$ -upper uniform hypergraph if for any pairwise disjoint sets  $U_1, \dots, U_k$  with  $|U_1| \geq \dots \geq |U_k| \geq \eta |V(H)|$ ,  $d_p(U_1, \dots, U_k) \leq D$ .

The following theorem is a sparse version of weak hypergraph regularity which will be the workhorse used to prove the main result for the random hypergraph. The sparse version of the regularity lemma [74] for graphs was discovered independently by Kohayakawa [57], and Rödl (see, for example, [14]), and subsequently improved by Scott [73]. The following is a straightforward generalization of their result for hypergraphs, which we state here without proof.

**Theorem 2.3.1** *For any given integers  $k \geq 2$ ,  $r \geq 1$ , and  $t_0 \geq 1$ , and real numbers  $\epsilon > 0$ ,  $D \geq 1$ , there are constants  $\eta = \eta(k, r, \epsilon, t_0, D) > 0$ ,  $T_0 = T_0(k, r, \epsilon, t_0, D) \geq t_0$ , and  $N_0 = N_0(k, r, \epsilon, t_0, D)$  such that any collection  $H_1, \dots, H_r$  of  $k$ -uniform hypergraphs on the same vertex set  $V$  with  $|V| \geq N_0$  that are all  $(\eta, p, D)$ -upper-uniform with respect to density  $0 < p \leq 1$  admits an equipartition (i.e. part sizes differ by at most 1) of  $V$  into  $t$  parts with  $t_0 \leq t \leq T_0$  such that all but at most  $\epsilon \binom{t}{k}$  of the  $k$ -tuples induce an  $(\epsilon, p)$ -regular  $k$ -tuple in each  $H_i$ .*

## 2.3.2 Large Components

We will use the following lemma to turn an  $(\epsilon, p)$ -regular  $k$ -tuple of some color into a large monochromatic subgraph.

**Lemma 2.3.2** *Let  $0 < \epsilon < 1/3$ , let  $k \geq 2$ , and let  $G = (V, E)$  be a  $k$ -partite hypergraph with vertex partition  $V = V_1 \cup V_2 \cup \dots \cup V_k$ . If for all  $\{U_1, \dots, U_k\}$ , with  $|U_i| \geq \epsilon |V_i|$  and  $U_i \subseteq V_i$*

for all  $i \in [k]$ , there exists an edge  $\{u_1, \dots, u_k\}$  with  $u_i \in U_i$  for all  $i \in [k]$ , then  $G$  contains a connected component  $H$  such that  $|V(H) \cap V_i| \geq (1 - \epsilon)|V_i|$  for all  $i \in [k]$ .

*Proof.* Start by choosing for each  $i \in [k]$  a set  $X_i \subseteq V_i$  with  $|X_i| = \lceil 3\epsilon|V_i| \rceil$  and let  $G'$  be the hypergraph induced by  $\{X_1, \dots, X_k\}$ . Suppose that no component of  $G'$  has intersection of size at least  $\epsilon|V_i|$  with any  $X_i$ . Choose  $t$  as small as possible so that there exist components  $H_1, \dots, H_t$  of  $G'$  such that for some  $\ell \in [k]$ ,  $|X_\ell \cap \bigcup_{i \in [t]} V(H_i)| \geq \epsilon|V_\ell|$ . By the minimality of  $t$ , we have that for all  $j \in [k]$ ,  $|X_j \cap \bigcup_{i \in [t]} V(H_i)| < 2\epsilon|V_j|$ . So let  $X'_\ell = X_\ell \cap \bigcup_{i \in [t]} V(H_i)$  and for all  $j \in [k] \setminus \{\ell\}$ , let  $X'_j = X_j \setminus \bigcup_{i \in [t]} V(H_i)$ . For all  $i \in [k]$ , we have  $|X'_i| \geq \epsilon|V_i|$ ; however, by the construction of the sets  $X'_1, \dots, X'_k$ , there is no edge in  $G'[X'_1 \cup \dots \cup X'_k]$  which violates the hypothesis.

So we may assume that some component  $H'$  of  $G'$  intersects, say  $V_1$ , in at least  $\epsilon|V_1|$  vertices. We will show that for each  $i$ ,  $|V(H') \cap X_i| \geq \epsilon|V_i|$ . Indeed, let us assume for a moment that there exists some  $j$  such that  $|V(H') \cap X_j| < \epsilon|V_j|$  and define for each  $i \in [k]$  the set  $Y_i$  according to the following rule. If  $|V(H') \cap X_i| \geq \epsilon|V_i|$ , then set  $Y_i = V(H') \cap X_i$ ; otherwise set  $Y_i = X_i \setminus V(H')$ . Clearly,  $|Y_i| \geq \epsilon|V_i|$ . By the hypothesis, there is an edge in the hypergraph  $G'[Y_1, \dots, Y_k]$  induced by  $\{Y_1, \dots, Y_k\}$ . But this cannot happen since  $Y_1 \subseteq V(H')$  and  $Y_j$  is disjoint from  $V(H')$ . Therefore, for each  $i$ ,  $|V(H') \cap X_i| \geq \epsilon|V_i|$ .

Suppose to the contrary and without loss of generality that  $|V(H') \cap V_k| < (1 - \epsilon)|V_k|$ . But as there must be an edge in  $G[V(H') \cap V_1, \dots, V(H') \cap V_{k-1}, V_k \setminus V(H')]$ , this is a contradiction. So we have the desired component  $H = H'$ .  $\square$

We can now prove the main theorem.

*Proof of Theorem 2.1.2.* Let  $r, k, \alpha$ , and  $\varphi = \varphi(r, k)$  be given. Set  $\alpha^* = \alpha/2$ . Let  $\epsilon^*$  and  $t_0$  be the constants guaranteed by the values of  $k, r$ , and  $\alpha^*$ . Let  $\epsilon < \min\{1/(2k), \alpha/(4\varphi), \epsilon^*\}$ . Thus, if  $\Gamma$  is a  $k$ -uniform hypergraph on  $t \geq t_0$  vertices with  $e(\Gamma) \geq (1 - \epsilon)\binom{t}{k} > (1 - \epsilon^*)\binom{t}{k}$ , then  $\text{mc}_r(\Gamma) \geq (\varphi - \alpha/2)t$ . Let  $H = (V, E) = \mathcal{H}^{(k)}(n, p)$ .



First observe that for any fixed positive  $\eta$  any sub-hypergraph  $H'$  of  $H$  is  $(\eta, p, 2)$ -upper uniform. Indeed, let  $U_1, \dots, U_k \subseteq V$  with  $|U_1| \geq \dots \geq |U_k| \geq \eta n$  be given. Then the expected number of edges in  $H$  having exactly one vertex in each  $U_i$  is  $|U_1| \cdot \dots \cdot |U_k| p \geq \eta^k n^k p$ . Thus (A.3), applied with  $\gamma = 1$ , implies that

$$\begin{aligned} \Pr(e(U_1, \dots, U_k) \geq 2|U_1| \cdot \dots \cdot |U_k| p) &\leq 2 \exp(-|U_1| \cdot \dots \cdot |U_k| p / 3) \\ &\leq 2 \exp(-\eta^k n^k p / 3) = o(2^{-kn}), \end{aligned}$$

since  $pn^{k-1} \rightarrow \infty$ . Consequently, a.a.s. the number of edges in  $H'$  is at most  $2|U_1| \cdot \dots \cdot |U_k| p$ . Finally, since the number of choices for  $U_i$ 's is trivially bounded from above by  $(2^n)^k = 2^{kn}$ , the union bound yields that a.a.s.  $H'$  is  $(\eta, p, 2)$ -upper uniform.

Apply Theorem 2.3.1 with  $k, r, t_0, \epsilon$ , as above, and  $D = 2$ . Let  $\eta, T_0$  and  $N_0$  be the constants that arise and assume that  $n \geq N_0$ . Let  $c$  be an  $r$ -coloring of the edges of  $H$  and let  $H_i$  be the subgraph of  $H$  induced by the  $i$ th color, that means,  $V(H_i) = V$  and  $E(H_i) = \{e \in E \mid c(e) = i\}$ . Then let  $t$  be the constant guaranteed by Theorem 2.3.1 for graphs  $H_i$  and let  $V_1 \cup \dots \cup V_t$  be the  $(\epsilon, p)$ -regular partition of  $V$ .

Let  $R$  be the  $k$ -uniform *cluster graph* with vertex set  $[t]$  where  $\{i_1, \dots, i_k\}$  is an edge if and only if  $V_{i_1}, \dots, V_{i_k}$  form an  $(\epsilon, p)$ -regular  $k$ -tuple. Color  $\{i_1, \dots, i_k\}$  in  $R$  by a majority color in the  $k$ -partite graph  $H[V_{i_1}, \dots, V_{i_k}]$ . Let  $H'$  be the sub-hypergraph colored by this color in  $H[V_{i_1}, \dots, V_{i_k}]$ . Observe that  $d_{H', p}(V_{i_1}, \dots, V_{i_k}) \geq 1/(2k)$ . Indeed, the Chernoff bound (A.3), applied with  $\gamma = 1/2$ , implies that a.a.s.  $e(V_{i_1}, \dots, V_{i_k}) \geq |V_{i_1}| \cdot \dots \cdot |V_{i_k}| p / 2$ . Thus,  $e_{H'}(V_{i_1}, \dots, V_{i_k}) \geq |V_{i_1}| \cdot \dots \cdot |V_{i_k}| p / (2k)$ , as required. Furthermore, since  $\epsilon < 1/(2k)$ , we also get that  $d_{H', p}(V_{i_1}, \dots, V_{i_k}) > \epsilon$ .

Clearly, we also have  $|E(R)| > (1 - \epsilon) \binom{t}{k}$ , so by the assumption, there is a monochromatic, say red, component of size at least  $(\varphi - \alpha/2)t$  in the cluster graph.

Now by Lemma 2.3.2 if  $V_i$  is contained in a red edge in the cluster graph  $R$ , at least

$(1 - \epsilon)|V_i|$  vertices inside  $V_i$  are in the same red component. Furthermore, since  $(1 - \epsilon)|V_i| > \frac{1}{2}|V_i|$ , any two red edges that intersect in  $R$  will correspond to red connected subgraphs that intersect in  $H$ , and thus are in the same red component in  $H$ . So, the  $(\varphi - \alpha/2)t$  vertices in the largest monochromatic component in  $R$  corresponds to a monochromatic component in  $H$  of order at least

$$\begin{aligned} (\varphi - \alpha/2)t(1 - \epsilon)|V_1| &\geq (\varphi - \alpha/2)t(1 - \epsilon)(1 - \epsilon)\frac{n}{t} \\ &= (\varphi - \alpha/2)(1 - 2\epsilon + \epsilon^2)n \geq (\varphi - \alpha/2 - 2\epsilon\varphi)n \geq (\varphi - \alpha)n, \end{aligned}$$

where the last inequality uses  $\epsilon < \alpha/(4\varphi)$ . □

### 2.3.3 Loose Cycles

We first generalize a result explicitly stated by Letzter [60, Corollary 2.1], but independently proved implicitly by Dudek and Prałat [22] and Pokrovskiy [69].

**Lemma 2.3.3** *Let  $H$  be a  $k$ -uniform  $k$ -partite graph with partite sets  $X_1, \dots, X_k$  such that  $|X_2| = \dots = |X_{k-1}| = m$  and  $|X_1| = |X_k| = m/2$  for some  $m$ . Then for all  $0 \leq \zeta \leq 1$  if there are no sets  $U_1 \subseteq X_1, \dots, U_k \subseteq X_k$  with  $|U_1| = \dots = |U_k| \geq \zeta m$  such that  $H[U_1, \dots, U_k]$  is empty, then there is a loose path on at least  $(1 - 4\zeta)m - 2$  edges.*

*Proof.* The proof is based on the depth first search algorithm. A similar idea for graphs was first noticed by Ben-Eliezer, Krivelevich and Sudakov [6, 7].

Let  $\zeta$  be given. We will proceed by a depth first search algorithm with the restriction that the vertices of degree 2 in the current path are in  $X_1 \cup X_k$ . We will let  $X_i^*$  be the current set of unexplored vertices in  $X_i$ . We will let  $X'_i$  be the current set of vertices which were added to the path at some point, but later rejected. So, initially we have  $X_i^* = X_i$  and  $X'_i = \emptyset$ . Start the algorithm by removing any vertex from  $X_1^*$  and adding it to the path  $P$ .

Suppose  $P$  is the current path. First consider the case where  $P$  has been reduced to a single vertex, say without loss of generality  $P = x_1 \in X_1$ . If there are no edges  $\{x_1, \dots, x_k\}$  where  $x_i \in X_i^*$  is unexplored for all  $2 \leq i \leq k$ , then we move  $x_1$  out of the path and back into  $X_1^*$  and start the algorithm again by choosing an edge in  $G[X_1^*, \dots, X_k^*]$  if we can and stopping the algorithm otherwise. If there is an edge  $\{x_1, \dots, x_k\}$  where  $x_i \in X_i^*$  is unexplored for all  $2 \leq i \leq k$ , then we add  $\{x_1, \dots, x_k\}$  to  $P$  and remove  $x_i$  from  $X_i^*$  for all  $2 \leq i \leq k$  and continue the algorithm from  $x_k$ .

Now assume  $P$  has at least one edge. Let the last edge of  $P$  be  $\{x_1, \dots, x_k\}$  and assume without loss of generality that  $x_1$  has degree 1 in  $P$  (i.e. is the current endpoint of the path). If there are no edges  $\{x_1, y_2, \dots, y_k\}$  where  $y_i \in X_i^*$  is unexplored for all  $2 \leq i \leq k$ , then we move  $x_i$  from the path to  $X'_i$  for all  $1 \leq i \leq k-1$  and continue the algorithm from  $x_k$ . If there is an edge  $\{x_1, y_2, \dots, y_k\}$  where  $y_i \in X_i^*$  is unexplored for all  $2 \leq i \leq k$ , then we add  $\{x_1, y_2, \dots, y_k\}$  to  $P$  and remove  $y_i$  from  $X_i^*$  for all  $2 \leq i \leq k$  and continue the algorithm from  $y_k$ .

Note that during every stage of the algorithm, there is no edge  $\{x_1, \dots, x_k\}$  where  $x_1 \in X'_1$  and  $x_i \in X_i^*$  for all  $2 \leq i \leq k$  and no edge  $\{y_1, \dots, y_k\}$  where  $y_k \in X'_k$  and  $y_i \in X_i^*$  for all  $2 \leq i \leq k$ . We also know that at every stage of the algorithm,  $|X'_2| = \dots = |X'_{k-1}| = |X'_1| + |X'_k|$  since every time a vertex from either  $X_1$  or  $X_k$  gets rejected, so do vertices from each  $X_i$ , and at no step does a vertex from both  $X_1$  and  $X_k$  get rejected. Furthermore,

$$|X_2^*| = \dots = |X_{k-1}^*| = |X_1^*| + |X_k^*| + 1$$

at every step where  $P \neq \emptyset$  since each time a vertex is removed from any  $X_i^*$ , for  $2 \leq i \leq k-1$ , exactly one vertex from  $X_1^* \cup X_k^*$  is removed as well, or one vertex is added to  $X_1^* \cup X_k^*$  then two are removed, with the exception of when the initial edge is selected. Note that at each step  $X_i = X_i^* \cup X'_i \cup (P \cap X_i)$  where the unions are disjoint.

Notice that this algorithm cannot stop while each  $|X_i^*| \geq \zeta m$ , since otherwise the  $X_i^*$  sets would violate the hypothesis.

After each step of the algorithm, either the difference between  $|X_1^*|$  and  $|X_1'|$  decreases by 1, or the difference between  $|X_k^*|$  and  $|X_k'|$  decreases by 1. So there is a stage in the algorithm where either  $|X_1'| = |X_1^*|$  or  $|X_k'| = |X_k^*|$  but not both. Suppose without loss of generality, we are at a stage where  $|X_1'| = |X_1^*|$  and  $|X_k'| < |X_k^*|$ . Since  $P$  always contains almost the same number of vertices in  $X_1$  as it does vertices in  $X_k$  (off by at most one), we have that the sums  $|X_1^*| + |X_1'|$  and  $|X_k^*| + |X_k'|$  differ by at most 1. This yields that

$$2|X_1^*| = |X_1^*| + |X_1'| \leq |X_k^*| + |X_k'| + 1 < 2|X_k^*| + 1$$

and so  $|X_1^*| \leq |X_k^*|$ . Thus,  $|X_2^*| = \dots = |X_{k-1}^*| = |X_1^*| + |X_k^*| + 1 \geq 2|X_1^*|$  for all  $1 \leq i \leq k$ . Thus, if  $|X_1'| \geq \zeta m$ , and consequently, each  $|X_i^*| \geq |X_1^*| = |X_1'| \geq \zeta m$ , we are done since  $H[X_1', X_2^*, \dots, X_k^*]$  has no edges, a contradiction. Otherwise, if  $|X_1'| < \zeta m$ , then

$$|P \cap X_1| = |X_1| - |X_1^*| - |X_1'| = m/2 - 2|X_1'| \geq m/2 - 2\zeta m.$$

But by our construction, each vertex in  $P \cap X_1$  corresponds to two edges in the path, except at most two vertices, so we have that there are at least  $m - 4\zeta m - 2$  edges in  $P$ .  $\square$

For a graph  $G$ , and hypergraph  $F$  with  $V(G) \subseteq V(F)$ , we say  $F$  is a *Berge- $G$*  if there is a bijection  $f : E(G) \rightarrow E(F)$  such that  $e \subseteq f(e)$  for all  $e \in E(G)$ . A Berge path  $(E_1, \dots, E_{\ell-1})$  is a *Berge- $P_\ell$*  where  $P_\ell = (e_1, \dots, e_{\ell-1})$ , and  $E_i = f(e_i)$ . We say  $F$  contains a *Berge- $G$*  if  $F$  contains a sub-hypergraph that is a *Berge- $G$* .

We say a hypergraph  $H$  has a Berge path, or *Berge- $P_\ell$*  if  $H$  contains a sequence of vertices and edges,  $v_1, e_1, v_2, e_2, \dots, v_{\ell-1}, e_{\ell-1}, v_\ell$ , where  $v_i, v_{i+1} \in e_i$  for each  $i \in \{1, 2, \dots, \ell-1\}$ . Note that a connected hypergraph has the property that between any two

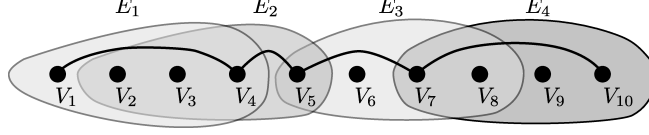


Figure 2.2: A Berge- $P_4$  path  $P = (E_1, E_2, E_3, E_4)$  between  $V_1$  and  $V_{10}$  with  $P_4 = (\{V_1, V_4\}, \{V_4, V_5\}, \{V_5, V_7\}, \{V_7, V_{10}\})$ . Observe that  $P$  contains no shorter Berge path between  $V_1$  and  $V_{10}$ .

vertices, there is a Berge path.

Let  $P = (E_1, \dots, E_\ell)$  be a Berge path in the cluster graph  $R$  of  $H$ . Assume that  $P$  contains no shorter Berge path that connects the two endpoints of  $P$ , which we will call  $V_1$  and  $V_s$ . Since there is no shorter Berge path, any two nonconsecutive edges  $E_i$  and  $E_j$  must be disjoint, and so there is a labeling  $V_1, \dots, V_s$  of the vertices of  $V(P)$  such that  $E_1 = \{V_1, \dots, V_k\}$ ,  $E_\ell = \{V_{s-k+1}, \dots, V_s\}$ , and for each  $i$  all the vertices of  $E_i \setminus E_{i+1}$  come before the vertices of  $E_i \cap E_{i+1}$ , which come before the vertices of  $E_{i+1} \setminus E_i$  (see Figure 2.2).

**Lemma 2.3.4** *Let  $P = (E_1, \dots, E_\ell)$  be a Berge path in the cluster graph  $R$  of  $H$  with  $|V(R)| = m$  and edges of density  $> \epsilon$ . Assume that  $P$  contains no shorter Berge path that connects the two endpoints of  $P$  and let  $V_1, \dots, V_s$  be defined as above. Let  $U_1 \subset V_1, \dots, U_s \subset V_s$  be such that  $|U_1|, |U_s| \geq \sqrt[k]{\epsilon m}$  and  $|U_2|, \dots, |U_{s-1}| \geq 2\sqrt[k]{\epsilon m}$ .*

*Then, there is a loose path in  $H$  that goes from a vertex in  $V_1$  to a vertex in  $V_s$  using only vertices from  $\bigcup_{i=1}^s U_i$ .*

*Proof.* Let us use induction on  $\ell$ . For  $\ell = 1$  then the statement follows from  $\epsilon$ -regularity. Now assume the statement holds for  $\ell - 1$  and consider such a path  $P$  of length  $\ell$ . Note that by construction of the sequence of vertices  $V_1, \dots, V_s$ , we have  $V_{s-k+1} \in E_{\ell-1} \cap E_\ell$ . Further,  $P - E_\ell$  is a shortest Berge path connecting  $V_1$  to  $V_{s-k+1}$  and  $U_1, \dots, U_{s-k+1}$  satisfy the requirements of the theorem, so by the inductive hypothesis, there is a loose path from  $V_1$  to  $V_{s-k+1}$  contained in  $\bigcup_{i=1}^{s-k+1} U_i$ . Let  $v_1 \in V_{s-k+1}$  be the last vertex in this path. If we replace  $U_{s-k+1}$  with  $U_{s-k+1} \setminus \{v_1\}$ , we can find another loose path from  $V_1$  to  $V_{s-k+1}$  that

does not use  $v_1$ . Since  $|U_{s-k+1}| \geq 2\sqrt[k]{\epsilon}m$ , we can repeat this process to find at least  $\sqrt[k]{\epsilon}m$  loose paths from  $V_1$  to  $V_{s-k+1}$  which all have distinct endpoints in  $V_{s-k+1}$ . Let  $X$  be the set of vertices in these at least  $\sqrt[k]{\epsilon}m$  paths. Observe that  $|X \cap U_s| = 0$  since  $P - E_\ell$  does not reach  $U_s$ , and for each  $s - k + 2 \leq i \leq s - 1$  we have that  $|X \cap U_i| \leq \sqrt[k]{\epsilon}$  since each path hits at most one vertex from each cluster, so for all  $s - k + 2 \leq i \leq s$ ,  $|U_i \setminus X| \geq \sqrt[k]{\epsilon}m$  and  $|U_{s-k+1} \cap X| \geq \sqrt[k]{\epsilon}m$ . So there is an edge in  $H[U_{s-k+1} \cap X, U_{s-k+2} \setminus X, \dots, U_s \setminus X]$ . This edge, along with whichever path from  $V_1$  to  $V_{s-k+1}$  is incident to it gives us the desired loose path from  $V_1$  to  $V_s$ .  $\square$

Now we prove the main result of this section.

*Proof of Theorem 2.1.5.* Let  $r, k, \alpha$ , and  $\psi = \psi(r, k)$  be given. Set  $\alpha^* = \alpha/2$ . Let  $\epsilon^*$  and  $t_0$  be the constants guaranteed by the values of  $k, r$ , and  $\alpha^*$ . Let  $\epsilon < \min\{1/(2k), (\alpha/(14\psi))^k, \epsilon^*\}$ . Thus, if  $\Gamma$  is a  $k$ -uniform hypergraph on  $t \geq t_0$  vertices with  $e(\Gamma) \geq (1 - \epsilon)\binom{t}{k} > (1 - \epsilon^*)\binom{t}{k}$ , then  $\Gamma$  contains a monochromatic connected loose cycle packing on at least  $(\psi - \alpha/2)t$  vertices for every  $r$ -edge-coloring.

Apply Theorem 2.3.1 with  $k, r, t_0, \epsilon$ , as above, and  $D = 2$ . Let  $\eta, T_0$  and  $N_0$  be the constants that arise. Let  $n \geq N_0$  be large. Let  $H = \mathcal{H}^{(k)}(n, p)$ . Color the edges of  $H$  with  $r$  colors. Let  $H_i$  be the sub-hypergraph of  $H$  colored  $i$  for  $1 \leq i \leq r$ . As in the proof of Theorem 2.1.2 it follows from the Chernoff bound (A.3) that each  $H_i$  is  $(\eta, p, D)$ -upper uniform. Then Theorem 2.3.1 gives us that  $H_1, \dots, H_r$  admits an  $(\epsilon, p)$ -regular partition.

Let  $R$  be the  $k$ -uniform cluster graph of  $H$  with clusters  $V_1, \dots, V_t$  each of size  $m = n/t$ . Color each edge  $\{V_{i_1}, \dots, V_{i_k}\}$  of  $R$  with a majority color in the  $k$ -partite graph  $H[V_{i_1}, \dots, V_{i_k}]$ . Let  $H'$  be the sub-hypergraph colored by this majority color in  $H[V_{i_1}, \dots, V_{i_k}]$ . Then again as in the proof of Theorem 2.1.2 we have  $d_{H', p}(V_{i_1}, \dots, V_{i_k}) \geq \epsilon$ .

Since  $t \geq t_0$ , and  $|E(R)| \geq (1 - \epsilon)\binom{t}{k}$ ,  $R$  contains a monochromatic connected loose cycle packing,  $C_1, \dots, C_\ell$ , on at least  $(\psi - \alpha/2)t$  clusters. Let  $f = \sum_{i=1}^\ell |E(C_i)| \geq (\psi -$

$\alpha/2)t/(k-1)$  be the number of edges in the packing. Let  $E_1, \dots, E_f$  be an enumeration of these edges. There is a bijection between the edges in the cycle packing and the vertices of degree 2 in the packing. We may assume without loss of generality that  $V_i$  is the vertex corresponding to  $E_i$  for each  $1 \leq i \leq f$ . Partition  $V_i$  into two equal sized sets  $V_{i,1}$  and  $V_{i,2}$  for each  $1 \leq i \leq f$ .

Fix  $1 \leq i \leq f$  for a moment. Let  $V_j \neq V_i$  be the second vertex of degree 2 in  $E_i$ . Set  $X_1 = V_{i,1}$ ,  $X_k = V_{j,2}$  and let  $X_2, \dots, X_{k-1}$  be the clusters in  $E_i \setminus \{V_i, V_j\}$ . Set  $\zeta = \sqrt[k]{\epsilon}$ . Then Lemma 2.3.3 gives us that there is a loose path, call it  $P_i$ , on exactly  $(1 - 4\sqrt[k]{\epsilon})m - 2$  edges in  $H'[X_1, \dots, X_k]$ . We can do this for each  $1 \leq i \leq f$  to find  $f$  long paths. Let  $U_{i,1}$  and  $U_{i,2}$  be the first and last  $\sqrt[k]{\epsilon}m$  vertices along  $P_i$  in  $V_i \cap V(P_i)$  respectively.

For each  $V_i$ , let  $\tau(V_i)$  denote the set of vertices in  $V_i$  that are not in any of the  $f$  long paths. Notice that if  $V_i$  is in the cycle packing,  $\tau(V_i) \geq 4\sqrt[k]{\epsilon}m$  by the length of each  $P_i$  and if  $V_i$  is not in the cycle packing,  $\tau(V_i) = |V_i| = m \geq 4\sqrt[k]{\epsilon}m$ . Let  $\mathcal{L}$  be the monochromatic component of  $R$  containing the cycle packing.

We now show how to find a loose path that connect the end of the path  $P_i$  to the beginning of  $P_{i+1}$  for each  $1 \leq i \leq f$ . Assume we have found vertex disjoint loose paths  $Q_1, \dots, Q_{i-1}$  in  $H'$  such that  $Q_j$  connects a vertex in  $U_{j,2} \subseteq V_j$  with a vertex in  $U_{j+1,1} \subseteq V_{j+1}$  and does not intersect any other vertices in  $\bigcup_{i=1}^f V(P_i)$ . Furthermore, let  $\mathcal{V}_{i-1} = \bigcup_{j=1}^{i-1} V(Q_j)$  and assume that  $|\mathcal{V}_{i-1} \cap \tau(V_j)| \leq 2(i-1)$  for each  $1 \leq j \leq t$ . Now we will construct a path  $Q_i$  as follows.

Let  $\Pi$  be the shortest Berge path from  $V_i$  to  $V_{i+1}$  in  $\mathcal{L}$ . Let  $s$  be the number of clusters in  $\Pi$ . Let  $V_{j_1} = V_i$  and  $V_{j_s} = V_{i+1}$  and let  $V_{j_2}, \dots, V_{j_{s-1}}$  be the other  $s-2$  clusters in  $\Pi$ . Let  $U_{j_1} = U_{i,2}$ ,  $U_{j_s} = U_{i+1,1}$  and  $U_{j_2} = \tau(V_{j_2}) \setminus \mathcal{V}_{i-1}, \dots, U_{j_{s-1}} = \tau(V_{j_{s-1}}) \setminus \mathcal{V}_{i-1}$ . Then we have  $|U_{j_1}|, |U_{j_s}| \geq \sqrt[k]{\epsilon}m$  (by definition of  $U_{i,2}$  and  $U_{i+1,1}$ ) and for each  $1 \leq \ell \leq s-1$ ,

$$|U_{j_\ell}| \geq \tau(V_{j_\ell}) - 2f \geq 4\sqrt[k]{\epsilon}m - 2\sqrt[k]{\epsilon}m = 2\sqrt[k]{\epsilon}m,$$

since  $f \leq \sqrt[k]{\epsilon} m$  for sufficiently large  $n$  (and so  $m$ ). Hence, by Lemma 2.3.4 we have a loose path  $Q_i$  from  $U_{i,2}$  to  $U_{i+1,1}$  that is disjoint from each short loose path found previously and that does not intersect any of the long paths except  $P_i$  and  $P_{i+1}$  at a single vertex in  $U_{i,2}$  and  $U_{i+1,1}$ , respectively. Finally, observe that for  $\mathcal{V}_i = \bigcup_{j=1}^i V(Q_j)$  we still have that

$$|\mathcal{V}_i \cap \tau(V_j)| = |\mathcal{V}_{i-1} \cap \tau(V_j)| + |V(Q_i) \cap \tau(V_j)| \leq 2(i-1) + 2 = 2i$$

for each  $1 \leq j \leq t$ , since if  $Q_i$  contained three or more vertices from a single cluster  $V_j$ , the corresponding Berge path  $\Pi$  would not have been a shortest Berge path.

Notice then these short paths  $Q_i$ 's together with the long paths  $P_i$ 's give us a long loose cycle,  $C$ . By our choice for each  $U_{i,1}$  and  $U_{i,2}$ , our cycle uses all the edges from each  $P_i$  except at most  $\sqrt[k]{\epsilon} m$  from each the beginning of the path and the end. There are  $f$  long paths, so the total number of edges in our cycle is at least

$$|E(C)| \geq \sum_{i=1}^f (|E(P_i)| - 2\sqrt[k]{\epsilon} m) \geq f(1 - 7\sqrt[k]{\epsilon})m \geq \frac{(\psi - \alpha/2)t}{k-1} (1 - 7\sqrt[k]{\epsilon})m \geq \frac{\psi - \alpha}{k-1} n,$$

where the last inequality uses  $\epsilon < (\alpha/(14\psi))^k$ . This immediately implies that  $|V(C)| \geq (\psi - \alpha)n$ , as required.  $\square$

The following result was proved by Gyárfás, Sárközy, and Szemerédi [47].

**Theorem 2.3.5** (Gyárfás, Sárközy, and Szemerédi, 2008, [47]) *Suppose that  $k$  is fixed and the edges of an almost complete  $k$ -uniform hypergraph on  $n$  vertices are 2-colored. Then there is a monochromatic connected diamond matching  $cD_k$  such that  $|V(cD_k)| = (1 + o(1))\frac{2k-2}{2k-1}n$ .*

Combining Theorem 2.3.5 with Theorem 2.1.5, we immediately get a proof of Corollary 2.1.6, which is best possible as observed in [47].

**Observation 2.3.6** *There exists a 2-coloring of the edges of the complete  $k$ -uniform hypergraph  $K_n^{(k)} = (V, E)$  such that the longest loose cycle covers no more than  $\frac{2k-2}{2k-1}n$  vertices.*



*Proof.* Choose any set  $S$  of  $\frac{2k-2}{2k-1}n$  vertices, and color all the edges completely inside  $S$  red, and the rest of the edges blue. Clearly the longest red loose cycle has no more than  $\frac{2k-2}{2k-1}n$  vertices. For the longest blue path, notice that each vertex in  $V \setminus S$  can be in at most two edges, and each edge must contain one such vertex. Thus the longest possible cycle is of length  $\frac{2}{2k-1}n$  or order  $\frac{2k-2}{2k-1}n$ .  $\square$

## 2.4 $k$ -Uniform Hypergraphs Colored with $k$ Colors

**Theorem 2.4.1** *Let  $k \geq 3$  and let  $0 < \epsilon < 16^{-k}$ . Then there exists some  $n_0$  such that for any  $k$ -uniform hypergraph  $G$  on  $n > n_0$  vertices with  $e(G) > (1 - \epsilon)\binom{n}{k}$ , we have  $\text{mc}_k(G) \geq (1 - 8\sqrt[k]{\epsilon})n$ .*

*Proof.* Let  $G$  be a  $k$ -uniform hypergraph on  $n$  vertices with  $e(G) > (1 - \epsilon)\binom{n}{k}$  and suppose that  $\text{mc}_k(G) < (1 - 8\sqrt[k]{\epsilon})n$ . Let  $G$  be  $k$ -colored in such a way that the largest monochromatic component of  $G$  is of size  $\text{mc}_k(G)$ .

First, for each  $i \in [k]$ , we find a partition  $\{A_i, B_i\}$  of  $V(G)$  such that  $(1 - 4\sqrt[k]{\epsilon})n \geq |A_i|, |B_i| \geq 4\sqrt[k]{\epsilon}n$  and there are no edges of color  $i$  which are incident with both a vertex in  $A_i$  and a vertex in  $B_i$ . To find such a partition, let  $C_i$  be the largest component of color  $i$ . By assumption  $|V(C_i)| < (1 - 8\sqrt[k]{\epsilon})n < (1 - 4\sqrt[k]{\epsilon})n$ . If  $|V(C_i)| \geq 4\sqrt[k]{\epsilon}n$ , then we can let  $A_i = V(C_i)$  and  $B_i = V(G) \setminus V(C_i)$ . If  $|V(C_i)| \leq 4\sqrt[k]{\epsilon}n$ , then some union of components of color  $i$ , call it  $C_i^*$ , will have the property that  $4\sqrt[k]{\epsilon}n \leq |V(C_i^*)| \leq 4\sqrt[k]{\epsilon}n + 4\sqrt[k]{\epsilon}n \leq n/2$ , since  $\epsilon < 1/16^k$ . Here we can let  $A_i = V(C_i^*)$  and  $B_i = V(G) \setminus V(C_i^*)$ .

We can also assume that  $|A_1 \cap A_2| \geq 2\sqrt[k]{\epsilon}n$  and  $|B_1 \cap B_2| \geq 2\sqrt[k]{\epsilon}n$ . Indeed, if  $|A_1 \cap A_2| < 2\sqrt[k]{\epsilon}n$ , then  $|B_1 \cap A_2| > 2\sqrt[k]{\epsilon}n$ . In this case, we can switch the roles of  $A_2$  and  $B_2$  and now we have  $|B_1 \cap B_2| > 2\sqrt[k]{\epsilon}n$  and  $|A_1 \cap A_2| = |A_1| - |A_1 \cap B_2| > |A_1| - 2\sqrt[k]{\epsilon}n \geq 2\sqrt[k]{\epsilon}n$ . Similarly, if  $|B_1 \cap B_2| < 2\sqrt[k]{\epsilon}n$ , switching the roles of  $A_2$  and  $B_2$  gives us the desired inequalities. Thus we can always assume that  $|A_1 \cap A_2| \geq 2\sqrt[k]{\epsilon}n$  and  $|B_1 \cap B_2| \geq 2\sqrt[k]{\epsilon}n$ .

Furthermore, we can assume that  $|A_1 \cap A_2 \cap \dots \cap A_k| \geq 2\sqrt[k]{\epsilon n}/2^{k-2}$ . Indeed, we already have that  $|A_1 \cap A_2| \geq 2\sqrt[k]{\epsilon n}$ . Now for a fixed  $i$  satisfying  $3 \leq i \leq k$  assume that  $|A_1 \cap \dots \cap A_{i-1}| \geq 2\sqrt[k]{\epsilon n}/2^{i-3}$ . Since  $\{A_i \cap (A_1 \cap \dots \cap A_{i-1}), B_i \cap (A_1 \cap \dots \cap A_{i-1})\}$  is a partition of  $A_1 \cap \dots \cap A_{i-1}$ , if  $A_i$  does not cover at least half of vertices in  $A_1 \cap \dots \cap A_{i-1}$ , then  $B_i$  does, so switching the roles of  $A_i$  and  $B_i$  if necessary will give us that  $|A_1 \cap \dots \cap A_i| \geq 2\sqrt[k]{\epsilon n}/2^{i-2}$ . Thus, we can recursively choose  $A_i$  such that  $|A_1 \cap A_2 \cap \dots \cap A_k| \geq 2\sqrt[k]{\epsilon n}/2^{k-2}$ .

Now observe that the number of  $k$ -tuples of distinct vertices  $(x_1, \dots, x_k)$  where  $x_1 \in A_1 \cap A_2 \cap \dots \cap A_k$ ,  $x_2 \in B_1 \cap B_2$ , and  $x_i \in B_i$  for all  $3 \leq i \leq k$ , is at least

$$2\sqrt[k]{\epsilon n}/2^{k-2} \cdot (2\sqrt[k]{\epsilon n} - 1) \cdot (4\sqrt[k]{\epsilon n} - 2) \cdots (4\sqrt[k]{\epsilon n} - k + 1) = 2^k \epsilon n^k + O(n^{k-1}). \quad (2.1)$$

Note that any such  $k$ -tuple intersects every set  $A_1, \dots, A_k, B_1, \dots, B_k$  and thus no such  $k$ -tuple can be an edge of  $G$  since if it were an edge of  $G$ , it would be colored, say by color  $i$ , and thus cannot contain vertices from both  $A_i$  and  $B_i$ . Since there are less than  $\epsilon \binom{n}{k}$  non-edges in  $G$ , there are less than  $\epsilon n^k$  ordered  $k$ -tuples of distinct vertices  $(x_1, \dots, x_k)$  where  $\{x_1, \dots, x_k\}$  is not an edge, which contradicts (2.1) for  $n$  sufficiently large.  $\square$

## 2.5 $k$ -Uniform Hypergraphs Colored with $k + 1$ Colors

The following proof makes use of some technical lemmas which can all be found in Section 2.6.

**Theorem 2.5.1** *Let  $k \geq 3$  and  $0 < \epsilon < \frac{1}{k^{5k} 2^{8k}}$ . If  $G$  is a  $k$ -uniform hypergraph on  $n$  vertices with  $\delta(G) \geq (1 - \epsilon) \binom{n-1}{k-1}$ , then  $\text{mc}_{k+1}(G) \geq (\frac{k}{k+1} - \sqrt{k^k \epsilon})n$ .*

*Proof.* Suppose that the edges of  $G$  are  $(k + 1)$ -colored. First note for all  $v \in V(G)$ , the link graph of  $v$  is a  $(k + 1)$ -colored  $(k - 1)$ -uniform hypergraph on  $n - 1$  vertices with at least  $(1 - \epsilon) \binom{n-1}{k-1}$  edges and thus by Lemma 2.6.3 (applied with  $k - 1$  in place of  $k$  and  $\ell = 2$ ), we get a monochromatic 1-core in the link graph of  $v$  of order at least  $(\frac{k-1}{k+1} - \sqrt{\epsilon})(n-1)$ . Since all

these vertices are connected via  $v$ , this gives us a monochromatic component containing  $v$  of order at least  $(\frac{k-1}{k+1} - \sqrt{\epsilon})(n-1) + 1 \geq (\frac{k-1}{k+1} - \sqrt{\epsilon})n$ . Now let  $C_1$  be the largest monochromatic component in  $G$ , and note that by the calculation above,

$$|C_1| \geq \left( \frac{k-1}{k+1} - \sqrt{\epsilon} \right) n.$$

For ease of reading, set  $\alpha = k^k \epsilon$ . For any vertex  $v \in V(G) \setminus C_1$ , we have

$$d(v, C_1) \geq \binom{|C_1|}{k-1} - \epsilon \binom{n-1}{k-1} \geq \left( 1 - \frac{k\epsilon}{\left(\frac{k-1}{k+1} - \sqrt{\epsilon}\right)^{k-1}} \right) \binom{|C_1|}{k-1} \geq (1-\alpha) \binom{|C_1|}{k-1}, \quad (2.2)$$

where the second inequality holds by Observation 2.6.1 (applied with  $U = C_1$  and  $\lambda = \frac{k-1}{k+1} - \sqrt{\epsilon}$ ) and the last inequality holds since  $\frac{k-1}{k+1} - \sqrt{\epsilon} \geq \frac{1}{k}$ .

Now choose a monochromatic component  $C_2$  so that  $|C_1 \cap C_2|$  is maximized. We claim that

$$|C_1 \cap C_2| \geq \left( \frac{k-1}{k} - \sqrt{\alpha} \right) n.$$

To see this, let  $v \in V(G) \setminus C_1$  and note that the link graph of  $v$  restricted to  $C_1$  is a  $(k-1)$ -uniform hypergraph on  $|C_1|$  vertices with at least  $(1-\alpha) \binom{|C_1|}{k-1}$  edges (by (2.2)). Furthermore, note that the link graph of  $v$  restricted to  $C_1$  only uses  $k$  colors since the color of  $C_1$  cannot show up by our choice of  $v \notin C_1$ . Thus by Lemma 2.6.3 (with  $k$  replaced by  $k-1$  and  $\ell = 1$ ) there is a monochromatic component  $C_2$  containing  $v$  such that  $|C_1 \cap C_2| \geq \left( \frac{k-1}{k} - \sqrt{\alpha} \right) |C_1|$ , thus proving the claim.

Now if  $V \setminus C_1 \subseteq C_2$ , then we have

$$|C_1| \geq |C_2| = n - |C_1| + |C_1 \cap C_2| \geq n - |C_1| + \left( \frac{k-1}{k} - \sqrt{\alpha} \right) |C_1|,$$

which implies

$$|C_1| \geq \frac{n}{\frac{k+1}{k} + \sqrt{\alpha}} \geq \left( \frac{k}{k+1} - \sqrt{\alpha} \right) n;$$

and we are done, so suppose not.

We now show that  $|C_1 \setminus C_2|$  must be small (in other words,  $|C_1 \cap C_2|$  must be big); more precisely, we will show that

$$|C_1 \setminus C_2| \leq 8^{2(k-1)\sqrt{\alpha}} |C_1|. \quad (2.3)$$

Let  $v \in V(G) \setminus (C_1 \cup C_2)$  and consider the link graph  $L'$  obtained from  $L(v, C_1)$  by deleting all edges entirely contained in  $C_1 \setminus C_2$ .

Notice that  $L(v, C_1)$  is a  $(k-1)$ -uniform hypergraph with at least  $(1-\alpha) \binom{|C_1|}{k-1}$  edges (by (2.2)). Furthermore,  $L'$  is  $(k-1)$ -colored since it cannot contain any edges whose color matches  $C_1$ , and if there was an edge in  $L(v, C_1)$  whose color matched  $C_2$ , it would need to be contained in  $C_1 \setminus C_2$ , but such edges do not exist in  $L'$ . Thus by Lemma 2.6.4 (applied with  $G' = L(v, C_1)$ ,  $G = L'$ ,  $A = C_1 \cap C_2$  and  $B = C_1 \setminus C_2$ ) either there is a monochromatic component containing  $v$  and at least  $(1 - 8^{2(k-1)\sqrt{\alpha}})|C_1|$  vertices of  $C_1$ , in which case (2.3) holds since  $C_2$  was assumed to be the component that maximizes  $|C_1 \cap C_2|$ , or there is a monochromatic component containing  $v$  and at least

$$|C_1 \cap C_2| - \sqrt{\alpha}|C_1| + \left( \frac{k-2}{k-1} - \frac{\sqrt{k-1}^{4(k-1)\sqrt{\alpha}}}{2^{k-2}} \right) |C_1 \setminus C_2| \quad (2.4)$$

vertices of  $C_1$ . Observe that it cannot be the case that

$$\left( \frac{k-2}{k-1} - \frac{\sqrt{k-1}^{4(k-1)\sqrt{\alpha}}}{2^{k-2}} \right) |C_1 \setminus C_2| > \sqrt{\alpha}|C_1|,$$

since this and (2.4) would imply that we have a monochromatic component containing  $v$  and

more than  $|C_1 \cap C_2|$  vertices of  $C_1$ , which contradicts the choice of  $C_2$ . So it must be that  $\left(\frac{k-2}{k-1} - \frac{\sqrt{k-1} \sqrt[4]{(k-1)\sqrt{\alpha}}}{2^{k-2}}\right) |C_1 \setminus C_2| \leq \sqrt{\alpha} |C_1|$ , which implies

$$|C_1 \setminus C_2| \leq \frac{\sqrt{\alpha} |C_1|}{\frac{k-2}{k-1} - \frac{\sqrt{k-1} \sqrt[4]{(k-1)\sqrt{\alpha}}}{2^{k-2}}}. \quad (2.5)$$

We further claim that

$$\frac{\sqrt{\alpha} |C_1|}{\frac{k-2}{k-1} - \frac{\sqrt{k-1} \sqrt[4]{(k-1)\sqrt{\alpha}}}{2^{k-2}}} < 8 \sqrt[2^{(k-1)}]{\alpha} |C_1|. \quad (2.6)$$

Indeed, first note that since  $\alpha < 2^{-8k}$  and  $k \geq 3$ , we get  $\sqrt[4]{(k-1)\sqrt{\alpha}} < 2^{-2k/(k-1)} \leq 1/8$ ,  $\frac{\sqrt{k-1}}{2^{k-2}} \leq 1$  and  $\frac{k-2}{k-1} \geq 1/2$ . Consequently,

$$\frac{k-2}{k-1} - \frac{\sqrt{k-1} \sqrt[4]{(k-1)\sqrt{\alpha}}}{2^{k-2}} \geq \frac{1}{2} - \frac{1}{8} > \frac{1}{8},$$

which together with  $\sqrt{\alpha} \leq \sqrt[2^{(k-1)}]{\alpha}$  yields (2.6). Thus by (2.5) and (2.6),

$$|C_1 \setminus C_2| < 8 \sqrt[2^{(k-1)}]{\alpha} |C_1|$$

as desired.

Now that we have established (2.3), we will show that any vertex in  $V(G) \setminus C_1$  is in some monochromatic component that has large intersection with  $C_1$ . More precisely, for all  $v \in V(G) \setminus C_1$ , there exist some monochromatic component  $C \ni v$  such that  $|C \cap C_1| \geq (1 - 9 \sqrt[2^{(k-1)}]{\alpha}) |C_1|$ . Indeed, let  $v \in V(G) \setminus C_1$ . If  $v \in C_2$ , then (2.3) implies that  $C = C_2$  suffices. Otherwise, as in (2.3) and (2.4) (with  $C_2$  replaced by  $C$ ) we get that there is a monochromatic component containing  $v$  and at least

$$\begin{aligned} |C_1 \cap C| - \sqrt{\alpha} |C_1| &= |C_1| - |C_1 \setminus C| - \sqrt{\alpha} |C_1| \\ &\geq (1 - 8 \sqrt[2^{(k-1)}]{\alpha} - \sqrt{\alpha}) |C_1| \geq (1 - 9 \sqrt[2^{(k-1)}]{\alpha}) |C_1| \end{aligned}$$

vertices of  $C_1$ .

Since there are only  $k$  possible colors for edges between  $C_1$  and  $V(G) \setminus C_1$ , there must exist at least  $(n - |C_1|)/k$  vertices in  $V(G) \setminus C_1$  that are all in some monochromatic component  $C^*$  together along with at least  $(1 - 9^{-2(k-1)\sqrt{\alpha}})|C_1|$  vertices of  $C_1$ . Thus,

$$|C_1| \geq |C^*| \geq (n - |C_1|)/k + (1 - 9^{-2(k-1)\sqrt{\alpha}})|C_1|,$$

where the first inequality holds by the maximality of  $C_1$ . So solving for  $C_1$  we finally obtain that

$$|C_1| \geq \frac{n}{k \left( \frac{1}{k} + 9^{-2(k-1)\sqrt{\alpha}} \right)} > \left( \frac{k}{k+1} - \sqrt{\alpha} \right) n,$$

where the final inequality follows from the upper bound on  $\epsilon$  (and so on  $\alpha$ ).  $\square$

**Corollary 2.5.2** *Let  $k \geq 3$  and  $0 < \epsilon < \frac{1}{k^{10k+2} 2^{16k+2}}$ . If  $G$  is a  $k$ -uniform hypergraph on  $n$  vertices with  $e(G) \geq (1 - \epsilon) \binom{n}{k}$ , then  $\text{mc}_{k+1}(G) \geq \left( \frac{k}{k+1} - 2k^{\frac{k+1}{2}} \sqrt[4]{\epsilon} \right) n$ .*

*Proof.* Apply Observation 2.6.2 to get a subgraph  $G' = (V', E')$  with  $|V'| \geq (1 - \sqrt{\epsilon})n$  and  $\delta(G') \geq (1 - 2k\sqrt{\epsilon}) \binom{|V'|-1}{k-1}$ . Now by Theorem 2.5.1, we have

$$\text{mc}_{k+1}(G) \geq \text{mc}_{k+1}(G') \geq \left( \frac{k}{k+1} - \sqrt{2} k^{\frac{k+1}{2}} \sqrt[4]{\epsilon} \right) (1 - \sqrt{\epsilon})n \geq \left( \frac{k}{k+1} - 2k^{\frac{k+1}{2}} \sqrt[4]{\epsilon} \right) n.$$

$\square$

## 2.6 Technical Lemmas

The first two observations are simple counting arguments, but it is convenient for us to state them explicitly.

**Observation 2.6.1** Let  $n \geq k \geq 2$ , let  $0 < \epsilon, \lambda \leq 1$ . If  $|U| = \lambda n \geq k^2$ , then

$$\binom{|U| - 1}{k - 1} - \epsilon \binom{n - 1}{k - 1} \geq \left(1 - \frac{k\epsilon}{\lambda^{k-1}}\right) \binom{|U| - 1}{k - 1}.$$

*Proof.* First we show that  $\lambda n \geq k^2$  implies that  $\frac{n - (k-1)}{\lambda n - (k-1)} \leq \frac{{}^{k-1}\sqrt{k}}{\lambda}$ . By Bernoulli's inequality (A.5) we get that

$$k^{k/(k-1)} = (1 + (k - 1))^{k/(k-1)} \geq 1 + k,$$

and equivalently,  ${}^{k-1}\sqrt{k} \geq 1 + 1/k$ . Thus,

$$\begin{aligned} {}^{k-1}\sqrt{k} &\geq 1 + \frac{1}{k} = 1 + \frac{k - 1}{k(k - 1)} \geq 1 + \frac{k - 1}{k^2 - k + 1} \\ &\geq 1 + \frac{k - 1}{\lambda n - k + 1} = \frac{\lambda n}{\lambda n - (k - 1)} \geq \frac{\lambda n - \lambda(k - 1)}{\lambda n - (k - 1)}. \end{aligned}$$

Now,

$$\begin{aligned} \binom{|U| - 1}{k - 1} - \epsilon \binom{n - 1}{k - 1} &= \binom{|U| - 1}{k - 1} - \epsilon \binom{|U| - 1}{k - 1} \cdot \frac{\binom{n-1}{k-1}}{\binom{|U|-1}{k-1}} \\ &\geq \binom{|U| - 1}{k - 1} - \epsilon \binom{|U| - 1}{k - 1} \left(\frac{n - (k - 1)}{\lambda n - (k - 1)}\right)^{k-1} \\ &\geq \left(1 - \frac{k\epsilon}{\lambda^{k-1}}\right) \binom{|U| - 1}{k - 1}. \end{aligned}$$

□

**Observation 2.6.2** Let  $k \geq 2$ , let  $0 < \eta < 1$ , and let  $0 < \epsilon \leq (1 - {}^{k-1}\sqrt{1/2})^{1/(1-\eta)}$ . Let  $G = (V, E)$  be a  $k$ -uniform hypergraph on  $n$  vertices with  $e(G) \geq (1 - \epsilon)\binom{n}{k}$ . For all  $U \subseteq V$  all but at most  $\epsilon^{1-\eta}n$  vertices have

$$d(v, U) \geq \binom{|U| - 1}{k - 1} - \epsilon^\eta \binom{n - 1}{k - 1}. \quad (2.7)$$

In particular, if  $(1 - \epsilon^{1-\eta})n \geq k^2$ , then there exists an induced subgraph  $G' = (V', E')$  with  $|V'| \geq (1 - \epsilon^{1-\eta})n$  and  $\delta(G') \geq (1 - 2k\epsilon^\eta) \binom{|V'|-1}{k-1}$ .

*Proof.* Let  $V^* = \{v \in V : d(v) < \binom{n-1}{k-1} - \epsilon^\eta \binom{n-1}{k-1}\}$ . We have

$$|V^*| \frac{\epsilon^\eta}{k} \binom{n-1}{k-1} \leq e(\bar{G}) \leq \epsilon \binom{n}{k},$$

which implies  $|V^*| \leq \epsilon^{1-\eta}n$ . So let  $U \subseteq V$  with  $|U| = \lambda n \geq k^2$  and note that for all  $v \in V \setminus V^*$  we have

$$d(v, U) \geq \binom{|U|-1}{k-1} - \epsilon^\eta \binom{n-1}{k-1}, \quad (2.8)$$

as required.

To see the final statement, let  $V' = V \setminus V^*$  and note that  $|V'| \geq (1 - \epsilon^{1-\eta})n$  by the calculation of  $|V^*|$  above. From (2.8) (applied with  $U = V'$ ) we have

$$\delta(G[V']) \geq \binom{|V'|-1}{k-1} - \epsilon^\eta \binom{n-1}{k-1}.$$

Now Observation 2.6.1 with  $\lambda = 1 - \epsilon^{1-\eta}$  yields

$$\delta(G[V']) \geq \left(1 - \frac{k\epsilon^\eta}{(1 - \epsilon^{1-\eta})^{k-1}}\right) \binom{|V'|-1}{k-1} \geq (1 - 2k\epsilon^\eta) \binom{|V'|-1}{k-1},$$

where the last inequality holds since  $1 - \epsilon^{1-\eta} \geq \sqrt[k-1]{1/2}$ .

□

**Lemma 2.6.3** *Let  $k \geq 2$ ,  $\ell \geq 1$ , and  $0 < \epsilon < 256^{-k}$ . Then there exists  $n_0 > 0$ , a sufficiently large integer, such that if  $G = (V, E)$  is a  $(k + \ell)$ -colored  $k$ -uniform hypergraph on  $n \geq n_0$  vertices with  $e(G) > (1 - \epsilon) \binom{n}{k}$ , then  $G$  contains a monochromatic 1-core on at least  $(\frac{k}{k+\ell} - \sqrt{\epsilon})n$  vertices.*

*Proof.* By (2.7) in Observation 2.6.2 (applied with  $\eta = 1/2$ ,  $\lambda = 1$ , and  $U = V$ ) at least



$(1 - \sqrt{\epsilon})n$  vertices have degree at least  $(1 - \sqrt{\epsilon})\binom{n-1}{k-1}$ . Call these vertices  $V'$ .

Suppose that the 1-core of each of the  $k + \ell$  colors has order less than  $(\frac{k}{k+\ell} - \sqrt{\epsilon})n$ ; that is, for each color  $i \in [k + \ell]$ , more than  $(\frac{\ell}{k+\ell} + \sqrt{\epsilon})n$  vertices are not incident with an edge of color  $i$ . Let  $X$  be the set of vertices that are not incident with at least  $\ell + 1$  colors. We claim that  $|X| \geq \frac{k+\ell}{k}\sqrt{\epsilon}n > \sqrt{\epsilon}n$ . Indeed, let  $T \subseteq V \times [k + \ell]$  be the set of ordered pairs where  $(v, t) \in T$  if the vertex  $v$  is not incident with any edge of color  $t$ . Since for each  $i \in [k + \ell]$ , there are more than  $(\frac{\ell}{k+\ell} + \sqrt{\epsilon})n$  vertices that are not incident with an edge of color  $i$ , we have that

$$|T| \geq (k + \ell) \left( \frac{\ell}{k + \ell} + \sqrt{\epsilon} \right) n = \ell n + \sqrt{\epsilon}(k + \ell)n. \quad (2.9)$$

Now each vertex of  $V \setminus X$  contributes no more than  $\ell$  ordered pairs to the set  $T$ , and each vertex in  $X$  can contribute no more than  $k + \ell$ , so we have that

$$|T| \leq (n - |X|)\ell + |X|(k + \ell) = \ell n + k|X|. \quad (2.10)$$

Combining (2.9) and (2.10) gives us that  $\ell n + k|X| \geq \ell n + \sqrt{\epsilon}(k + \ell)n$ , which implies that  $|X| \geq \frac{k+\ell}{k}\sqrt{\epsilon}n \geq \sqrt{\epsilon}n$ , which proves the claim.

By the bounds on  $|X|$  and  $|V'|$  there exists  $v \in V' \cap X$ , and the link graph of  $v$  is a  $(k - 1)$ -uniform hypergraph on  $n - 1$  vertices with at least  $(1 - \sqrt{\epsilon})\binom{n-1}{k-1}$  edges which is colored with at most  $k - 1$  colors and thus by Theorem 2.4.1, there exists a monochromatic component on at least

$$(1 - 8\sqrt[2k]{\epsilon})(n - 1) \geq \left( \frac{k}{k + \ell} - \sqrt{\epsilon} \right) n$$

vertices, where the above inequality holds by the choice of  $\epsilon$ . So, the 1-core of some color must have order at least  $(\frac{k}{k+\ell} - \sqrt{\epsilon})n$ .  $\square$

**Lemma 2.6.4** *Let  $k \geq 2$  and  $0 < \epsilon \leq 512^{-k}$ . Then there exists a sufficiently large integer*

$n_0$  such that the following holds. Let  $G$  be a graph obtained from a  $k$ -uniform hypergraph  $G'$  on  $n \geq n_0$  vertices with  $e(G') \geq (1 - \epsilon) \binom{n}{k}$  by letting  $\{A, B\}$  be a partition of  $V(G')$  with  $|A| > \sqrt{\epsilon}n$ , and deleting all edges which lie entirely inside  $B$ . Then for any  $k$ -edge-coloring of  $G$  there exists a monochromatic 1-core in  $G$  on at least

$$\min \left\{ (1 - 8 \sqrt[2k]{\epsilon})n, |A| - \sqrt{\epsilon}n + \left( \frac{k-1}{k} - \frac{\sqrt{k} \sqrt[4k]{\epsilon}}{2^{k-1}} \right) |B| \right\}$$

vertices.

*Proof.* By (2.7) in Observation 2.6.2 (applied to  $G'$  with parameters  $\eta = 1/2$ ,  $\lambda = 1$ , and  $U = V$ ) we get that all but at most  $\sqrt{\epsilon}n$  vertices in  $G'$  have degree at least  $(1 - \sqrt{\epsilon}) \binom{n-1}{k-1}$ . Let  $A'$  be the vertices in  $A$  which have degree at least  $(1 - \sqrt{\epsilon}) \binom{n-1}{k-1}$  in  $G$  and note that since the degrees of the vertices in  $A$  are the same in  $G$  as in  $G'$ , we have that  $|A'| \geq |A| - \sqrt{\epsilon}n > 0$ . First suppose that there exists a color  $i \in [k]$  and a vertex  $v \in A'$  such that  $v$  is not contained in the 1-core of color  $i$ . Then the link graph of  $v$  in  $G$  is a  $(k-1)$ -colored  $(k-1)$ -graph on  $n-1$  vertices with at least  $(1 - \sqrt{\epsilon}) \binom{n-1}{k-1}$  edges. By Theorem 2.4.1, we have a monochromatic component in the link graph of  $v$  of order at least  $(1 - 8 \sqrt[2k]{\epsilon})(n-1)$  which together with  $v$  gives a monochromatic component (and so a 1-core) in  $G$  of order  $(1 - 8 \sqrt[2k]{\epsilon})(n-1) + 1 \geq (1 - 8 \sqrt[2k]{\epsilon})n$ . Otherwise every vertex in  $A'$  is contained in the 1-core of every color. Now if  $|A'| \geq (1 - 8 \sqrt[2k]{\epsilon})n$ , then we have a 1-core of the desired size, so suppose  $|A'| < (1 - 8 \sqrt[2k]{\epsilon})n$ , which implies

$$|B| = n - |A| \geq n - |A'| - \sqrt{\epsilon}n \geq 4 \sqrt[2k]{\epsilon}n.$$

Let  $v \in A'$  and consider the link graph of  $v$  restricted to  $B$ , which is a  $(k-1)$ -uniform hypergraph on  $|B|$  vertices colored with  $k$  colors which, by Observation 2.6.1 (applied with

$U = B$ ,  $\lambda = |B|/n \geq 4 \sqrt[2k]{\epsilon}$ , has at least

$$\begin{aligned} \binom{|B|}{k-1} - \sqrt{\epsilon} \binom{n-1}{k-1} &\geq \left(1 - \frac{k\sqrt{\epsilon}}{\lambda^{k-1}}\right) \binom{|B|}{k-1} \\ &\geq \left(1 - \frac{k\sqrt{\epsilon}}{(4 \sqrt[2k]{\epsilon})^{k-1}}\right) \binom{|B|}{k-1} = \left(1 - \frac{k \sqrt[2k]{\epsilon}}{4^{k-1}}\right) \binom{|B|}{k-1} \end{aligned}$$

edges. Now by Lemma 2.6.3 (applied to the link graph of  $v$  with  $k$  replaced by  $k-1$  and  $\ell = 1$ ), there is a monochromatic 1-core of color  $i$  which contains at least  $\left(\frac{k-1}{k} - \frac{\sqrt{k} \sqrt[4k]{\epsilon}}{2^{k-1}}\right) |B|$  vertices of  $B$ , and since every vertex in  $A'$  is in the 1-core of color  $i$ , the total size of the 1-core of color  $i$  is at least

$$|A'| + \left(\frac{k-1}{k} - \frac{\sqrt{k} \sqrt[4k]{\epsilon}}{2^{k-1}}\right) |B| \geq |A| - \sqrt{\epsilon} n + \left(\frac{k-1}{k} - \frac{\sqrt{k} \sqrt[4k]{\epsilon}}{2^{k-1}}\right) |B|.$$

□

## 2.7 Concluding Remarks

The most satisfactory results of this chapter are for  $k$ -colored or  $(k+1)$ -colored  $k$ -uniform hypergraphs. In order to obtain them, we extended Theorem 2.1.1 to almost complete hypergraphs. However, for complete hypergraphs more is known.

**Theorem 2.7.1** (Füredi and Gyárfás, 1991, [36]) *Let  $k, r \geq 2$  and let  $q$  be the smallest integer such that  $r \leq q^{k-1} + q^{k-2} + \dots + q + 1$ . Then  $\text{mc}_r(K_n^{(k)}) \geq \frac{n}{q}$ . This is sharp when  $q^k$  divides  $n$ ,  $r = q^{k-1} + q^{k-2} + \dots + q + 1$ , and an affine space of dimension  $k$  and order  $q$  exists.*

**Theorem 2.7.2** (Gyárfás and Haxell, 2009, [46])  *$\text{mc}_5(K_n^{(3)}) \geq 5n/7$  and  $\text{mc}_6(K_n^{(3)}) \geq 2n/3$ . Furthermore these are tight when  $n$  is divisible by 7 and 6 respectively.*

It would be interesting to extend these results to the nearly complete setting, from which

the random analogs would then follow by applying Theorem 2.1.4. It is worth noting that these results follow by applying Füredi's fractional transversal method which does not seem to extend to non-complete graphs in a straightforward manner.

# Chapter 3

## Rainbow Hamilton Cycles in Random Hypergraphs

### 3.1 Introduction

The results of this chapter are joint work with Andrzej Dudek and Alan Frieze [17].

Suppose that  $k > \ell \geq 1$ . An  $\ell$ -Hamilton cycle  $C$  in a  $k$ -uniform hypergraph  $H = (V, \mathcal{E})$  on  $n$  vertices is a collection of  $m_\ell = n/(k - \ell)$  edges of  $H$  such that for some cyclic order of  $[n]$  every edge consists of  $k$  consecutive vertices and for every pair of consecutive edges  $E_{i-1}, E_i$  in  $C$  (in the natural ordering of the edges) we have  $|E_{i-1} \cap E_i| = \ell$  (see Figure 3.1). Thus, in every  $\ell$ -Hamilton cycle the sets  $C_i = E_i \setminus E_{i-1}$ ,  $i = 1, 2, \dots, m_\ell$ , are a partition of  $V$  into sets of size  $k - \ell$ . Hence,  $m_\ell = n/(k - \ell)$ . We thus always assume, when discussing  $\ell$ -Hamilton cycles, that this necessary condition,  $k - \ell$  divides  $n$ , is fulfilled. In the literature, when  $\ell = k - 1$  we have a *tight* Hamilton cycle and when  $\ell = 1$  we have a *loose* Hamilton cycle, analogous to the definition of loose cycle given in Chapter 2.

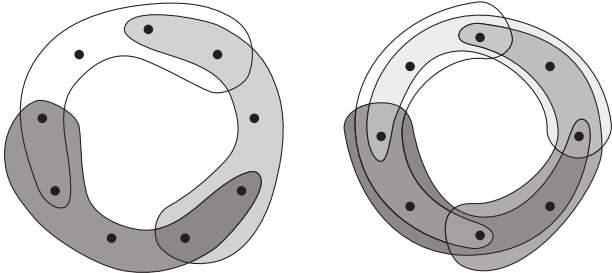


Figure 3.1: A 2-Hamilton and a 3-Hamilton 5-uniform cycle.

The threshold for the existence of Hamilton cycles in the random graph  $\mathcal{G}(n, p)$  has been known for many years, see, e.g., [2], [9] and [58]. Recently these results were extended to hypergraphs, see, e.g., [3, 19, 21, 20, 28, 31, 32, 42, 64, 67]. Below we summarize some of them.

In the following and throughout the chapter,  $\omega = \omega(n)$  can be any function tending to infinity with  $n$ . All logarithms in this dissertation will be base  $e$ , unless otherwise specified.

**Theorem 3.1.1** (Dudek and Frieze, 2013, [19]) *Let  $\varepsilon > 0$  be fixed. Then:*

- (i) *For all integers  $k > \ell \geq 2$ , if  $p \leq (1 - \varepsilon)e^{k-\ell}/n^{k-\ell}$ , then a.a.s.  $\mathcal{H}^{(k)}(n, p)$  is not  $\ell$ -Hamiltonian .*
- (ii) *For all integers  $k > \ell \geq 3$ , there exists a constant  $K = K(k)$  such that if  $p \geq K/n^{k-\ell}$  and  $n$  is a multiple of  $k - \ell$  then a.a.s.  $\mathcal{H}^{(k)}(n, p)$  is  $\ell$ -Hamiltonian.*
- (iii) *If  $k > \ell = 2$  and  $p \geq \omega/n^{k-2}$  and  $n$  is a multiple of  $k - 2$ , then a.a.s.  $\mathcal{H}^{(k)}(n, p)$  is 2-Hamiltonian.*
- (iv) *For all  $k \geq 4$ , if  $p \geq (1 + \varepsilon)e/n$ , then  $\mathcal{H}^{(k)}(n, p)$  is  $(k - 1)$ -Hamiltonian a.a.s., i.e. it contains a tight Hamilton cycle.*

In particular, this theorem shows that  $e/n$  is the sharp threshold for the existence of a tight Hamilton cycle in a  $k$ -uniform hypergraph, when  $k \geq 4$ . As it was explained in [19], quite surprisingly, the proof of (ii)-(iv) in Theorem 3.1.1 is based on the second moment method (see Appendix A).

**Theorem 3.1.2** ([21, 20, 28, 31]) *Fix  $k \geq 3$  and suppose that  $n$  is a multiple of  $k - 1$ . Let  $p \geq \omega(\log n)/n^{k-1}$ . Then, a.a.s.  $\mathcal{H}^{(k)}(n, p)$  contains a loose Hamilton cycle.*

Thus,  $(\log n)/n^{k-1}$  is the asymptotic threshold for the existence of loose Hamilton cycles. This is because if  $p \leq (1 - \varepsilon)(k - 1)!(\log n)/n^{k-1}$  and  $\varepsilon > 0$  is constant, then a.a.s.  $\mathcal{H}^{(k)}(n, p)$  contains isolated vertices.

In this chapter we study the existence of rainbow Hamilton cycles in the random graph  $\mathcal{H}^{(k)}(n, p)$  with independently colored edges. Let  $\mathcal{H}^{(k)}(n, p, \kappa)$  denote a *randomly colored random hypergraph*, constructed on the vertex set  $[n]$  by taking each  $k$ -tuple independently with probability  $p$ , and then independently coloring it with a random color from the set  $[\kappa]$ . We also denote  $\mathcal{H}^{(2)}(n, p, \kappa)$  by  $\mathcal{G}(n, p, \kappa)$ . Rainbow properties of  $\mathcal{G}(n, p, \kappa)$  attracted a considerable amount of attention, see, e.g., [5, 15, 29, 33, 30].

Here we only focus on rainbow Hamilton cycles, which are Hamilton cycles where every edge of the cycle receives a different color. Improving the previous results of Cooper and Frieze [15] and Frieze and Loh [33], Ferber and Krivelevich [29] determined the very sharp threshold for the existence of the rainbow Hamilton cycle in  $\mathcal{G}(n, p, \kappa)$  assuming nearly optimal number of colors.

**Theorem 3.1.3** (Ferber and Krivelevich, 2016, [29]) *Let  $\varepsilon > 0$ ,  $\kappa = (1 + \varepsilon)n$  and let  $p = (\log n + \log \log n + \omega)/n$ . Then, a.a.s.  $G_{n,p,\kappa}$  contains a rainbow Hamilton cycle.*

For expressions such as  $\kappa = (1 + \varepsilon)n$  that clearly have to be an integer, we round up or down but do not specify which: the reader can choose either one, without affecting the argument.

Ferber and Krivelevich [29] were the first to study rainbow Hamilton cycles in  $\mathcal{H}^{(k)}(n, p)$ . They showed the following. (Recall that  $m_\ell = n/(k - \ell)$  is the number of edges in an  $\ell$ -Hamilton cycle.)

**Theorem 3.1.4** (Ferber and Krivelevich, 2016, [29]) *Let  $k > \ell \geq 1$  be integers. Suppose that  $n$  is a multiple of  $k - \ell$ . Let  $p \in [0, 1]$  be such that a.a.s.  $\mathcal{H}^{(k)}(n, p)$  contains an  $\ell$ -Hamilton cycle. Then, for every  $\varepsilon = \varepsilon(n) \geq 0$ , letting  $\kappa = (1 + \varepsilon)m_\ell$  and  $q = \kappa p / (\varepsilon m_\ell + 1)$  we have that a.a.s.  $\mathcal{H}^{(k)}(n, p, \kappa)$  contains a rainbow  $\ell$ -Hamilton cycle.*

Observe that if  $\varepsilon$  is a constant, then by losing a multiplicative constant in the threshold, a.a.s. a rainbow  $\ell$ -Hamilton exists. By combining this result with Theorems 3.1.1 and 3.1.2 one can obtain some explicit values of  $q$ . However, for small  $\varepsilon$  (including  $\varepsilon = 0$ ) Theorem 3.1.4

does not provide optimal  $q$ . In our results we focus on the case when  $\kappa = m_\ell$ . (But we also allow more colors.) Here we state our first result.

**Theorem 3.1.5** *Let  $k > \ell \geq 2$  and  $\varepsilon > 0$  be fixed. Let  $c \geq 1/(k - \ell)$  and  $\kappa = cn$ . Then:*

(i) *For all integers  $k > \ell \geq 2$ , if*

$$p \leq \begin{cases} (1 - \varepsilon)e^{k-\ell+1}/n^{k-\ell} & \text{if } c = 1/(k - \ell) \\ (1 - \varepsilon) \left(\frac{c-1/(k-\ell)}{c}\right)^{(k-\ell)c-1} e^{k-\ell+1}/n^{k-\ell} & \text{if } c > 1/(k - \ell), \end{cases}$$

*then a.a.s.  $\mathcal{H}^{(k)}(n, p, \kappa)$  is not rainbow  $\ell$ -Hamiltonian.*

(ii) *For all integers  $k > \ell \geq 3$ , there exists a constant  $K = K(k)$  such that if  $p \geq K/n^{k-\ell}$  and  $n$  is a multiple of  $k - \ell$  then a.a.s.  $\mathcal{H}^{(k)}(n, p, \kappa)$  is rainbow  $\ell$ -Hamiltonian.*

(iii) *If  $k > \ell = 2$  and  $p \geq \omega/n^{k-2}$  and  $n$  is a multiple of  $k - 2$ , then a.a.s.  $\mathcal{H}^{(k)}(n, p, \kappa)$  is rainbow 2-Hamiltonian.*

(iv) *For all  $k \geq 4$ , if*

$$p \geq \begin{cases} (1 + \varepsilon)e^2/n & \text{if } c = 1 \\ (1 + \varepsilon) \left(\frac{c-1}{c}\right)^{c-1} e^2/n & \text{if } c > 1, \end{cases}$$

*then a.a.s.  $\mathcal{H}^{(k)}(n, p, \kappa)$  is rainbow  $(k-1)$ -Hamiltonian, i.e. it contains a rainbow tight Hamilton cycle.*

Consequently, if  $k \geq 4$ , then

$$p = \begin{cases} e^2/n & \text{if } c = 1 \\ \left(\frac{c-1}{c}\right)^{c-1} e^2/n & \text{if } c > 1 \end{cases}$$



is the sharp threshold for the existence of a rainbow tight Hamilton cycle. Furthermore, observe that  $\lim_{c \rightarrow 1^+} \left(\frac{c-1}{c}\right)^{c-1} = 1$ . Thus, in (iv) the case  $c > 1$  approaches the case  $c = 1$  in the continuous way. Finally, also observe that  $\lim_{c \rightarrow \infty} \left(\frac{c-1}{c}\right)^{c-1} = 1/e$ . Hence, when  $c$  tends to infinity (that means that each edge receives a different color) the threshold function is  $e/n$ , which is consistent with Theorem 3.1.1. The proof of Theorem 3.1.5 modifies the proof of Theorem 3.1.1.

We also establish a similar result for loose Hamilton cycles. Recall that a loose Hamilton cycle of order  $n$  has exactly  $n/(k-1)$  edges. So for a rainbow loose Hamilton cycle we always need at least  $n/(k-1)$  colors. Here we only consider this most restrictive case with  $\kappa = n/(k-1)$ .

**Theorem 3.1.6** *Fix  $k \geq 3$  and suppose that  $n$  is a multiple of  $k-1$ . Let  $\kappa = n/(k-1)$  and  $p \geq \omega(\log n)/n^{k-1}$ . Then, a.a.s.  $\mathcal{H}^{(k)}(n, p, \kappa)$  contains a rainbow loose Hamilton cycle.*

The proof is a modification of the proof of Theorem 3.1.2.

**Some notation:** For sequences  $A_n, B_n, n \geq 1$  we write  $A_n \approx B_n$  to mean that  $A_n = (1 + o(1))B_n$  as  $n \rightarrow \infty$ . Similarly, we write  $A_n \lesssim B_n$  to mean that  $A_n \leq (1 + o(1))B_n$  as  $n \rightarrow \infty$ .

## 3.2 $\ell$ -Hamilton Cycles for $\ell \geq 2$

We now prove Theorem 3.1.5 The proof modifies the proof of Theorem 3 from [19].

Let  $([n], \mathcal{E})$  be a  $k$ -uniform hypergraph. A permutation  $\pi$  of  $[n]$  is  $\ell$ -Hamilton cycle inducing if

$$E_\pi(i) = \{\pi((i-1)(k-\ell) + j) : j \in [k]\} \in \mathcal{E} \text{ for all } i \in [n/(k-\ell)].$$

(We use the convention  $\pi(n+r) = \pi(r)$  for  $r > 0$ .) Let the term *hamperm* refer to such a

permutation.

Let  $Y$  be the random variable that counts the number of rainbow hamperms  $\pi$  for  $\mathcal{H}^{(k)}(n, p, \kappa)$ . Every  $\ell$ -Hamilton cycle induces at least one hamperm and so we can concentrate on estimating  $\Pr(Y > 0)$ .

Observe that

$$\mathbf{E}(Y) = n! \cdot p^{n/(k-\ell)} \cdot \frac{(\kappa)_{n/(k-\ell)}}{\kappa^{n/(k-\ell)}},$$

where  $(x)_t = x(x-1)\cdots(x-t+1)$  is the *falling factorial*. This is because  $\pi$  induces an  $\ell$ -Hamilton cycle if and only if a certain  $n/(k-\ell)$  edges are present and are colored rainbow.

Now let  $c > 1/(k-\ell)$ . Then, by Stirling's formula (A.6) we get

$$\begin{aligned} \mathbf{E}(Y) &= n! p^{n/(k-\ell)} \frac{\kappa!}{\kappa^{n/(k-\ell)} (\kappa - n/(k-\ell))!} \\ &\approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n p^{n/(k-\ell)} \frac{\sqrt{\frac{\kappa}{\kappa - n/(k-\ell)}} \left(\frac{\kappa}{e}\right)^\kappa}{\kappa^{n/(k-\ell)} \left(\frac{\kappa - n/(k-\ell)}{e}\right)^{\kappa - n/(k-\ell)}} \\ &= \sqrt{\frac{2\pi n \kappa}{\kappa - n/(k-\ell)}} \left(\frac{np^{1/(k-\ell)}}{e^{1+1/(k-\ell)}} \cdot \left(\frac{\kappa}{\kappa - n/(k-\ell)}\right)^{\kappa/n - 1/(k-\ell)}\right)^n \\ &= \sqrt{\frac{2\pi n \kappa}{\kappa - n/(k-\ell)}} \left(\frac{np^{1/(k-\ell)}}{e^{1+1/(k-\ell)}} \cdot \left(\frac{c}{c - 1/(k-\ell)}\right)^{c-1/(k-\ell)}\right)^n. \end{aligned}$$

Similarly for  $c = 1/(k-\ell)$  we get

$$\mathbf{E}(Y) \approx 2\pi n \sqrt{\frac{1}{k-\ell}} \left(\frac{np^{1/(k-\ell)}}{e^{1+1/(k-\ell)}}\right)^n.$$

Thus, if

$$p \leq \begin{cases} (1-\varepsilon)e^{k-\ell+1}/n^{k-\ell} & \text{if } c = 1/(k-\ell) \\ (1-\varepsilon) \left(\frac{c-1/(k-\ell)}{c}\right)^{(k-\ell)c-1} e^{k-\ell+1}/n^{k-\ell} & \text{if } c > 1/(k-\ell), \end{cases}$$

then  $\mathbf{E}(Y) = o(1)$ . By the first moment method (see Appendix A), this verifies part (i).

Now we prove parts (ii)-(iv) by the second moment method. First observe that if

$$p \geq \begin{cases} (1 + \varepsilon)e^{k-\ell+1}/n^{k-\ell} & \text{if } c = 1/(k - \ell) \\ (1 + \varepsilon) \left( \frac{c-1/(k-\ell)}{c} \right)^{(k-\ell)c-1} e^{k-\ell+1}/n^{k-\ell} & \text{if } c > 1/(k - \ell), \end{cases}$$

then  $\mathbf{E}(Y) \rightarrow \infty$  together with  $n$ .

Fix a hamperm  $\pi$ . Let  $H(\pi) = (E_\pi(1), E_\pi(2), \dots, E_\pi(m_\ell))$  be the Hamilton cycle induced by  $\pi$ . Then let  $N(b, a)$  be the number of permutations  $\pi'$  such that  $|E(H(\pi)) \cap E(H(\pi'))| = b$  and  $E(H(\pi)) \cap E(H(\pi'))$  consists of  $a$  edge disjoint paths. Here a path is a maximal sub-sequence  $F_1, F_2, \dots, F_q$  of the edges of  $H(\pi)$  such that  $F_i \cap F_{i+1} \neq \emptyset$  for  $1 \leq i < q$ . The set  $\bigcup_{j=1}^q F_j$  may contain other edges of  $H(\pi)$ . Observe that  $N(b, a)$  does not depend on  $\pi$ .

Now,

$$\begin{aligned} \frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} &\leq \frac{n!N(0, 0)p^{2n/(k-\ell)} \left( \frac{(\kappa)_{n/(k-\ell)}}{\kappa^{n/(k-\ell)}} \right)^2}{\mathbf{E}(Y)^2} \\ &\quad + \sum_{b=1}^{n/(k-\ell)} \sum_{a=1}^b \frac{n!N(b, a)p^{2n/(k-\ell)-b}}{\mathbf{E}(Y)^2} \cdot \frac{(\kappa)_{n/(k-\ell)}}{\kappa^{n/(k-\ell)}} \cdot \frac{(\kappa - b)_{n/(k-\ell)-b}}{\kappa^{n/(k-\ell)-b}}. \end{aligned}$$

Since trivially  $N(0, 0) \leq n!$ , we get

$$\frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} \leq 1 + \sum_{b=1}^{n/(k-\ell)} \sum_{a=1}^b \frac{n!N(b, a)p^{2n/(k-\ell)-b}}{\mathbf{E}(Y)^2} \cdot \frac{(\kappa)_{n/(k-\ell)}}{\kappa^{n/(k-\ell)}} \cdot \frac{(\kappa - b)_{n/(k-\ell)-b}}{\kappa^{n/(k-\ell)-b}}.$$

Let  $X$  be the number of  $\ell$ -hamperms in  $\mathcal{H}^{(k)}(n, p)$ . Then,

$$\mathbf{E}(X) = n!p^{n/(k-\ell)} \quad \text{and} \quad \mathbf{E}(Y) = \mathbf{E}(X) \cdot \frac{(\kappa)_{n/(k-\ell)}}{\kappa^{n/(k-\ell)}}.$$

Consequently,

$$\begin{aligned}
\frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} &\leq 1 + \sum_{b=1}^{n/(k-\ell)} \sum_{a=1}^b \frac{n!N(b,a)p^{2n/(k-\ell)-b}}{\mathbf{E}(X)^2} \cdot \frac{(\kappa)_{n/(k-\ell)}}{\kappa^{n/(k-\ell)}} \cdot \frac{(\kappa-b)_{n/(k-\ell)-b}}{\kappa^{n/(k-\ell)-b}} \cdot \left( \frac{\kappa^{n/(k-\ell)}}{(\kappa)_{n/(k-\ell)}} \right)^2 \\
&= 1 + \sum_{b=1}^{n/(k-\ell)} \sum_{a=1}^b \frac{N(b,a)p^{n/(k-\ell)-b}}{\mathbf{E}(X)} \cdot \kappa^b \cdot \frac{(\kappa-b)_{n/(k-\ell)-b}}{(\kappa)_{n/(k-\ell)}} \\
&= 1 + \sum_{b=1}^{n/(k-\ell)} \sum_{a=1}^b \frac{N(b,a)p^{n/(k-\ell)-b}}{\mathbf{E}(X)} \cdot \kappa^b \cdot \frac{(\kappa-b)!}{\kappa!} \\
&\leq 1 + \sum_{b=1}^{n/(k-\ell)} \sum_{a=1}^b \frac{N(b,a)p^{n/(k-\ell)-b}}{\mathbf{E}(X)} \cdot e^b \left( \frac{\kappa-b}{\kappa} \right)^{\kappa-b}. \tag{3.1}
\end{aligned}$$

**Part (ii):**  $\ell \geq 3$

We trivially bound  $\left(\frac{\kappa-b}{\kappa}\right)^{\kappa-b} \leq 1$ . It was shown in [19] (equation (10)) that

$$\frac{N(b,a)p^{n/(k-\ell)-b}}{\mathbf{E}(X)} \lesssim \left( \frac{2k!ke^k}{n^{k-\ell}p} \right)^b \frac{1}{n^{a(\ell-2)}}.$$

Thus,

$$\begin{aligned}
\frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} &\leq 1 + \sum_{b=1}^{n/(k-\ell)} \sum_{a=1}^b \frac{N(b,a)p^{n/(k-\ell)-b}}{\mathbf{E}(X)} \cdot e^b \\
&\lesssim 1 + \sum_{b=1}^{n/(k-\ell)} \sum_{a=1}^b \left( \frac{2k!ke^k}{n^{k-\ell}p} \right)^b \frac{1}{n^{a(\ell-2)}} \cdot e^b \\
&\leq 1 + \sum_{b=1}^{n/(k-\ell)} \sum_{a=1}^b \left( \frac{2k!ke^{k+1}}{n^{k-\ell}p} \right)^b \frac{1}{n^a}.
\end{aligned}$$

Set  $K = 4k!ke^{k+1}$  and  $p = K/n^{k-\ell}$ . Thus,

$$\frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} \leq 1 + \sum_{b=1}^{n/(k-\ell)} \sum_{a=1}^b \frac{1}{2^b} \cdot \frac{1}{n^a} \leq 1 + \left( \sum_{b=1}^n \frac{1}{2^b} \right) \left( \sum_{a=1}^n \frac{1}{n^a} \right) \approx 1.$$

**Part (iii):**  $\ell = 2$

Let  $p \geq \omega/n^{k-2}$ . Similarly as in the previous case

$$\begin{aligned} \frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} &\leq 1 + \sum_{b=1}^{n/(k-2)} \sum_{a=1}^b \left( \frac{2k!ke^k}{n^{k-2}p} \right)^b \cdot e^b \\ &\leq 1 + \sum_{b=1}^{n/(k-2)} \sum_{a=1}^b \left( \frac{2k!ke^{k+1}}{\omega} \right)^b \\ &\leq 1 + \sum_{b=1}^{n/(k-2)} b \left( \frac{2k!ke^{k+1}}{\omega} \right)^b \approx 1. \end{aligned}$$

**Part (iv):**  $\ell = k - 1$  (tight cycles)

If  $c = 1$  (that means  $\kappa = n$ ), then we trivially bound  $\left(\frac{\kappa-b}{\kappa}\right)^{\kappa-b} \leq 1$ . Otherwise, we use a simple fact.

**Claim 3.2.1** *Let  $\kappa = cn$ , where  $c > 1$ . Then,*

$$\max_{0 < b \leq n} \left( \frac{\kappa - b}{\kappa} \right)^{\frac{\kappa-b}{b}} = \left( \frac{\kappa - n}{\kappa} \right)^{\frac{\kappa-n}{n}} = \left( \frac{c-1}{c} \right)^{c-1}.$$

*Proof of the claim.* Let  $x = b/n$ . Note that since  $1 \leq b \leq n$ ,  $x \in (0, 1]$  and  $c > x$ . Then

$$\left( \frac{\kappa - b}{\kappa} \right)^{\frac{\kappa-b}{b}} = \left( \frac{c - b/n}{c} \right)^{\frac{c-b/n}{b/n}} = \left( \frac{c-x}{c} \right)^{\frac{c-x}{x}}.$$

Taking the derivative gives us

$$\frac{d}{dx} \left( \frac{c-x}{c} \right)^{\frac{c-x}{x}} = -\frac{c}{x^2} \left( \frac{c-x}{c} \right)^{\frac{c-x}{x}} \left( \log \left( \frac{c-x}{c} \right) + \frac{x}{c} \right).$$

Since  $\frac{c}{x^2} \left(\frac{c-x}{c}\right)^{\frac{c-x}{x}} > 0$ , we have

$$\operatorname{sgn} \left( \frac{d}{dx} \left( \frac{c-x}{c} \right)^{\frac{c-x}{x}} \right) = -\operatorname{sgn} \left( \log \left( \frac{c-x}{c} \right) + \frac{x}{c} \right) = -\operatorname{sgn} \left( \log \left( 1 - \frac{x}{c} \right) + \frac{x}{c} \right)$$

and since  $\log \left( 1 - \frac{x}{c} \right) < \log e^{-\frac{x}{c}} = -\frac{x}{c}$  we get  $\log \left( 1 - \frac{x}{c} \right) + \frac{x}{c} < 0$ . Thus

$$\frac{d}{dx} \left( \frac{c-x}{c} \right)^{\frac{c-x}{x}} > 0$$

for  $0 < x \leq 1$  and  $c > x$ . Thus  $\left(\frac{c-x}{c}\right)^{\frac{c-x}{x}}$  is maximized at  $x = 1$  in our domain, which corresponds to  $b = n$ , proving the claim.  $\square$

Due to (3.1) and the above claim we obtain

$$\frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} \leq \begin{cases} 1 + \sum_{b=1}^n \sum_{a=1}^b \frac{N(b,a)p^{n-b}}{\mathbf{E}(X)} \cdot e^b, & \text{if } c = 1 \\ 1 + \sum_{b=1}^n \sum_{a=1}^b \frac{N(b,a)p^{n-b}}{\mathbf{E}(X)} \cdot \left( e \left( \frac{c-1}{c} \right)^{c-1} \right)^b, & \text{if } c > 1. \end{cases}$$

Moreover, it was shown in [19] (equation (13)) that for  $k \geq 4$ ,

$$\sum_{b=1}^n \sum_{a=1}^b \frac{N(b,a)p^{n-b}}{\mathbf{E}(X)} \leq \frac{2c_k k! e^{k-1}}{n^{k-3}} \exp \left\{ \frac{2k! e^{k-1}}{n^{k-4}} \right\} \sum_{b=1}^n \left( \frac{e}{np} \right)^b$$

for some positive constant  $c_k$  that depends on  $k$  only. Thus,

$$\frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} \leq \begin{cases} 1 + \frac{2c_k k! e^{k-1}}{n^{k-3}} \exp \left\{ \frac{2k! e^{k-1}}{n^{k-4}} \right\} \sum_{b=1}^n \left( \frac{e}{np} \right)^b \cdot e^b, & \text{if } c = 1 \\ 1 + \frac{2c_k k! e^{k-1}}{n^{k-3}} \exp \left\{ \frac{2k! e^{k-1}}{n^{k-4}} \right\} \sum_{b=1}^n \left( \frac{e}{np} \right)^b \cdot \left( e \left( \frac{c-1}{c} \right)^{c-1} \right)^b, & \text{if } c > 1. \end{cases}$$

Hence, both for  $c = 1$ ,  $p \geq \frac{(1+\varepsilon)e^2}{n}$  and for  $c > 1$ ,  $p \geq (1 + \varepsilon) \left(\frac{c-1}{c}\right)^{c-1} \frac{e^2}{n}$ , we get that

$$\frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} \leq 1 + \frac{2c_k k! e^{k-1}}{n^{k-3}} \exp\left\{\frac{2k! e^{k-1}}{n^{k-4}}\right\} \sum_{b=1}^n \frac{1}{(1+\varepsilon)^b} \approx 1.$$

In all three cases we showed that  $\frac{\mathbf{E}(Y^2)}{\mathbf{E}(Y)^2} \lesssim 1$ . Thus, by the second moment method, this completes the proof of Theorem 3.1.5.

### 3.3 Loose Hamilton Cycles

We now prove Theorem 3.1.6

Let  $n = (k-1)m$  and assume that  $m$  is even. Clearly,  $m = m_1 = \kappa$ . In this case the proof is a straightforward modification of the proof of Theorem 2 from [21].

Let  $X = [m]$  and  $Y = [m+1, n]$  and  $Z = [n+1, n+m]$ . Given  $H = \mathcal{H}^{(k)}(n, p, m)$  we define the  $(k+1)$ -uniform hypergraph  $\Gamma$  with vertex set  $[n]$  and an edge  $\phi(e)$  for each edge  $e = \{x_1, x_2, y_1, \dots, y_{k-2}\}$  of  $H$  that satisfies  $|e \cap X| = 2$ . Here  $x_1, x_2 \in X$  and  $y_i \in Y$ ,  $1 \leq i \leq k-2$ . We then let  $\phi(e) = \{x_1, x_2, y_1, \dots, y_{k-2}, c(e) + n\}$ , where  $c(e)$  is the color of  $e$  and  $c(e) + n \in Z$ . The proof in [21] can be adapted (and therefore we need to assume that  $m$  is even) to show that a.a.s.  $\Gamma$  contains a loose Hamilton cycle where consecutive edges intersect in vertices of  $X$ . We will provide some details of the adaptation of the proof now.

Suppose that  $p = \omega(\log n)/n^{k-1}$ , where  $\omega = o(\log n)$  and  $\omega \rightarrow \infty$ . Let  $M = \binom{n}{k} p$  and

consider a random  $(k+1)$ -uniform hypergraph  $K$  with approximately  $M' \approx M$  edges. Then

$$\begin{aligned}
\Pr(\exists e_1, e_2 \in E(K) : |e_1 \cap e_2| = k) &\leq \binom{n}{k+1} \binom{k+1}{k} n \frac{\binom{n}{M'-2}^{(k+1)-2}}{\binom{n}{M'}} \\
&\leq n^{k+2} \left( \frac{M'}{\binom{n}{k+1}} \right)^2 \\
&\leq n^{k+1} \left( \frac{2(k+1)! \omega n \log n}{n^{k+1}} \right)^2 \\
&= o(1).
\end{aligned}$$

In this way we can justify viewing  $\mathcal{H}^{(k)}(n, p, m)$  as a random  $(k+1)$ -uniform hypergraph. This would be a problem if the latter model gave an edge more than one color.

Let  $m = 2m_1$ . Then  $m_1$  will replace  $m$  in the proof in [21]. The proof in [21] involves proving that a.a.s.  $\mathcal{H}^{(k)}(n, p)$  contains a loose Hamilton cycle that respects a certain vertex partition. Such a Hamilton cycle will consist of  $2m_1$  edges of the form  $\{x_i, x_{i+1}, y_{i,1}, \dots, y_{i,\kappa}\}$ , where  $\kappa = k-2$ ,  $1 \leq i \leq 2m_1$ ,  $x_{2m_1+1} = x_1$ ,  $\{x_1, \dots, x_{2m_1}\} = X$  and  $\{y_{1,1}, \dots, y_{2m_1,\kappa}\} = Y$ . This is done as follows: we choose a large positive integer  $d$ . Let  $\mathcal{X}$  be a set of size  $2dm_1$  representing  $d$  copies of each  $x \in X$ . Denote the  $j$ th copy of  $x \in X$  by  $x^{(j)} \in \mathcal{X}$  and let  $\mathcal{X}_x = \{x^{(j)}, j = 1, 2, \dots, d\}$ . Then let  $X_1, X_2, \dots, X_d$  be a uniform random partition of  $\mathcal{X}$  into  $d$  sets of size  $2m_1$ . Define  $\psi_1 : \mathcal{X} \rightarrow X$  by  $\psi_1(x^{(j)}) = x$  for all  $j$  and  $x \in X$ . Similarly, we let  $\mathcal{Y}$  be a set of size  $d\kappa m_1$  representing  $d/2$  copies of each  $y \in Y$ . Denote the  $j$ th copy of  $y \in Y$  by  $y^{(j)} \in \mathcal{Y}$  and let  $\mathcal{Y}_y = \{y^{(j)}, j = 1, 2, \dots, d/2\}$ . Then let  $Y_1, Y_2, \dots, Y_d$  be a uniform random partition of  $\mathcal{Y}$  into  $d$  sets of size  $\kappa m_1$ . Define  $\psi_2 : \mathcal{Y} \rightarrow Y$  by  $\psi_2(y^{(j)}) = y$  for all  $y \in Y$ . Finally, let  $\psi : \binom{\mathcal{X}}{2} \times \binom{\mathcal{Y}}{\kappa} \rightarrow X^2 \times Y^\kappa$  be such that  $\psi(\nu_1, \nu_2, \xi_1, \xi_2, \dots, \xi_\kappa) = (\psi_1(\nu_1), \psi_1(\nu_2), \psi_2(\xi_1), \psi_2(\xi_2), \dots, \psi_2(\xi_\kappa))$ .

All we need do is add a set  $\mathcal{Z}$  of size  $dm_1$  representing  $d/2$  copies of each  $z \in Z$ . We denote the  $j$ th copy of  $z \in Z$  by  $z^{(j)} \in \mathcal{Z}$  and let  $\mathcal{Z}_z = \{z^{(j)}, j = 1, 2, \dots, d/2\}$ . Then let



$Z_1, Z_2, \dots, Z_d$  be a uniform random partition of  $\mathcal{Z}$  into  $d$  sets of size  $m_1$ . Define  $\psi_3 : \mathcal{Z} \rightarrow Z$  by  $\psi_3(z^{(j)}) = z$  for all  $z \in Z$ . We then modify  $\psi$  so that  $\psi : \binom{\mathcal{X}}{2} \times \binom{\mathcal{Y}}{\kappa} \times \mathcal{Z} \rightarrow X^2 \times Y^\kappa \times Z$  be such that  $\psi(\nu_1, \nu_2, \xi_1, \xi_2, \dots, \xi_\kappa, \zeta) = (\psi_1(\nu_1), \psi_1(\nu_2), \psi_2(\xi_1), \psi_2(\xi_2), \dots, \psi_2(\xi_\kappa), \psi_3(\zeta))$ . After this the proof in [21] can be carried out with straightforward modifications involving adding a component for members of  $Z$ .

For the proof in [21], we use the fact that  $m$  is even. We can easily remove this requirement by using an idea of Ferber [28]. In particular, one can follow his proof of Theorem 1.2 to show that  $\Gamma$  contains a loose Hamilton cycle in this case.

### 3.4 Concluding Remarks

The results of Theorem 3.1.5 are in some sense more satisfactory than the result in Theorem 3.1.6. For example, we establish the sharp threshold for rainbow tight cycles with any number of colors.

On the other hand, there is still much work to be done on loose Hamilton cycles. The sharp threshold for the existence of loose Hamilton cycles in  $\mathcal{H}^{(k)}(n, p)$  is not known, so this problem would need to be solved before these results could be extended to  $\mathcal{H}^{(k)}(n, p, \kappa)$ . In addition, we only consider loose rainbow Hamilton cycles here where the number of colors used is exactly the number we need to have a rainbow Hamilton cycle. It would be interesting to investigate this case when the number of colors,  $\kappa > n/(k - 1)$ .

# Chapter 4

## Saturation of Berge Hypergraphs

### 4.1 Introduction

The results of this chapter are joint work with Nathan Graber, Pamela Kirkpatrick, Abhishek Methuku, and Eric C. Sullivan [25].

Given simple graphs  $G$  and  $F$ , we say  $G$  is  $F$ -saturated if  $G$  does not contain  $F$  as a subgraph, but  $G + e$  contains  $F$  as a subgraph for each  $e \in E(\overline{G})$ . The maximum possible number of edges in a graph  $G$  on  $n$  vertices that is  $F$ -saturated is known as the *Turán number* or the extremal number of  $F$ , and is denoted  $\text{ex}(n, F)$ . In 1907, Mantel proved one of the first results on extremal numbers, finding them for triangles [62]. Mantel's result was generalized for all complete graphs in 1941 by Turán [77].

On the other end of the spectrum, the minimum number of edges of an  $F$ -saturated graph on  $n$  vertices is known as the *saturation number* of  $F$  and is denoted  $\text{sat}(n, F)$ . Saturation numbers in graphs were first studied by Erdős, Hajnal and Moon in [51], where they determined  $\text{sat}(n, K_m)$  and the  $K_m$ -saturated graphs that achieve that saturation number. In [56], Kászonyi and Tuza determine the saturation number for paths, stars and matchings, and provide a general upper bound.

**Theorem 4.1.1** (Kászonyi and Tuza, 1986, [56])

1. Let

$$a_k = \begin{cases} 3 \cdot 2^{t-1} - 2 & \text{if } k = 2t \\ 4 \cdot 2^{t-1} - 2 & \text{if } k = 2t + 1 \end{cases}.$$

If  $n \geq a_k$  and  $k \geq 6$ , then  $\text{sat}(n, P_k) = n - \left\lfloor \frac{n}{a_k} \right\rfloor$ .

2. Let  $S_t = K_{1,t-1}$  denote a star on  $t$  vertices. Then,

$$\text{sat}(n, S_t) = \begin{cases} \binom{t-1}{2} + \binom{n-t+1}{2} & \text{if } t \leq n \leq \frac{3t-3}{2} \\ \left\lceil \frac{(t-2)n}{2} - \frac{(t-1)^2}{8} \right\rceil & \text{if } \frac{3t-3}{2} \leq n \end{cases}.$$

3. For  $n \geq 3t - 3$ ,  $\text{sat}(n, tK_2) = 3t - 3$ .

The concept of saturation has been extended to hypergraphs. Given a  $k$ -uniform hypergraph  $F$ , one can define  $F$ -saturated and  $\text{sat}_k(n, F)$  analogously to the graph case. In this chapter, we will study the  $k$ -uniform saturation numbers for Berge hypergraphs.

The classical definition of a hypergraph cycle due to Berge is the following. A Berge cycle of length  $k$  is an alternating sequence of distinct vertices and hyperedges of the form  $v_1, h_1, v_2, h_2, \dots, v_k, h_k, v_1$  where  $v_i, v_{i+1} \in h_i$  for each  $i \in \{1, 2, \dots, k-1\}$  and  $v_k, v_1 \in h_k$  and is denoted  $\text{Berge-}C_k$ . Note that Berge paths, defined in Chapter 2 are defined analogously. In this chapter, whenever we refer to a path in a hypergraph, unless otherwise stated, this will refer to a Berge path.

Gerbner and Palmer [40] gave the following natural generalization of the definitions of Berge cycles and Berge paths. Let  $F$  be a graph and  $H$  a hypergraph. We say  $H$  is a *Berge- $F$*  if there is a bijection  $\phi : E(F) \rightarrow E(H)$  such that  $e \subseteq \phi(e)$  for all  $e \in E(F)$ . This can be thought of as expanding each edge of  $F$  to an edge of  $H$  or shrinking each edge of  $H$  down to an edge of  $F$ . For a graph  $F$  we denote the set of all  $k$ -uniform hypergraphs without isolated vertices that are a Berge- $F$  by  $\mathcal{B}_k(F)$ . Note that  $\mathcal{B}_k(F)$  is finite for any finite graph  $F$ . We say a hypergraph  $H$  *contains* a Berge- $F$  if  $H$  contains a subhypergraph that is a Berge- $F$ , and that  $H$  is Berge- $F$ -free otherwise. For example, the hypergraph in Figure 4.1 contains a Berge- $P_5$ , but is Berge- $P_6$ -free.

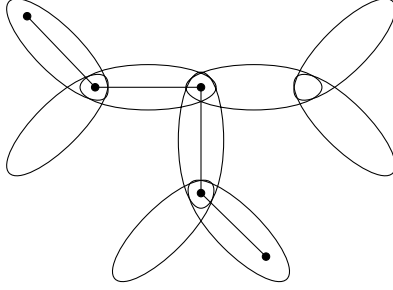


Figure 4.1: A hypergraph  $H$  that contains a Berge- $P_5$ .

Turán-type extremal problems for hypergraphs in the Berge sense have attracted considerable attention [59, 48, 10, 50, 37, 76, 27, 39, 66], where the goal is to determine the maximum possible number of hyperedges in a Berge- $F$ -free hypergraph, called the Turán number of Berge- $F$ . The contents of this chapter will study the concept of saturation for Berge hypergraphs.

A hypergraph  $H$  is Berge- $F$ -saturated if it does not contain any member of  $\mathcal{B}_k(F)$ , but  $H + e$  does for any edge  $e \notin E(H)$ . We denote the minimum number of edges in a  $k$ -uniform Berge- $F$ -saturated hypergraph by  $\text{sat}_k(n, \text{Berge-}F)$ . Hence  $\text{sat}_k(n, \text{Berge-}F) = \text{sat}_k(n, \mathcal{B}_k(F))$ . In [68], Pikhurko showed that  $\text{sat}_k(n, \mathcal{F}) = O(n^{k-1})$  for any finite family of  $k$ -uniform hypergraphs  $\mathcal{F}$ . So we have that  $\text{sat}_k(n, \text{Berge-}F) = O(n^{k-1})$ .

Turán numbers for Berge cycles and Berge paths were found by Győri et al. [49, 16]. For general Berge hypergraphs, some Turán numbers were determined by Gerbner and Palmer [40]. They showed that if  $H$  is Berge- $F$ -free where  $|e| \geq |V(F)|$  for all  $e \in E(H)$ , then  $|E(H)| \leq \text{ex}(n, F)$ . Since any  $k$ -uniform hypergraph that is Berge- $F$ -saturated is also Berge- $F$ -free, we see that  $\text{sat}_k(n, \text{Berge-}F) \leq \text{ex}(n, F) = O(n^2)$  when  $k \geq |V(F)|$ .

Despite intensive research concerning Turán numbers for Berge hypergraphs, no work has been done for saturation numbers for Berge hypergraphs. Determining the minimum number of edges that a hypergraph can have and be Berge- $F$ -saturated is different from classical saturation in several ways. In some cases, the  $k$ -uniform hypergraph obtained by

adding  $k - 2$  new vertices to each edge of a minimally  $F$ -saturated graph is also Berge- $F$ -saturated, but not minimal and in other cases it is not even Berge- $F$ -saturated. For example, the five cycle  $C_5$  is  $K_3$ -saturated, however if we add new vertices to each edge of  $C_5$  to form a 3-uniform hypergraph, then it is not Berge- $K_3$ -saturated. Also, if one adds  $k - 2$  new vertices to each edge in the  $P_m$ -saturated graph given by Kászonyi and Tuza in [56] to form a hypergraph, the resulting hypergraph will be Berge- $P_m$ -saturated, but not minimal.

We will explore the saturation number of Berge hypergraphs for many classes of graphs. In Section 4.5, we determine the saturation numbers for Berge triangles, cycles, matchings, and stars. The results of this section are summarized in the following theorem. It is worth noting that the results that only contain an upper bound may not be tight, but they establish linearity in each case.

**Theorem 4.1.2** *Let  $k \geq 3$ ,  $\ell \geq 1$  and  $m \geq 4$ .*

- a. *For all  $n \geq k + 1$ ,  $\text{sat}_k(n, \text{Berge-}K_3) = \lceil \frac{n-1}{k-1} \rceil$ .*
- b. *For all  $k \geq m - 1$  and  $n > m(k - (m - 2)) + (m - 2)$ ,  $\text{sat}_k(n, \text{Berge-}C_m) \leq \lceil \frac{n-m+2}{k-m+2} \rceil$ .*
- c. *If  $k = m - 2$  and  $n \geq m^2$ ,  $\text{sat}_k(n, \text{Berge-}C_m) \leq \lfloor \frac{n-1}{m-2} \rfloor \binom{m-1}{k} + \frac{(n-1) \bmod (m-2)}{k-2}$ .*
- d. *For all  $k \leq m - 3$ ,  $\ell = \max\{m/2 + 1, k + 1\}$  and  $n \geq \ell^2$ ,  
 $\text{sat}_k(n, \text{Berge-}C_m) \leq \lfloor \frac{n-1}{\ell-1} \rfloor \binom{\ell}{k} + ((n-1) \bmod (\ell-1)) \binom{\ell}{k-1}$ .*
- e. *For all  $n \geq k(\ell - 1)$ ,  $\text{sat}_k(n, \text{Berge-}\ell K_2) = \ell - 1$ .*
- f. *For all  $n \geq k^2$ ,  $\text{sat}_k(n, \text{Berge-}K_{1,k+1}) = n - (k - 1)$ .*
- g. *For all  $k \leq m - 1$ ,  $\text{sat}_k(n, \text{Berge-}K_{1,m}) \leq \lceil \frac{n}{m} \rceil \binom{m}{k}$ .*

Our main result, however, determines  $\text{sat}_k(n, \text{Berge-}P_m)$  where  $P_m$  is the simple graph path on  $m$  vertices. Let

$$a_m^{(k)} = \min\{|E(T)| \mid T \text{ is a } k\text{-uniform Berge-}P_m\text{-saturated linear tree on } \geq k + 1 \text{ vertices}\}.$$

In Sections 4.2, 4.3 and 4.4 we find exactly the value of  $a_m^{(k)}$  for all  $k \geq 3$ ,  $k \neq 5$  and all  $m \geq 10$  and also establish the following theorem.

**Theorem 4.2.3** *Let  $k \geq 3$  with  $k \neq 5$ . Let  $m \geq 10$ , and  $n \geq (k-1)a_m^{(k)} + k - 1$ . Then*

$$\frac{1}{k-1} \left( n - \left\lfloor \frac{n-k+2}{(k-1)a_m^{(k)} + 1} \right\rfloor - k + 2 \right) \leq \text{sat}_k(n, \text{Berge-}P_m),$$

and

$$\text{sat}_k(n, \text{Berge-}P_m) \leq \left\lceil \frac{1}{k-1} \left( n - \left\lfloor \frac{n}{(k-1)a_m^{(k)} + 1} \right\rfloor \right) \right\rceil.$$

It should be noted that the bounds given here differ by at most three. The case of uniformity  $k = 5$  for paths has some complications not present for other values of  $k$ , so this case will not be covered here.

## 4.2 Berge Saturation Numbers for Paths

### 4.2.1 A Lower Bound for $\text{sat}_k(n, \text{Berge-}P_m)$

We give a lower bound for  $\text{sat}_k(n, \text{Berge-}P_m)$  in terms of  $a_m^{(k)}$ .

**Theorem 4.2.1** *Let  $k \geq 3$ ,  $m \geq 10$  and  $n \geq (k-1)a_m^{(k)} + k - 1$ . Then*

$$\text{sat}_k(n, \text{Berge-}P_m) \geq \frac{1}{k-1} \left( n - \left\lfloor \frac{n-k+2}{(k-1)a_m^{(k)} + 1} \right\rfloor - k + 2 \right).$$

*Proof.* Let  $H_0$  be a minimal  $k$ -uniform Berge- $P_m$ -saturated hypergraph on  $n$  vertices. Observe that  $H_0$  cannot have more than  $k-1$  isolated vertices.

**Case 1.**  $H_0$  contains a single edge,  $e$  as a component.

If  $H_0$  contains another component, say  $T$  that is a linear tree, then there must be a path of length at least  $m-2$  starting at each vertex of  $T$ , since otherwise we can add an

edge containing that vertex and  $k - 1$  vertices in  $e$  without creating a Berge- $P_m$ . Then if we consider a vertex in the center of  $T$ , we have that this vertex has a path of length at least  $m - 2$  away from it, and since it is in the center, a second path of at least length  $m - 3$  away from it, and these paths can share at most one edge, so putting these two together, we have a path of length at least  $2m - 6 > m - 1$ , so  $T$  contains a Berge- $P_m$ , contradicting saturation. Thus,  $e$  is the only component of  $H_0$  that is a linear tree. Since any  $k$ -uniform hypergraph on  $n_0$  vertices, no component of which is a linear tree, has at least  $\frac{n_0}{k-1}$  edges, we have

$$|E(H_0)| \geq \frac{n-k}{k-1} + 1 = \frac{n-1}{k-1} > \frac{1}{k-1} \left( n - \left\lfloor \frac{n-k+2}{(k-1)a_m^{(k)} + 1} \right\rfloor - k + 2 \right).$$

**Case 2.**  $H_0$  has no isolated edges and at least  $k - 1$  isolated vertices.

If  $H_0$  has a leaf, then we can add an edge containing the  $k - 1$  isolates and the vertex of degree greater than 1 in the leaf without creating a Berge- $P_m$ . Thus  $H_0$  would not be saturated. Then every component in  $H_0$  that is not an isolate cannot be a linear tree since any non-trivial linear tree contains a leaf. Again, since any  $k$ -uniform hypergraph on  $n_0$  vertices, no component of which is a linear tree, has at least  $\frac{n_0}{k-1}$  edges and since  $n \geq (k-1)a_m^{(k)} + k - 1$ , we have

$$|E(H_0)| \geq \frac{n - (k-1)}{k-1} \geq \frac{1}{k-1} \left( n - \left\lfloor \frac{n-k+2}{(k-1)a_m^{(k)} + 1} \right\rfloor - k + 2 \right).$$

**Case 3.**  $H_0$  has no isolated edges, and no more than  $k - 2$  isolated vertices.

Assume that  $H_0$  has  $c \geq 1$  connected components  $C_1, \dots, C_c$  and assume without loss of generality that the first  $t$  of these are linear trees for some  $0 \leq t \leq c$ . If a linear tree has  $a_m^{(k)}$  edges, then it has  $b = (k-1)a_m^{(k)} + 1$  vertices. There are at most  $k - 2$  isolates in  $H_0$ , so there are at least  $t - k + 2$  non-trivial trees in  $H_0$ . This implies that  $(t - k + 2)b \leq n - k + 2$ ,

or  $t \leq \lfloor \frac{n-k+2}{b} \rfloor + k - 2$ . For  $i \leq t$ ,  $|E(C_i)| = \frac{|V(C_i)|-1}{k-1}$ , and for  $i > t$ ,  $|E(C_i)| \geq \frac{|V(C_i)|}{k-1}$ . Thus

$$\begin{aligned} |E(H_0)| &\geq \sum_{i=1}^t \frac{|V(C_i)|-1}{k-1} + \sum_{i=t+1}^c \frac{|V(C_i)|}{k-1} = \frac{1}{k-1}(n-t) \\ &\geq \frac{1}{k-1} \left( n - \left\lfloor \frac{n-k+2}{b} \right\rfloor - k + 2 \right). \end{aligned}$$

□

The rest of the chapter will be dedicated to determining  $a_m^{(k)}$ . In section 4.3 we will give constructions for  $k$ -uniform Berge- $P_m$ -saturated linear trees and in section 4.4 we will show that these constructions are minimal. The results of these sections will imply the following theorem.

**Theorem 4.2.2** *Let  $m \geq 10$ . If  $m = 4s + r$  for  $1 \leq r \leq 4$ , then*

$$a_m^{(3)} = (3+r)2^s - 5.$$

*If  $m = 6s + r$  for  $0 \leq r \leq 5$ , then*

$$a_m^{(4)} = (6+r)2^s - 8.$$

*If  $k \geq 6$ , then*

$$a_m^{(k)} = \begin{cases} 2^{s+1} + 2^s + 2^{s-1} + 2^{s-2} - 6 & \text{if } m = 4s, \\ 2^{s+2} + 2^{s-1} - 6 & \text{if } m = 4s + 1, \\ 2^{s+2} + 2^s - 6 & \text{if } m = 4s + 2, \\ 2^{s+2} + 2^{s+1} + 2^{s-1} - 6 & \text{if } m = 4s + 3. \end{cases}$$

*Proof.* Theorems 4.4.13, 4.4.14 and 4.4.15 give us the lower bounds on  $a_m^{(k)}$ , while Lemma 4.3.7 and observations 4.3.2, 4.3.4 and 4.3.6 give us the matching upper bounds. □



Theorems 4.2.1 and 4.2.2 give us a lower bound on  $\text{sat}_k(n, \text{Berge-}P_m)$ , while Construction 4.3.8, Observation 4.3.9 and Lemma 4.3.10 give us an upper bound. These bounds are summarized in the following theorem.

**Theorem 4.2.3** *Let  $k \geq 3$  with  $k \neq 5$ . Let  $m \geq 10$ , and  $n \geq (k-1)a_m^{(k)} + k - 1$ . Then*

$$\frac{1}{k-1} \left( n - \left\lfloor \frac{n-k+2}{(k-1)a_m^{(k)} + 1} \right\rfloor - k + 2 \right) \leq \text{sat}_k(n, \text{Berge-}P_m),$$

and

$$\text{sat}_k(n, \text{Berge-}P_m) \leq \left\lceil \frac{1}{k-1} \left( n - \left\lfloor \frac{n}{(k-1)a_m^{(k)} + 1} \right\rfloor \right) \right\rceil.$$

It is the belief of the authors that the upper bound in the previous theorem should be correct. In the proof of Theorem 4.2.1, we allow for the possibility for a minimal saturated graph to contain up to  $k-2$  isolated vertices, but this may not be possible. If it can be shown that there are no isolated vertices in the minimal construction, the upper and lower bounds will match. This leads to the following conjecture:

**Conjecture 4.2.4** *Let  $k \geq 3$ ,  $m \geq 10$ ,  $n \geq (k-1)a_m^{(k)} + 1$ . Then*

$$\text{sat}_k(n, \text{Berge-}P_m) = \left\lceil \frac{1}{k-1} \left( n - \left\lfloor \frac{n}{(k-1)a_m^{(k)} + 1} \right\rfloor \right) \right\rceil.$$

### 4.3 Constructing an Upper Bound on $a_m^{(k)}$

In this section, we give a construction for a Berge- $P_m$  saturated linear tree,  $T_m^{(k)}$ , for  $k = 3, 4$  and  $k \geq 6$  and for all  $m \geq 10$ , and we show how to use  $T_m^{(k)}$  to create a saturated graph on  $n$  vertices for any  $n \geq (k-1)a_m^{(k)} + 1$ . In section 4.4 we will show that our trees are minimal, or more precisely,  $|E(T_m^{(k)})| = a_m^{(k)}$ .

### 4.3.1 Constructing the Tree $T_m^{(k)}$

We begin determining the upper bound for  $\text{sat}_3(n, \text{Berge-}P_m)$  by constructing 3-uniform linear trees that are Berge- $P_m$  saturated for  $m \geq 8$ .

**Construction 4.3.1** For  $m \geq 8$ , let  $T_m^{(3)}$  be the linear tree built as follows:

1. Start with a central vertex  $v$  in an edge,  $e$ .
2. Add a pendant edge to both vertices in  $e - \{v\}$  to begin the two main branches of the tree. Call these the left and right sides of the tree.
3. To each vertex added in the previous step, append a pendant edge.
4. Add a pendant edge to one degree 1 vertex in each of the edges added in step 3.
5. If  $m$  is odd, repeat steps 3 and 4 until there are  $\frac{m-3}{2}$  levels of the linear tree after the initial edge  $e$ . If  $m$  is even, repeat steps 3 and 4 until there are  $\frac{m-4}{2}$  levels after the initial edge  $e$ , and then add an additional level to the left side of the tree.

Figures 4.2 and 4.3 show the constructions for  $m = 8$  and  $m = 11$  respectively.

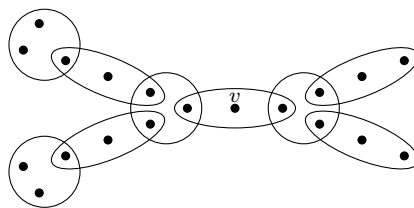


Figure 4.2: The saturated tree  $T_8^{(3)}$ .

**Observation 4.3.2** If  $m \geq 8$ ,  $1 \leq r \leq 4$  and  $m = 4s + r$ , then

$$|E(T_m^{(3)})| = (3 + r)2^s - 5.$$

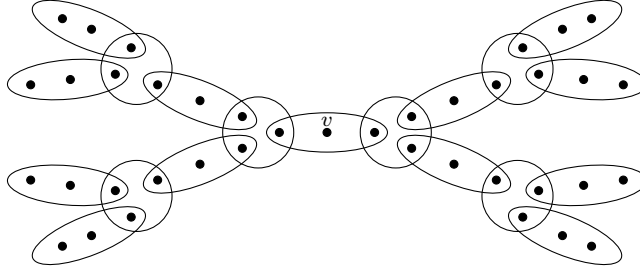


Figure 4.3: The saturated tree  $T_{11}^{(3)}$ .

Next, we provide the tree construction for uniformity  $k = 4$

**Construction 4.3.3** For  $m \geq 10$ , let  $T_m^{(4)}$  be the linear tree built as follows:

1. Start with a central vertex,  $v$  and two edges that intersect only at  $v$ .
2. Add a pendent edge to one degree 1 vertex in each of the edges added in the previous step to begin the two main branches of the tree. Call these the left and right sides of the tree.
3. Add a pendant edge to each of two degree 1 vertices in each of the edges added in the previous step.
4. Add a pendant edge to one degree 1 vertex in each of the edges added in the previous step.
5. Repeat steps 2,3,4 until the tree has  $3 \lfloor \frac{m}{6} \rfloor - 1$  levels after the initial vertex.

For  $m = 0 \pmod{6}$ , we do not need any additional steps.

For  $m = 1 \pmod{6}$ , repeat step 3 on the left side of the tree.

For  $m = 2 \pmod{6}$ , repeat step 3.

For  $m = 3 \pmod{6}$ , repeat step 3, and repeat step 4 on the left side of the tree.

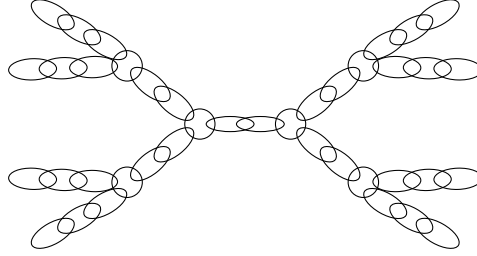


Figure 4.4: The saturated tree  $T_{18}^{(4)}$ .

For  $m = 4 \pmod 6$ , repeat steps 3 and 4.

For  $m = 5 \pmod 6$ , repeat steps 3 and 4, and step 4 again on the left side of the tree.

See Figure 4.4 for an example of the construction for  $m = 18$ .

**Observation 4.3.4** For  $m \geq 10$ , let  $0 \leq r \leq 5$  such that  $m = 6s + r$ . Then

$$|E(T_m^{(4)})| = (6 + r)2^s - 8.$$

Finally, we present a construction for  $T_m^{(k)}$  with  $k \geq 6$ . The construction for when  $m$  is divisible by 4 is distinct from other divisibilities. We give the more general construction first.

**Construction 4.3.5** For  $m \geq 8$  and  $k \geq 6$ , let  $T_m^{(k)}$  be the  $k$ -uniform linear tree built as follows:

For  $m = 1, 2, 3 \pmod 4$ :

1. Start with a central vertex,  $v$  and two edges that intersect only at  $v$  that begin the two main branches of the tree. Call these the left and right sides.
2. Add two pendant edges to one degree 1 vertex in each of the edges added in the previous step.

3. Add a pendant edge to one degree one vertex in each of the edges added in the previous step.

4. For  $m = 1 \pmod 4$ , repeat steps 2 and 3 until there are  $\frac{m-5}{2} + 1$  levels after  $v$ , and repeat step 3 for one of the two main branches.

For  $m = 2 \pmod 4$ , repeat steps 2 and 3 until there are  $\frac{m-6}{2} + 1$  levels after  $v$  and then repeat step 3.

For  $m = 3 \pmod 4$ , repeat steps 2 and 3 until there are  $\frac{m-7}{2} + 1$  levels after  $v$ , and repeat steps 2 and then 3 for one of the two main branches and just step 3 for the other.

For  $m = 4s$ :

1. Start with a central vertex,  $v$  and attach three edges to create a star. These will begin the three main branches of the construction.

2. Add a pendant edge to one degree 1 vertex in each of the edges added in the previous step.

3. Add two pendant edges to one degree 1 vertex in each of the edges added in the previous step.

4. Repeat steps 2 and 3 until there are  $\frac{m-4}{2}$  levels after the initial vertex and then repeat step 2. Note: For  $m = 8$ , the construction does not use step 3.

See Figures 4.5 and 4.6 for  $T_8^{(k)}$  and  $T_{13}^{(k)}$ , respectively.

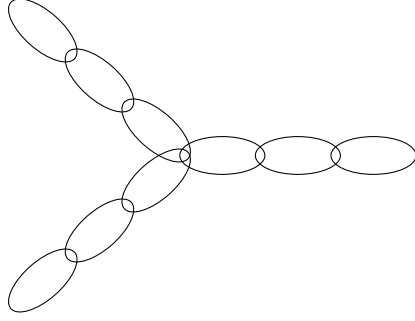


Figure 4.5: The saturated tree  $T_8^{(k)}$ .

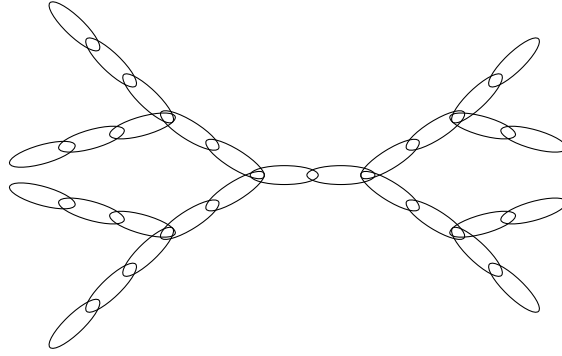


Figure 4.6: The saturated tree  $T_{13}^{(k)}$ .

**Observation 4.3.6** *If  $k \geq 6$  and  $m \geq 8$ , then*

$$|E(T_m^{(k)})| = \begin{cases} 2^{s+1} + 2^s + 2^{s-1} + 2^{s-2} - 6 & \text{if } m = 4s, \\ 2^{s+2} + 2^{s-1} - 6 & \text{if } m = 4s + 1, \\ 2^{s+2} + 2^s - 6 & \text{if } m = 4s + 2, \\ 2^{s+2} + 2^{s+1} + 2^{s-1} - 6 & \text{if } m = 4s + 3. \end{cases}$$

### 4.3.2 The Saturation of $T_m^{(k)}$

We now provide a proof that  $T_m^{(k)}$  is Berge- $P_m$ -saturated for all  $k \geq 6$  and  $m \geq 10$ . As the saturation of  $T_m^{(3)}$  and  $T_m^{(4)}$  follow from almost the same arguments that we will make for the

case  $k \geq 6$ , we provide a sketch of how to adapt the  $k \geq 6$  proof for these cases.

A vertex of degree 3 in  $T_m^{(k)}$  will be called a *branch vertex* and an edge that contains three vertices of degree 2 will be called a *branch edge*. Note that  $T_m^{(3)}$  and  $T_m^{(4)}$  contain branch edges but not branch vertices, while  $T_m^{(k)}$  with  $k \geq 6$  contains branch vertices but not branch edges. With these definition in hand, we now prove the following:

**Lemma 4.3.7** *For all  $k \geq 3$ ,  $k \neq 5$  and  $m \geq 10$ ,  $T_m^{(k)}$  is Berge- $P_m$ -saturated.*

*Proof.* It is clear from the construction of  $T_m^{(k)}$  that the longest path stretches from a leaf on the left to a leaf on the right, or in the case  $k \geq 6$  and  $4|m$ , a leaf in one of the three main branches to a leaf in another main branch. This path is of length  $m - 2$ , so  $T_m^{(k)}$  is Berge- $P_m$ -free. Now we show that any edge added will create a Berge- $P_m$ .

We first consider the case when  $k \geq 6$  and  $4 \nmid m$ . Let  $e$  be any edge in  $\overline{T_m^{(k)}}$ . For each vertex  $v \in V(T_m^{(k)})$ , let  $\ell(v)$  and  $r(v)$  be the length of the shortest path that starts at  $v$  and ends on a leaf on the left, or right side of  $T_m^{(k)}$  respectively. Note that for each  $v$   $\ell(v) + r(v) = m - 2$  if  $d(v) \geq 2$  and  $\ell(v) + r(v) = m - 1$  if  $d(v) = 1$  since the edge containing  $v$  will be used both on the left path and the right path.

Let  $x, y \in e$  with  $\min\{\ell(x), r(x)\} \leq \min\{\ell(y), r(y)\}$ . Let  $P$  be a Berge- $P_{m-1}$  containing  $x$ . If  $y$  is not contained in an edge of  $P$ , then we can traverse the longer part of  $P$ , hitting at least  $m - 2 - \min\{\ell(x), r(x)\}$  edges until we hit  $x$ , then use  $e$  to jump to  $y$ , and then take a path from  $y$  on  $\min\{\ell(y), r(y)\}$  edges to a leaf that does not contain  $x$ . This path has length at least

$$m - 2 - \min\{\ell(x), r(x)\} + 1 + \min\{\ell(y), r(y)\} \geq m - 1.$$

Thus this path contains a desired Berge- $P_m$  in  $T_m^{(k)} + e$ .

Then we may assume that each pair  $x, y$  in  $e$  with  $\min\{\ell(x), r(x)\} \leq \min\{\ell(y), r(y)\}$  has that  $y$  is in every longest path containing  $x$ . This implies that all the vertices in  $e$

are in a single longest path  $P^*$ . Further, this implies that every vertex is on one side of the construction or in the center since if we had vertices properly on both sides of the construction, there would be a Berge- $P_{m-1}$  that contains a vertex on one side, but not the vertex on the other side due to the branching structure of  $T_m^{(k)}$ . Say that the vertices are all on the left side or in the center. Let  $x^* \in e$  be such that  $\ell(x^*) = \min\{\ell(v) \mid v \in e\}$  and let  $y^*$  be such that  $\ell(y^*) = \max\{\ell(v) \mid v \in e\}$ .

If  $d(x^*, y^*) \geq 4$ , then there is a branch vertex between  $x^*$  and  $y^*$  that is not adjacent to  $x^*$ . We can traverse  $P^*$  until we get to  $y^*$  picking up  $r(y^*)$  edges, then hop to  $x^*$ , then traverse  $P^*$  backwards until we hit the last branch vertex of  $P^*$  before  $y^*$  hitting at least  $d(x^*, y^*) - 2$  edges, then take this branch down to a leaf using at least  $\ell(y^*) - 2$  edges, giving us a path of length at least

$$r(y^*) + 1 + d(x^*, y^*) - 2 + \ell(y^*) - 2 \geq r(y^*) + \ell(y^*) + 1 \geq m - 1.$$

Thus we can assume that  $d(x^*, y^*) \leq 3$ .

Note that this preceding path also works if  $d(x^*, y^*) = 3$  as long as either  $d(y^*) = 1$ , since then  $r(y^*) + \ell(y^*) = m - 1$ , or  $d(y^*) = 2$  since then we hit  $d(x^*, y^*) - 1$  edges backtracking along  $P^*$  and  $\ell(y^*) - 1$  edges going down to a leaf since the branching point is only one edge away from  $y^*$ . Thus we may assume  $d(x^*, y^*) \leq 3$  and if equality holds,  $d(y^*) = 3$ .

If  $e$  contains a pair of vertices  $u_1$  and  $u_2$  such that  $d(u_1) \geq 2$ ,  $d(u_2) = 1$  and  $d(u_1, u_2) = 1$ , or such that  $d(u_1) = d(u_2) = 1$  and  $d(u_1, u_2) = 2$ , then  $e$  can be added into any path of length  $m - 2$  that contains the edge containing  $u_1$  and  $u_2$  by traversing the path until  $u_1$ , then using  $e$  to jump from  $u_1$  to  $u_2$ , then traversing the rest of the path. Thus we are done unless  $d(x^*, y^*) \geq 3$  since there is no way for two adjacent edges to contain all  $k$  vertices without having a pair of vertices that can play the roles of  $u_1$  and  $u_2$  above.

Then we have that  $d(x^*, y^*) = 3$  and  $d(y^*) = 3$ . Let  $e'_1$ ,  $e'_2$  and  $e'_3$  be the edges in the



path between  $x^*$  and  $y^*$  with  $x^* \in e'_1$  and  $y^* \in e'_3$ . Now,  $e \subset e'_1 \cup e'_2 \cup e'_3$ . We can assume that there are no vertices from  $e$  in  $e'_3 \setminus e'_2$  aside from  $y$  since any such vertex along with  $y$  would satisfy the conditions on  $u_1$  and  $u_2$  above.

It could be the case that  $e'_1$  is a leaf, as in the left diagram of Figure 4.7. In this case,  $d(x^*) = 1$ , so any vertex in  $e'_2 \setminus e'_3$  along with  $x^*$  would satisfy the conditions on  $u_1$  and  $u_2$  above, and if any vertex in  $e'_1 \setminus e'_2$  aside from  $x^*$  was in  $e$ , then we could extend a path of length  $m - 2$  that ends at  $x^*$  by 1 edge using  $e$  to jump from  $x^*$  to the other vertex in  $e'_1 \cap e$ . Finally any degree 1 vertex in  $e'_3$  along with  $y^*$  satisfy the conditions on  $u_1$  and  $u_2$  above, so none of these vertices can be in  $e$ . This only leaves the vertex in  $e'_2 \cap e'_3$ , but  $k \geq 6$  giving us a contradiction with the size of the edges  $e$ . Thus we can assume  $e'_1$  is not a leaf, as in the right hand diagram of Figure 4.7.

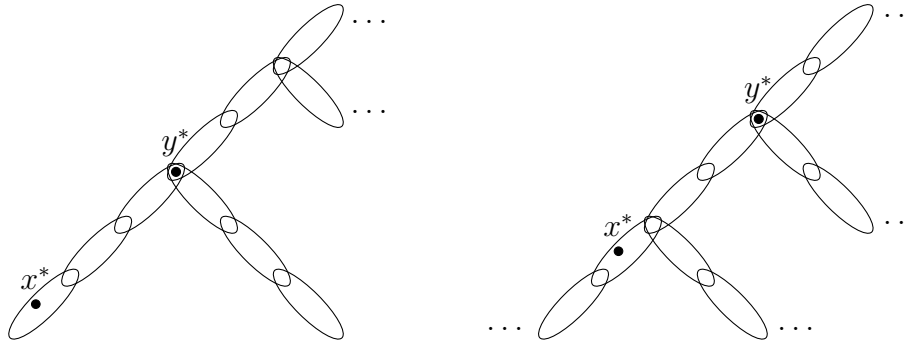


Figure 4.7: The two cases when  $d(x^*, y^*) = 3$  and  $d(y^*) = 3$ .

If  $d(x^*) \geq 2$  in  $T_m^{(k)}$ , then by the preceding argument, there is at most one vertex in  $e$  at distance 1 from  $x^*$ , and if  $d(x^*) = 1$ , then by the preceding argument, there are at most  $k - 3$  vertices from  $e$  at distance 1 from  $x^*$  since they all must be degree 1 vertices. Thus in all cases, there are at least two vertices not in the same edge as  $x^*$ . One is  $y^*$ , call the other  $w^*$ .

Since  $y^*$  has degree 3 and  $d(y^*, w^*) \leq 2$ , there is a branch vertex between  $w^*$  and  $x^*$ , and  $d(w^*) \leq 2$  since this branch vertex is distance 1 from  $x^*$ .

Then we can traverse  $P^*$  until we hit  $w^*$  picking up  $r(w^*)$  edges, then jump to  $x^*$ , pick up a single edge going to the branch vertex, then take a path down from the branching point using  $\ell(w^*) - 2$  edges if  $d(w^*) = 1$  and  $d(w^*, y^*) = 1$  or  $\ell(w^*) - 1$  otherwise. If  $d(w^*) = 1$  and  $d(w^*, y^*) = 1$ , then this gives us a path on

$$r(w^*) + 1 + 1 + \ell(w^*) - 2 = r(w^*) + \ell(w^*) = m - 1$$

edges and otherwise, a path on

$$r(w^*) + 1 + 1 + \ell(w^*) - 1 = r(w^*) + \ell(w^*) + 1 \geq m - 1$$

edges, a desired path in either case. Thus  $T_m^{(k)}$  is saturated when  $4 \nmid k$ .

When  $k \geq 6$  and  $4 \mid k$ , the proof follows from a similar argument. The only difference is that instead of having two sides, a left side and a right side,  $T_m^{(k)}$  has three branches coming from the center. Thus instead of having  $\ell(v)$  and  $r(v)$ , we can label the three branches with integers 1, 2 and 3, and define  $\ell_i(v)$  for  $1 \leq i \leq 3$  by the distance from  $v$  to a leaf in branch  $i$ . This will give us that all the vertices of  $e$  are on one path inside one main branch. From here, the proof is identical.

For  $k = 3$  and  $k = 4$ , again the proof follows similarly. We can argue that all the vertices of  $e$  must lie on one path and all on one side of the tree in exactly the same ways as  $k \geq 6$ . Define  $x^*$  and  $y^*$  in the same way as in the proof of  $k \geq 6$ . For  $k = 3$ , we can establish that  $d(x^*, y^*) \leq 3$  the same as before, but with branch edges in place of branch vertices. In the case of  $k = 4$ , we cannot prove that  $d(x^*, y^*) \leq 3$  as easily since the branch edges in  $T_m^{(4)}$  are further away from each other than branch edges in  $T_m^{(3)}$ . Even so, using essentially the same idea we can establish  $d(x^*, y^*) \leq 5$ . From here, for both  $k = 3$  and  $k = 4$ , there are only a few small cases to work out similar to the end of the proof for  $k \geq 6$ .  $\square$

It is worth noting that in Lemma 4.3.7 we assume  $m \geq 10$  even though we define some constructions with  $m < 10$ , such as  $T_8^{(3)}$  or  $T_9^{(172)}$ . These constructions are also saturated; we choose  $m \geq 10$  to simplify the statement of Lemma 4.3.7 since we do not provide constructions for  $k = 4$  and  $m < 10$ .

### 4.3.3 Using $T_m^{(k)}$ to Construct the Saturated Hypergraph $H_{n,m}^{(k)}$

We will use the tree  $T_m^{(k)}$  as the building block to construct a Berge- $P_m$  saturated hypergraph on  $n$  vertices which we will call  $H_{n,m}^{(k)}$ .

**Construction 4.3.8** *Let  $k \geq 3$ ,  $k \neq 5$  and  $m \geq 10$ .*

*Let  $r = n \bmod ((k-1)|E(T_m^{(k)})|+1)$  and let  $H$  be the hypergraph on  $n$  vertices that contains  $\lfloor \frac{n}{(k-1)|E(T_m^{(k)})|+1} \rfloor$  components isomorphic to  $T_m^{(k)}$  and  $r$  isolates,  $v_1, \dots, v_r$ . If  $r > 0$ , we will show how to incorporate these vertices into one of the copies of  $T_m^{(k)}$ , which we will call  $T'$ .*

*First, we will create new leaves in  $T'$  using the isolates in  $H$ . Let  $v' \in V(T')$  be the vertex of degree 2 in some leaf of  $T'$  and for each  $0 \leq i \leq \lfloor \frac{r}{k-1} \rfloor - 1$ , let  $e_i = \{v', v_{(k-1)i+1}, v_{(k-1)i+2}, \dots, v_{(k-1)i+k}\}$ . Then let  $H'$  be a hypergraph with  $V(H') = V(H)$  and  $E(H') = E(H) \cup \bigcup_{i=0}^{\lfloor \frac{r}{k-1} \rfloor - 1} e_i$ .*

*Then  $H'$  has  $r' = r \bmod (k-1)$  isolates. If  $r' > 0$ , the method for including the remaining  $r'$  vertices is dependent upon the uniformity, but for each  $k$  we will add one additional edge,  $e'$ , which will be incident to the same edges as some specific edge  $e^* \in T_m^{(k)}$ .*

*For  $k = 3$ , we must have  $r' = 1$ . This vertex can be added by “cloning” the central edge of  $T'$  meaning we form an edge using the remaining new vertex and the two vertices of degree 2 in the center edge of  $T'$ . In this case  $e^*$  is the central edge. For  $k = 4$ , consider the left three initial edges,  $e_1, e_2$ , and  $e_3$ , labeling from the left. Let  $e'$  be the edge containing the  $r'$  isolates, the vertex of degree 2 in  $e_1 \cap e_2$ , and  $3 - r' \geq 1$  vertices of degree 1 in  $e_3$ . In this*

case  $e^* = e_2$ . For  $k \geq 6$ , let  $e'$  be the edge containing the  $r'$  isolates, the degree 3 vertex in one of the center edges, and  $k-(r'+1)$  vertices of degree 1 from the other center edge as in Figure 4.8. In this case  $e^*$  is as pictured in Figure 4.8.

Then we let  $H_{n,m}^{(k)} = H' + e'$ .

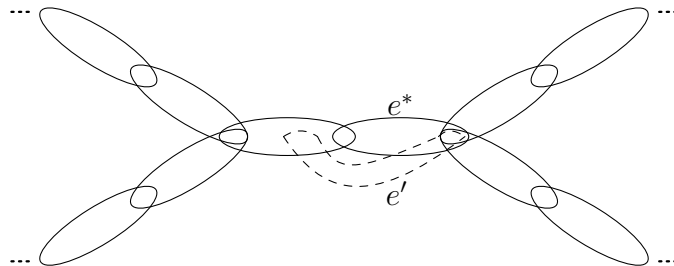


Figure 4.8: The hypergraph  $T'$  when  $k \geq 6$ .

**Observation 4.3.9** Let  $k \geq 3$  with  $k \neq 5$ , and let  $m \geq 10$ . Then for all  $n \geq (k - 1)|E(T_m^{(k)})| + 1$ ,

$$|E(H_{n,m}^{(k)})| = \left\lceil \frac{1}{k-1} \left( n - \left\lfloor \frac{n}{(k-1)|E(T_m^{(k)})|+1} \right\rfloor \right) \right\rceil$$

The preceding observation follows from the fact that  $H_{n,m}^{(k)}$  has  $\left\lfloor \frac{n}{(k-1)|E(T_m^{(k)})|+1} \right\rfloor$  components, at most one of which is not a linear tree, and if  $H_{n,m}^{(k)}$  does contain a component that is not a linear tree, this component is minimal in the sense that there are no connected  $k$ -uniform hypergraphs on the same vertex set with fewer edges.

**Lemma 4.3.10**  $H_{n,m}^{(k)}$  is saturated for all  $k \geq 3$  with  $k \neq 5$  and all  $m \geq 10$ .

*Proof.* Fix  $k$  and  $m$ . Let  $H = H_{n,m}^{(k)}$ .  $H$  consists of several components isomorphic to  $T_m^{(k)}$ , and one component  $T'$ , which is a copy of  $T_m^{(k)}$  that may have been modified by adding copies of a particular leaf at the vertex  $v'$  and possibly adding an edge near the center called  $e'$ .

We first show  $H$  contains no Berge- $P_m$ . Indeed, from Lemma 4.3.7 we know that no component isomorphic to  $T_m^{(k)}$  contains a Berge- $P_m$ , so we can restrict our attention to  $T'$ . Since there already was a leaf at  $v'$ , any added leaves could not create a Berge- $P_m$ . If  $e' \in E(H)$ , consider a longest path,  $P$  in  $T'$  using  $e'$ . Let  $e^*$  be as described in Construction 4.3.8. If  $P$  does not use  $e^*$ , then we can replace  $e'$  with  $e^*$  to create a path that only uses edges in  $T_m^{(k)}$ , with the same length, and thus is not a Berge- $P_m$ . On the other hand if  $P$  uses both  $e^*$  and  $e'$ , for  $k = 3$ , both vertices of degree 3 in  $e'$  will be used by  $P$ , and so one end of  $P$  must be in  $e^*$  or  $e'$ . Similarly, for  $k = 4$  or  $k \geq 6$ ,  $P$  must use the vertex of degree 3 in  $e^* \cap e'$  and the other center edge that is incident to both  $e^*$  and  $e'$ , and so one end of  $P$  must be  $e^*$  or  $e'$ . Thus, if  $P$  uses both  $e^*$  and  $e'$ ,  $P$  has length at most  $\lfloor \frac{m}{2} \rfloor + 2$ , which for  $m \geq 10$  is less than  $m - 1$ . Thus  $H$  contains no Berge- $P_m$ .

Now we will show that for all  $e \in E(\overline{H})$ ,  $H + e$  contains a Berge- $P_m$ . By lemma 4.3.7, each  $T_m^{(k)}$  is saturated, and so any edge added that is contained completely in one of the  $T_m^{(k)}$  results in a Berge- $P_m$ . It remains to show that  $T'$  is saturated, and that any edge added between components results in a Berge- $P_m$ .

Consider adding an edge  $e \in \overline{T'}$  to  $T'$ . If  $e$  contains only vertices in the underlying  $T_m^{(k)}$ , then  $T' + e$  contains a Berge- $P_m$ . If  $e$  contains vertices from more than one leaf incident with  $v'$ , any Berge- $P_{m-1}$  that has one of those leaves as a terminal edge can be extended in length by 1, thus we can assume  $e$  is incident with at most one leaf incident with  $v'$ , and without loss of generality, we can assume that leaf is the one in  $T_m^{(k)}$ . Thus, the only case we need still check is when  $e$  contains a vertex in  $e'$  that is not in the underlying  $T_m^{(k)}$ .

If  $e$  contains distinct vertices in  $e^*$  and  $e'$ , then any Berge- $P_{m-1}$  that contains  $e^*$  in the underlying  $T_m^{(k)}$  can be extended to a Berge- $P_m$  containing all the previous edges and both  $e'$  and  $e^*$  since  $e'$  is incident to the same edges as  $e$ . Otherwise,  $e'$  acts like  $e^*$ , so any edge incident with vertices from  $e'$  will create a Berge- $P_m$  just as if the edge was incident with vertices from  $e^*$ . Thus,  $T'$  is saturated.

Finally, we show that any edge added between components results in a Berge- $P_m$ . If we add an edge between any two components of  $H$ , then since every vertex of  $T_m^{(k)}$ , or  $T'$  is in some Berge- $P_{m-1}$ , each vertex is the initial vertex of a path of length at least  $\lceil \frac{m-2}{2} \rceil$ , and so we can construct a path of length at least  $2 \lceil \frac{m-2}{2} \rceil + 1 \geq m - 1$ . Thus, in all cases, the addition of an edge creates a Berge- $P_m$ , so  $H$  is Berge- $P_m$ -saturated.  $\square$

We have constructed Berge- $P_m$  saturated hypergraphs. In the next section, we will show that these hypergraphs are edge minimal.

## 4.4 Lower Bound on $a_m^{(k)}$

This section will establish a lower bound on  $a_m^{(k)}$ , the least number of edges in any  $k$ -uniform Berge- $P_m$ -saturated linear tree on at least  $k + 1$  vertices for  $k = 3$ ,  $k = 4$  and  $k \geq 6$ . The structure of such trees is highly dependent on the value of  $k$ , so this section essentially provides three proofs, one for each case. These proofs are very similar to each other, but have non-trivial differences. That being said, to get a feeling for how these proofs work, a reader does not need to read this entire section.

All cases rely on lemmas 4.4.1 and 4.4.3. Furthermore, in all cases we need to rule out the possibility that a minimal Berge- $P_m$ -saturated tree does not contain a Berge- $P_{m-1}$ . This is done in lemmas 4.4.7, 4.4.11 and 4.4.12. Aside from these lemmas, the proofs for each case are independent of the others. If the reader does not wish to read all three cases, the authors recommend one of the following two reading paths in addition to the five lemmas above:

The simplest case to understand is  $k = 3$ . This case follows from lemmas 4.4.6 and 4.4.8 and concludes in Theorem 4.4.13.

If the reader instead wishes to read the most general case of  $k \geq 6$ , this consists of the lemmas 4.4.4 and 4.4.10 and concludes in Theorem 4.4.15. The reader may also wish to

read each of the paragraphs at the beginning of subsections 4.4.1, 4.4.2 and 4.4.3 as these paragraphs foreshadow what follows in each subsection.

Now we can begin to establish the lower bound. Before we prove the lower bound on  $a_m^{(k)}$ , we need to develop a few elementary structural lemmas and make a few observations about Berge- $P_m$ -saturated graphs.

**Lemma 4.4.1** *Let  $k \geq 3$ . If  $H$  is a  $k$ -uniform Berge- $P_m$ -saturated linear tree on  $n \geq k + 1$  vertices, and  $e_1$  and  $e_2$  are a pair of adjacent edges in  $H$ , then there is a pair of vertex-disjoint paths of length  $\alpha$  and  $\beta$  that start at vertices in  $e_1 \cup e_2$  and do not use either edge  $e_1$  or  $e_2$  such that  $\alpha + \beta \geq m - 4$ .*

*Proof.* Let  $e_1$  and  $e_2$  be a pair of adjacent edges. Let  $e$  be any edge in  $\overline{H}$  such that each vertex of  $e$  is in either  $e_1$  or  $e_2$ . Consider  $H + e$ . Due to saturation, there exists a Berge- $P_m$  in  $H + e$  that uses the edge  $e$ . In addition to  $e$ , this Berge- $P_m$  could possibly use edges  $e_1$ ,  $e_2$  and edges from two edge-disjoint paths leaving  $e_1$  or  $e_2$ , but no more edges. Thus if the longest such paths are of length  $\alpha$  and  $\beta$ , we must have that  $\alpha + \beta + 3 \geq m - 1$  or  $\alpha + \beta \geq m - 4$ .

To see that these two paths are vertex-disjoint, assume that the only pair of paths long enough to satisfy the length requirement attach at the same vertex  $x \in e_1 \cup e_2$ . Consider an edge  $e' \subseteq (e_1 \cup e_2) \setminus \{x\}$ . Then,  $H + e'$  will not contain a Berge- $P_m$  since we cannot use  $e'$  while also using both paths starting at  $x$ , since this would require us to visit  $x$  twice in our path. □

**Observation 4.4.2** *Let  $k \geq 3$ . If  $H$  is a  $k$ -uniform Berge- $P_m$ -saturated linear tree on  $n \geq k + 1$  vertices, then  $H$  contains a Berge- $P_{m-2}$  (a path of length  $m - 3$ ).*

This follows immediately from applying Lemma 4.4.1 to any pair of edges in  $H$ .

**Lemma 4.4.3** *Let  $k \geq 3$ . If  $H$  is a  $k$ -uniform Berge- $P_m$ -saturated linear tree on  $n \geq k + 1$  vertices, then every edge in  $H$  that is not a leaf is in a Berge- $P_{m-2}$ .*

*Proof.* Let  $e$  be an edge of  $H$  that is not a leaf. Since  $e$  is not a leaf, it must be in a path of length at least 3, say with neighboring edges  $e'$  and  $e''$ . Apply Lemma 4.4.1 to edge  $e$  and  $e'$ . Then we have paths of length  $\alpha_1$  and  $\beta_1$  leaving distinct vertices in  $e \cup e'$  with  $\alpha_1 + \beta_1 \geq m - 4$ . If either of these paths attach at a vertex in  $e \setminus e'$ , then we are done since regardless of where the second path attaches, we can create a path using  $e$  and both the path of length  $\alpha_1$  and  $\beta_1$ , which is of length at least  $m - 3$ .

Otherwise both paths attach to vertices in  $e'$ . We now apply Lemma 4.4.1 to  $e$  and  $e''$  to get paths of length  $\alpha_2$  and  $\beta_2$  leaving distinct vertices in  $e \cup e''$ . If either one leaves a vertex in  $e \setminus e''$ , then we are done by the argument in the preceding paragraph, so assume both attach to vertices in  $e''$ . Without loss of generality let  $\alpha_1 \geq \beta_1$  and  $\alpha_2 \geq \beta_2$  be the longer paths generated by the two applications of Lemma 4.4.1. We note that  $\alpha_i \geq \lceil \frac{m-4}{2} \rceil$  for  $i \in \{1, 2\}$ . Thus, there is a path through  $e$  that traverses each of these longer paths, so it has length at least  $\lceil \frac{m-4}{2} \rceil + \lceil \frac{m-4}{2} \rceil + 1 \geq m - 3$ .  $\square$

#### 4.4.1 Branching Lemmas

If we consider a loose path on  $m - 2$  edges, it is clear this will not be Berge- $P_m$ -saturated since we can add edges that contain only vertices of degree 2 or edges that contain vertices of degree 1 from two non-consecutive non-leaf edges. Due to these restrictions, any path in a Berge- $P_m$ -saturated graph must branch fairly often. This section establishes lemmas that characterize the minimal necessary branching of a Berge- $P_m$ -saturated tree. We will find that the minimal amount of branching depends on if the path we are currently considering contains a Berge- $P_{m-1}$  or not.

##### Branching with a Berge- $P_{m-1}$

If  $H$  contains a Berge- $P_{m-1}$ , the amount of branching necessary to retain saturation is dictated by the uniformity  $k$ . Our first lemma is true for all  $k \geq 3$ , but is trivial for  $k = 3$



and can be refined for  $k = 4$ . Also in this section is the refinement for  $k = 4$  and a non-trivial branching lemma for  $k = 3$ .

**Lemma 4.4.4** *Let  $k \geq 3$ . Let  $H$  be a  $k$ -uniform Berge- $P_m$ -saturated linear tree on  $n \geq k + 1$  vertices. Let  $e_1, e_2$  and  $e_3$  be sequential edges in some Berge- $P_{m-1}$   $P$ . Let there be  $\alpha$  edges in  $P$  preceding  $e_1$  attached at a vertex  $x \in e_1$ , and  $\beta$  edges following  $e_3$  with  $\alpha \geq \beta \geq 0$ . Let  $y \in e_1 \cap e_2$  and  $z \in e_2 \cap e_3$ . Then one of the following is satisfied:*

(a) *There is a path of length at least  $\beta$  which is edge-disjoint from  $P$  that attaches to a vertex in  $e_2$*

or

(b) *there are paths of length at least  $\beta$  edge-disjoint from  $P$  attaching to at least  $k - 1$  vertices in  $X = (e_1 \cup e_3) \setminus e_2$ .*

*Proof.* For  $k = 3$ ,  $|X| = |(e_1 \cup e_3) \setminus e_2| = 4$  and the subpaths of  $P$  of length  $\alpha$  and  $\beta$  attach to  $k - 1 = 2$  vertices in  $X$ , so (b) holds.

Now consider,  $k \geq 4$ . Observe that due to the length of  $P$ ,  $\alpha + \beta + 3 = m - 2$ . Toward a contradiction, assume that there is no path of length at least  $\beta$  leaving  $e_2$  that is edge disjoint from  $P$ , and there are at least  $k$  vertices in  $(e_1 \cup e_3) \setminus \{y, z\}$  that do not have paths of length  $\beta$  attached to them. Consider the edge  $e$  made up of these  $k$  vertices as shown in Figure 4.9. Since  $H$  is saturated,  $H + e$  contains a Berge- $P_m$  that uses  $e$ .

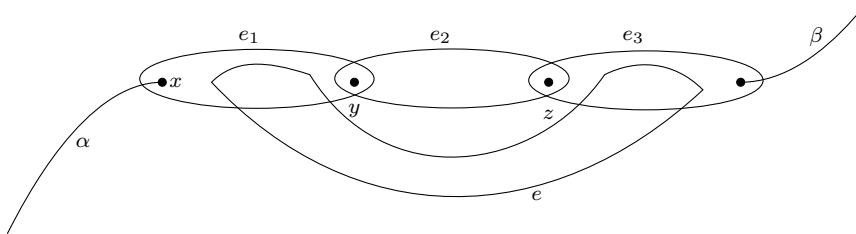


Figure 4.9: The graph  $H + e$  as described in Lemma 4.4.4.

This Berge- $P_m$  can use at most all four of the edges  $e_1, e_2, e_3$  and  $e$  and two paths that each start at some vertex in  $e_1 \cup e_2 \cup e_3$ . By the maximality of  $P$ , these two paths are

of length at most  $\alpha$ , and if one was of a length  $\gamma < \beta$ , then the longest path using  $e$  is of length at most  $\alpha + \gamma + 4 < \alpha + \beta + 4 = m - 1$ , a contradiction, so they are of length at least  $\beta$ . Thus, the Berge- $P_m$  cannot have used some path that entered at  $e_2$  since there are no paths of length  $\beta$  attached to a vertex in  $e_2$ . If the Berge- $P_m$  used two paths that both entered at  $e_1$  or both entered at  $e_3$ , then the edge,  $e_1$  or  $e_3$  respectively would have to be used twice since the vertices in  $e$  are not attached to paths of length  $\beta$ .

Then the remaining possibility is that the Berge- $P_m$  uses one path entering at  $e_1$ , and a second path leaving at  $e_3$ . By the maximality of  $P$ , the first path is of length at most  $\alpha$  and the second path at most  $\beta$ . Note that there is no way to traverse these paths while also using both  $e$  and  $e_2$ , so this path is of length at most  $\alpha + \beta + 3 < m - 1$ , a contradiction. In all cases we reach a contradiction, so either (a) or (b) occurs.  $\square$

**Corollary 4.4.5** *Let  $H$  be a 4-uniform Berge- $P_m$ -saturated linear tree on  $n \geq 10$  vertices. Let  $e_1, e_2$  and  $e_3$  be sequential edges in some Berge- $P_{m-1}$ , say  $P$ . Let there be  $\alpha$  edges in  $P$  preceding  $e_1$  attached at a vertex  $x$ , and  $\beta$  edges following  $e_3$  attached at a vertex  $y$  with  $\alpha \geq \beta \geq 0$ . Then there is a path of length at least  $\beta$ , edge disjoint from  $P$  that starts at a vertex in  $(e_1 \cup e_2 \cup e_3) \setminus \{x, y\}$ .*

*Proof.* Apply Lemma 4.4.4 to edges  $e_1, e_2$  and  $e_3$  in path  $P$ . If (a) is satisfied, we are done. If (b) is satisfied, then we have paths of length  $\beta$  leaving at least  $k - 1 = 3$  vertices in  $(e_1 \cup e_3) \setminus e_2$ . The subpaths of  $P$  of length  $\alpha$  and  $\beta$  can count for two of these, but there must be a third path on at least  $\beta$  edges, edge-disjoint from  $P$ , attaching to a vertex other than  $x$  or  $y$ , so in all cases we are done.  $\square$

**Lemma 4.4.6** *Let  $H$  be a 3-uniform Berge- $P_m$  saturated linear tree that contains a Berge- $P_{m-1}$  on  $n \geq 4$  vertices. Let  $e_1 = x_1x_2x_3$  and  $e_2 = x_3x_4x_5$  be sequential edges in some Berge- $P_{m-1}$ , say  $P$ , with  $\alpha$  edges preceding  $e_1$ , attaching at vertex  $x_1$  and  $\beta$  edges after  $e_2$ , attaching at  $x_5$  with  $\alpha \geq \beta$ . Then there exists a path of length at least  $\beta$  edge disjoint from*

$P$  that starts at a vertex in  $\{x_2, x_3, x_4\}$ .

*Proof.* By the length of  $P$ , we have that  $\alpha + \beta + 2 = m - 2$ . Let  $e = \{x_1, x_3, x_5\}$ . Then by saturation  $H + e$  contains a Berge- $P_m$  that uses the edge  $e$ . Observe that any such Berge- $P_m$  cannot traverse a path ending at  $x_1$ , then use all three edges  $e_1$ ,  $e_2$  and  $e$ , then traverse a path leaving  $x_5$  by the choice of  $e$ , so it must use a path starting at a vertex in  $\{x_2, x_3, x_4\}$ , say of length  $\gamma$ . Then this path can traverse a path of length at most  $\alpha$ , the three edges  $e_1$ ,  $e_2$  and  $e$ , then the path of length  $\gamma$ , so  $\alpha + \gamma + 3 \geq m - 1 = \alpha + \beta + 3$ , so  $\gamma \geq \beta$  as desired.  $\square$

### Branching with no Berge- $P_{m-1}$

If  $H$  does not contain a Berge- $P_{m-1}$ , then we get the same minimal amount of branching for all uniformities.

**Lemma 4.4.7** *Let  $H = (V, E)$  be a  $k$ -uniform Berge- $P_m$ -saturated linear tree with  $k \geq 3$ , and let  $e_1$  be an edge that is not in any Berge- $P_{m-1}$ . Let  $P$  be a longest path in  $H$  that contains  $e_1$ . Assume there are  $\alpha$  edges in  $P$  that precede  $e_1$  and attach via  $x \in e_1$  and assume there are  $\beta$  edges that follow  $e_1$ , and attach via  $y \in e_1$  with  $\alpha \geq \beta$ . Then there is a path of length  $\beta$  edge-disjoint from  $P$  attached to a vertex in  $e_1 \setminus \{x\}$*

*Proof.* Observe that by the length of  $P$ ,  $\alpha + \beta + 1 \leq m - 3$ . Assume to the contrary that we do not have a path of length  $\beta$  leaving  $e_1 \setminus \{x\}$ , edge-disjoint from  $P$ . Let  $e_2$  be the first edge along  $P$  after  $e_1$ . By the maximality of  $P$  and our assumption, the longest path leaving a vertex in  $e_1 \cup e_2 \setminus \{x\}$  is of length  $\beta - 1$ . Thus by Lemma 4.4.1, applied to  $e_1$  and  $e_2$ ,  $\alpha + \beta - 1 \geq m - 4$ , contradicting  $\alpha + \beta + 1 \leq m - 3$ .  $\square$

## 4.4.2 Counting Lemmas

Now that we have established the minimal amount of branching in a Berge- $P_m$ -saturated linear tree, we can begin counting the number of edges in a saturated tree  $H$ . Given an edge  $e$  that is not in the center of  $H$ , these counting lemmas will give a lower bound on the number of edges in the component of  $H - e$  that does not contain the center of  $H$ . As with the branching lemmas we need to separate into cases based on if  $H$  contains a Berge- $P_{m-1}$  or not.

### Counting with a Berge- $P_{m-1}$

As with the branching lemmas, when  $H$  contains Berge- $P_{m-1}$ , we must consider different cases for  $k = 3$ ,  $k = 4$  and  $k \geq 6$ .

**Lemma 4.4.8** *Let  $H$  be a 3-uniform Berge- $P_m$ -saturated linear tree on  $n \geq 4$  vertices. Let  $e$  be an edge of  $H$  and  $P$  be a Berge- $P_{m-1}$  through  $e$ . Let  $\alpha$  be the number of edges preceding  $e$  in  $P$  and  $\beta$  the number of edges after  $e$  attached at  $v \in e$  with  $\alpha \geq \beta$ . Let  $P'$  be the path on  $\beta$  edges starting at  $v$  contained in  $P$ . Let  $X$  be the set of all vertices  $x$  such that the path between  $x$  and  $v$  uses at least one edge of  $P'$ . Then*

$$|E(X \cup \{v\})| \geq \begin{cases} 2^{\ell+1} - 2 & \text{if } \beta = 2\ell \\ 2^{\ell+1} + 2^\ell - 2 & \text{if } \beta = 2\ell + 1. \end{cases}$$

*Proof.* We will proceed by strong induction on  $\beta$ . For  $0 \leq \beta \leq 2$  the result holds simply counting the edges on  $P'$ . Now let  $\beta \geq 3$  and assume that the result holds for all paths of length less than  $\beta$ .

Let  $e_1$  and  $e_2$  be the first two edges of the path  $P'$  starting at  $v \in e$ , in that order. Then there is a Berge- $P_{m-1}$  containing  $e_1$  and  $e_2$  with  $\alpha + 1$  edges preceding  $e_1$  and  $\beta - 2$  edges after  $e_2$  with  $\alpha + 1 > \beta - 2$ . Then by Lemma 4.4.6 there is a path of length at least

$\beta - 2$  leaving  $e_1$  or  $e_2$ , edge disjoint from the Berge- $P_{m-1}$ . Then we have two paths of length  $\beta - 2$  leaving  $e_1 \cup e_2$ , and  $e_1$  and  $e_2$  all in  $X$ . By induction, this gives us that if  $\beta = 2\ell$ , then

$$|X| \geq 2(2^\ell - 2) + 2 = 2^{\ell+1} - 2,$$

and if  $\beta = 2\ell + 1$ ,

$$|X| \geq 2(2^\ell + 2^{\ell-1} - 2) + 2 = 2^{\ell+1} + 2^\ell - 2.$$

□

**Lemma 4.4.9** *Let  $H$  be a 4-uniform Berge- $P_m$ -saturated linear tree on  $n \geq 5$  vertices. Let  $e$  be an edge of  $H$  and let  $P$  be a Berge- $P_{m-1}$  that uses  $e$ . Assume there are  $\alpha$  edges preceding  $e$  in  $P$  and  $\beta$  edges following  $e$  attached at  $v \in e$  with  $\alpha \geq \beta$ . Let  $P'$  be the path on  $\beta$  edges starting at  $v$  contained in  $P$ . Let  $X$  be the set of all vertices  $x$  such that the path between  $x$  and  $v$  uses at least one edge of  $P'$ . Then*

$$|E(X \cup \{v\})| \geq \begin{cases} 2^{\ell+1} + 2^\ell - 3 & \text{if } \beta = 3\ell \\ 2^{\ell+2} - 3 & \text{if } \beta = 3\ell + 1 \\ 2^{\ell+2} + 2^\ell - 3 & \text{if } \beta = 3\ell + 2. \end{cases}$$

*Proof.* We will proceed by strong induction based on the value of  $\beta$  modulo 3. For  $0 \leq \beta \leq 3$  the result holds simply counting the edges on  $P'$ . Let  $\beta \geq 4$  and assume the result holds for all paths of shorter length.

Let  $e_1, e_2$  and  $e_3$  be the first three edges of  $P'$  in that order. Then there is a Berge- $P_{m-1}$  containing  $e_1, e_2$  and  $e_3$  with  $\alpha + 1$  edges preceding  $e_1$  and  $\beta - 3$  edges after  $e_3$  with  $\alpha + 1 > \beta - 3$ . By Corollary 4.4.5 there is a path of length at least  $\beta - 3$  leaving a vertex in  $e_1 \cup e_2 \cup e_3$ , edge disjoint from the Berge- $P_{m-1}$ . Then  $X$  contains two paths of length  $\beta - 3$  leaving  $e_1 \cup e_2 \cup e_3$ , and also the edges  $e_1, e_2$  and  $e_3$ . By induction, this gives us that

if  $\beta = 3\ell$ , then

$$|X| \geq 2(2^\ell + 2^{\ell-1} - 3) + 3 = 2^{\ell+1} + 2^\ell - 3.$$

If  $\beta = 3\ell + 1$ , then

$$|X| \geq 2(2^{\ell+1} - 3) + 3 = 2^{\ell+2} - 3.$$

If  $\beta = 3\ell + 2$ , then

$$|X| \geq 2(2^{\ell+1} + 2^{\ell-1} - 3) + 3 = 2^{\ell+2} + 2^\ell - 3.$$

□

**Lemma 4.4.10** *Let  $k \geq 6$ . Let  $H$  be a  $k$ -uniform Berge- $P_m$ -saturated linear tree on  $n \geq k+1$  vertices. Let  $e$  be an edge and  $P$  a Berge- $P_{m-1}$  in  $H$  that uses  $e$ . Assume there are  $\alpha$  edges preceding  $e$  in  $P$  and  $\beta$  edges following  $e$  attached at  $v \in e$  with  $\alpha \geq \beta$ . Let  $P'$  be the path on  $\beta$  edges starting at  $v$  contained in  $P$ . Let  $X$  be the set of all vertices  $x$  such that the path between  $x$  and  $v$  uses at least one edge of  $P'$ . Then*

$$|E(X \cup \{v\})| \geq \begin{cases} 2^{\ell+1} - 2 & \text{if } \beta = 2\ell \\ 2^{\ell+1} + 2^{\ell-1} - 2 & \beta = 2\ell + 1. \end{cases}$$

*Proof.* For  $1 \leq \beta \leq 3$  the result holds simply counting the edges on  $P'$ . Let  $\beta = 4$ . We want to show that we have at least 6 edges. Let  $e_1, e_2, e_3$  and  $e_4$  be the edges in  $P'$  and assume there is at most one more edge,  $e_5$ , incident with these edges. Since  $\beta = 4$ ,  $e_5$  is not incident to  $e_4$ . If  $e_5$  is incident to  $w$  in  $e_1$  or  $e_3$ , or if there is no fifth edge  $e_5$ , then adding the edge  $e'$  which consists of  $k$  vertices of degree 1 in  $(e_1 \cup e_2) \setminus \{w\}$  does not create a Berge- $P_m$  in  $H + e'$ . If  $e_5$  is incident to a degree 1 vertex,  $w$ , in  $e_2$  then adding the edge  $e' = e_5 \setminus \{w\} \cup \{u\}$  where  $u$  is another degree 1 vertex in  $e_2$  does not create a Berge- $P_m$  in  $H + e'$  contradicting the assumption that  $H$  is Berge- $P_m$ -saturated. Thus we have at least 6 edges, so the result holds for  $\beta = 4$ .

Let us proceed by strong induction on  $\beta$  based on parity. Let  $\beta \geq 5$  and assume the result holds for paths of shorter length.

Let  $e_1, e_2$  and  $e_3$  be the first three edges in  $P'$ . There is a path of length  $\beta - 3$  leaving  $e_3$  as shown in Figure 4.10. Then by Lemma 4.4.4 there is either a path of length  $\beta - 3$  leaving  $e_2$  or there are at least  $k - 2 \geq 4$  paths of length  $\beta - 3$  leaving  $(e_1 \cup e_3) \setminus (e_2 \cup \{v\})$ . If there are  $k - 2$  paths of length  $\beta - 3$  leaving  $(e_1 \cup e_3) \setminus (e_2 \cup \{v\})$ , then by our inductive hypothesis we have

$$|E(X \cup \{v\})| \geq \begin{cases} (k - 2)(2^{\ell-1} + 2^{\ell-3} - 2) + 3 & \text{if } \beta = 2\ell \\ (k - 2)(2^\ell - 2) + 3 & \text{if } \beta = 2\ell + 1. \end{cases}$$

Note that since  $k \geq 6$  and  $\ell \geq 2$ , we have

$$|E(X \cup \{v\})| \geq \begin{cases} 2^{\ell+1} + 2^{\ell-1} - 5 & \text{if } \beta = 2\ell \\ 2^{\ell+2} - 5 & \text{if } \beta = 2\ell + 1. \end{cases}$$

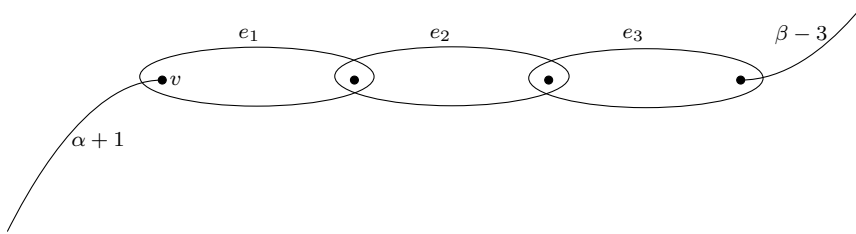


Figure 4.10: The edges  $e_1, e_2$  and  $e_3$  as described in Lemma 4.4.10.

If we do not have these  $k - 2$  paths, then there is a path of length  $\beta - 3$  leaving  $e_2$ . If this path cannot be extended to a path of length  $\beta - 2$ , then the edges in it are in no Berge- $P_{m-1}$  so by Lemma 4.4.11 this path contributes at least  $2^{\beta-3} - 1$  edges. In addition to this, we have the path of length  $\beta - 3$  leaving  $e_3$ . Then by our inductive hypothesis, we

have

$$|E(X \cup \{v\})| \geq \begin{cases} 2^{2\ell-3} - 1 + 2^{\ell-1} + 2^{\ell-3} - 2 + 3 = 2^{2\ell-3} + 2^{\ell-1} + 2^{\ell-3} & \text{if } \beta = 2\ell \\ 2^{2\ell-2} - 1 + 2^\ell - 2 + 3 = 2^{2\ell-2} + 2^\ell & \text{if } \beta = 2\ell + 1. \end{cases}$$

If instead the path can be extended to a path of length  $\beta - 2$ , then this path and the path of length  $\beta - 3$  give us two paths of length  $\beta - 2$  leaving  $e_2$ , which give us

$$|E(X \cup \{v\})| \geq \begin{cases} 2(2^\ell - 2) + 2 = 2^{\ell+1} - 2 & \text{if } \beta = 2\ell \\ 2(2^\ell + 2^{\ell-2} - 2) + 2 = 2^{\ell+1} + 2^{\ell-1} - 2 & \text{if } \beta = 2\ell + 1. \end{cases}$$

It is easy to check that this final case is minimal for all  $\beta \geq 5$ , giving us the result.  $\square$

### Counting with no Berge- $P_{m-1}$

Since there is only one branching lemma for all uniformities when  $H$  does not contain a Berge- $P_{m-1}$ , there is also only one counting lemma for this case.

**Lemma 4.4.11** *Let  $k \geq 3$ . Let  $H$  be a  $k$ -uniform Berge- $P_m$ -saturated linear tree on  $n \geq k+1$  vertices. Let  $e$  be an edge and  $P$  be a Berge- $P_{m-2}$  that is the longest path using  $e$ . Assume there are  $\alpha$  edges preceding  $e$  in  $P$  and  $\beta$  edges following  $e$  attached at  $v \in e$  with  $\alpha \geq \beta$ . Let  $P'$  be the path on  $\beta$  edges starting at  $v$  contained in  $P$  and let  $e_1$  be the first edge of  $P'$  (so  $v \in e_1$ ). Let  $X$  be the set of all vertices  $x$  such that the path between  $x$  and  $v$  uses at least one edge of  $P'$ . If  $P'$  is the longest path that does not use edge  $e$  but contains  $e_1$ , then the number of edges in  $X \cup \{v\}$  is at least  $2^\beta - 1$ .*

*Proof.* Let us proceed via induction. For  $\beta = 1$  the result is trivial. Let  $\beta \geq 1$  and assume the result is true for paths of length  $\beta - 1$ . Then, by Lemma 4.4.7, we have a path edge-disjoint from  $P'$  attached to a vertex in  $e_1$  of length  $\beta - 1$ . By our inductive hypothesis this



path contributes at least  $2^{\beta-1} - 1$  edges. Add this to the  $2^{\beta-1} - 1$  edges given by the path  $P' - e_1$  of length  $\beta - 1$  and the one edge  $e_1$ , we get that the number of edges is at least  $2(2^{\beta-1} - 1) + 1 = 2^\beta - 1$ .  $\square$

### 4.4.3 Central Structure and Final Count

With branching lemmas and counting lemmas in hand, we now must determine the central structure of a Berge- $P_{m-1}$ -saturated linear tree with a minimum number of edges looks like. We can then use this central structure and the counting lemmas to get a lower bound on the size of these trees. We first consider the case when  $H$  has no Berge- $P_{m-1}$ , and show that this is not optimal. We then give minimal edge counts for general Berge- $P_m$ -saturated trees.

#### Final count with no Berge- $P_{m-1}$

The following lemma gives a lower bound for a tree  $H$  with no Berge- $P_{m-1}$ . We will see that this case is far from optimal.

**Lemma 4.4.12** *Let  $m \geq 10$  and  $k \geq 3$ . Let  $H$  be a Berge- $P_m$  saturated linear tree on at least  $k + 1$  vertices and let  $H$  contain no Berge- $P_{m-1}$ . Then*

$$|E(H)| \geq \begin{cases} 3(2^{\ell-2}) - 2 & \text{if } m = 2\ell \\ 2^\ell - 2 & \text{if } m = 2\ell + 1. \end{cases}$$

*Proof.* By Observation 4.4.2 there exists some longest path  $P$  that is a Berge- $P_{m-2}$  in  $H$ .

First consider the case where  $m = 2\ell$  is even. Then the longest path is of odd length  $m - 3 = 2\ell - 3$ . Let  $e$  be the central edge of this path. By Lemma 4.4.7 with  $\alpha = \beta = \ell - 2$ , we have three paths of length  $\ell - 2$  leaving  $e$ . By Lemma 4.4.11 the existence of each path

guarantees at least  $2^{\ell-2} - 1$  edges, so adding these to the one edge  $e$ , we have

$$|E(H)| \geq 3(2^{\ell-2} - 1) + 1 = 3(2^{\ell-2}) - 2.$$

Now consider the case when  $m = 2\ell + 1$  is odd. Then the longest path is of even length  $m - 3 = 2\ell - 2$ . Let  $e_1$  and  $e_2$  be the two middle edges of  $P$ . These each have a path of length  $\ell - 2$  leaving them. By Lemma 4.4.7, there are either two paths of length at least  $\ell - 2$  edge-disjoint from  $P$  attached to  $e_1$  and  $e_2$ , call this situation (i) (see Figure 4.11), or one path of length at least  $\ell - 2$  attached to  $e_1 \cap e_2$ , call this situation (ii) (see Figure 4.12).

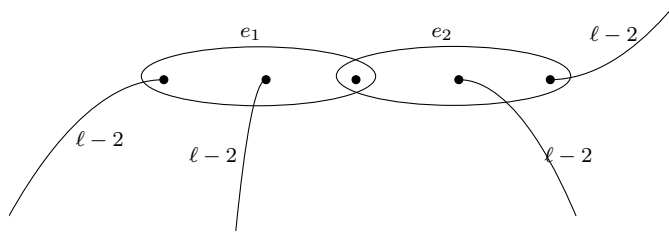


Figure 4.11: Edges  $e_1$  and  $e_2$  when  $m = 2\ell + 1$  in situation (i) of Lemma 4.4.12.

If we are in situation (i), Lemma 4.4.11 guarantees that each of these four paths contribute at least  $2^{\ell-2} - 1$  edges. These paths along with the two edges  $e_1$  and  $e_2$  gives us that

$$|E(H)| \geq 4(2^{\ell-2} - 1) + 2 = 2^\ell - 2.$$

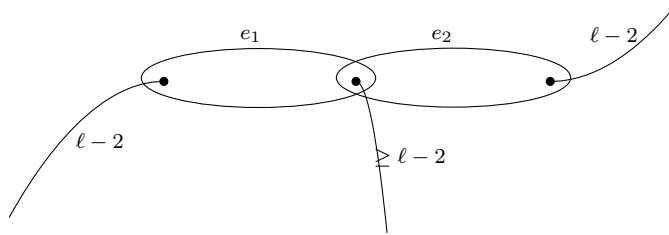


Figure 4.12: Edges  $e_1$  and  $e_2$  when  $m = 2\ell + 1$  in situation (ii) of Lemma 4.4.12.

If we are in situation (ii), Lemma 4.4.3 gives us that the edges in the path,  $P'$  attached

to  $e_1 \cap e_2$  must be contained in a Berge- $P_{m-2}$ , so  $P'$  must have length at least  $\ell - 1$ . Now by Lemma 4.4.11, the two paths of length  $\ell - 2$  contribute  $2^{\ell-2} - 1$  edges while the one path of length  $\ell - 1$  contributes  $2^{\ell-1} - 1$  edges. Thus we have that

$$|E(H)| \geq 2(2^{\ell-2} - 1) + 2^{\ell-1} - 1 + 2 = 2^\ell - 1 > 2^\ell - 2.$$

□

### Final count with a Berge- $P_{m-1}$

This section provides lower bounds of edge counts for general Berge- $P_m$ -saturated trees for uniformities  $k = 3$ ,  $k = 4$  and  $k \geq 6$ .

**Theorem 4.4.13** *Let  $H$  be a 3-uniform Berge- $P_m$ -saturated linear tree with at least 4 vertices. For  $m \geq 10$ , let  $m = 4s + r$  with  $1 \leq r \leq 4$ . Then*

$$|E(H)| \geq (3 + r)2^s - 5.$$

*Proof.* First, observe that if  $H$  does not contain a Berge- $P_{m-1}$ , then Lemma 4.4.12 implies our result. Therefore, we will assume that  $H$  does contain a Berge- $P_{m-1}$ .

Let  $m = 2\ell$ . Let  $P$  be a Berge- $P_{m-1}$  in  $H$ . Then  $P$  is of length  $2\ell - 2$ . Let  $e_1$  and  $e_2$  be the two central edges of  $P$ . Then there are  $\ell - 2$  edges preceding  $e_1$  and following  $e_2$ , so by Lemma 4.4.6, we have that there must be a third path of length at least  $\ell - 2$  attached to a vertex in  $e_1$  or  $e_2$ , say  $e_1$ . Then just considering paths away from  $e_1$ , we have a path of length  $\ell - 1$  and two paths of length  $\ell - 2$ , all edge disjoint and not using the edge  $e_1$ .

If  $m = 4s + 4$ , then  $\ell = 2s + 2$ , so  $\ell - 1 = 2s + 1$  is odd and  $\ell - 2 = 2s$  is even. Then by Lemma 4.4.8, from the path of length  $\ell - 1$  we have at least  $2^{s+1} + 2^s - 2$  edges and from the two paths of length  $\ell - 2$ , we have at least  $2(2^{s+1} - 2)$  edges. In addition to these, we

have  $e_1$ . Thus we get that

$$|E(H)| \geq 1 + 2^{s+1} + 2^s - 2 + 2(2^{s+1} - 2) = 7(2^s) - 5$$

If  $m = 4s + 2$ , then  $\ell = 2s + 1$ , so  $\ell - 1 = 2s$  is even and  $\ell - 2 = 2(s - 1) + 1$  is odd. Then by Lemma 4.4.8, from the path of length  $\ell - 1$  we have at least  $2^{s+1} - 2$  edges and from the two paths of length  $\ell - 2$ , we have at least  $2^s + 2^{s-1} - 2$  edges. In addition to these, we have  $e_1$ . Thus we get that

$$|E(H)| \geq 1 + 2^{s+1} - 2 + 2(2^s + 2^{s-1} - 2) = 5(2^s) - 5.$$

Now let  $m = 2\ell + 1$ . Let  $P$  be a Berge- $P_{m-1}$  in  $H$ . Then  $P$  is of length  $2\ell - 1$ . Let  $e_1$  be the central edge in  $P$ , and let  $e_2$  be the edge immediately after  $e_1$  in  $P$ . Then there are  $\ell - 1$  edges preceding  $e_1$  and  $\ell - 2$  edges after  $e_2$ , so by Lemma 4.4.6 there is another path of length at least  $\ell - 2$  coming from a vertex in either  $e_1$  or  $e_2$ .

If this path comes from the edge  $e_2$ , then  $e_2$  has a path of length  $\ell$  and two paths of length  $\ell - 2$  coming from it, all edge disjoint and not using the edge  $e_2$ . Call this situation (i).

If instead this path is coming from  $e_1$ , then  $e_1$  has a path of length at least  $\ell - 2$  and two paths of length  $\ell - 1$  coming from it, all edge disjoint and not using the edge  $e_1$ . It may end up that there are actually three paths of length  $\ell - 1$  leaving  $e_1$ . If this is the case, call it situation (ii).

If we are not in situation (i) or (ii), then there are two paths of length exactly  $\ell - 1$  leaving  $e_1$  and the third longest path is of length exactly  $\ell - 2$ . Observe that the edges of this last path are not in any Berge- $P_{m-1}$  in  $H$ . Call this situation (iii).

If  $m = 4s + 1$  and we have situation (i), then  $\ell = 2s$  and  $\ell - 2 = 2(s - 1)$  are both even. Then by Lemma 4.4.8, from the path of length  $\ell$  we have at least  $2^{s+1} - 2$  edges and

from the two paths of length  $\ell - 2$ , we have at least  $2^s - 2$  edges. In addition to these, we have  $e_2$ . Thus we get that

$$|E(H)| \geq 1 + 2^{s+1} - 2 + 2(2^s - 2) = 4(2^s) - 5$$

If  $m = 4s + 1$  and we have situation (ii), then  $\ell - 1 = 2(s - 1) + 1$  is odd. Then by Lemma 4.4.8 from each path of length  $\ell - 1$ , we have at least  $2^s + 2^{s-1} - 2$  edges, and we also have the edge  $e_1$ . Thus

$$|E(H)| \geq 1 + 3(2^s + 2^{s-1} - 2) = 2^{s+1} = 2^{s+2} + 2^{s-1} - 5$$

If  $m = 4s + 1$  and we have situation (iii), then  $\ell - 1 = 2(s - 1) + 1$  is odd. Then by Lemma 4.4.8 from each path of length  $\ell - 1$ , we have at least  $2^s + 2^{s-1} - 2$  edges. Furthermore, since the path of length  $\ell - 2 = 2s - 2$  is not in any Berge- $P_{m-1}$ , by Lemma 4.4.11 we have at least  $2^{2s-2} - 1$  edges. We also have the edge  $e_1$ , giving us

$$|E(H)| \geq 1 + 2(2^s + 2^{s-1} - 2) + 2^{2s-2} - 1 = 2^{2s-2} + 2^{s+1} + 2^s - 4.$$

Thus for  $m = 4s + 1$ , since  $m \geq 10$ , implying  $s \geq 2$ , we get the lowest bound from situation (i), giving our result.

Now let us consider  $m = 4s + 3$ . If we have situation (i), then  $\ell = 2s + 1$  and  $\ell - 2 = 2(s - 1) + 1$  are odd. Then by Lemma 4.4.8, from the path of length  $\ell$  we have at least  $2^{s+1} + 2^s - 2$  edges and from each path of length  $\ell - 2$ , we have at least  $2^s + 2^{s-1} - 2$  edges. In addition to these, we have  $e_2$ . Thus we get that

$$|E(H)| \geq 1 + 2^{s+1} + 2^s - 2 + 2(2^s + 2^{s-1} - 2) = 6(2^s) - 5$$

If  $m = 4s + 3$  and we have situation (ii),  $\ell - 1 = 2s$  is even. Then by Lemma 4.4.8, we have at least  $2^{s+1} - 2$  edges from each of the three paths of length  $\ell - 1$ . In addition to these, we have  $e_1$ . Thus we get that

$$|E(H)| \geq 1 + 3(2^{s+1} - 2) = 6(2^s) - 5$$

If  $m = 4s + 3$  and we have situation (iii), then  $\ell - 1 = 2s$  is even. Then by Lemma 4.4.8 from each path of length  $\ell - 1$ , we have at least  $2^{s+1} - 2$  edges. Furthermore, since the path of length  $\ell - 2 = 2s - 1$  is not in any Berge- $P_{m-1}$ , by Lemma 4.4.8 we have at least  $2^{2s-1} - 1$  edges. We also have the edge  $e_1$ , giving us

$$|E(H)| \geq 1 + 2(2^{s+1} - 2) + 2^{2s-1} - 1 = 2^{2s-1} + 2^{s+2} - 4.$$

Thus in all cases we have that for  $m = 4s + 3$ , since  $m \geq 10$ , and consequently  $s \geq 2$ ,  $|E(H)| \geq 6(2^s) - 5$ .

□

**Theorem 4.4.14** *Let  $H$  be a 4-uniform Berge- $P_m$  saturated linear tree with at least 5 vertices. For  $m \geq 10$ , let  $0 \leq r \leq 5$  such that  $m = 6s + r$ . Then*

$$|E(H)| \geq (6 + r)2^s - 8.$$

*Proof.* If  $H$  does not contain a Berge- $P_{m-1}$ , then Lemma 4.4.12 implies our result, so we will assume that  $H$  does have a Berge- $P_{m-1}$ .

Now, whenever counting edges, if we have a path away from the center of length exactly  $\beta$  whose edges are not in any Berge- $P_{m-1}$ , then by Lemma 4.4.11 we get that this path contributes at least  $2^\beta - 1$ . If instead this path was length  $\beta + 1$  with edges in some Berge- $P_{m-1}$ , then by Lemma 4.4.9, the path contributes  $2^{\beta+1} - 3$  edges if  $\beta = 3\ell$ ,  $2^{\beta+2} + 2^\ell - 3$

edges if  $\beta = 3\ell + 1$  and  $2^{\ell+2} + 2^{\ell+1} - 3$  edges if  $\beta = 3\ell + 2$ . Observe that if  $\beta \geq 2$ , we get a lower or equal count in all cases with the path of length  $\beta + 1$ . Thus any time we have a long path in no Berge- $P_{m-1}$ , we are justified in counting its contribution as if it were longer and in a Berge- $P_{m-1}$ . This fact is crucial to the argument and will be used many times, so we will refer to the argument in this paragraph as (\*).

Now, let us assume  $m = 2\ell$  is even. Let  $P$  be a Berge- $P_{m-1}$  in  $H$ . Note that  $P$  has even length  $2\ell - 2$ . Let  $e_1, e_2, e_3$  and  $e_4$  be the four central edges of  $P$ , appearing in that order. Then there are paths of length  $\ell - 3$  leaving some vertex in  $e_1$  and some vertex in  $e_4$ . Applying Corollary 4.4.5 on edges  $e_2, e_3$  and  $e_4$ , we get that there is a path of length at least  $\ell - 3$  leaving a vertex in  $e_2 \cup e_3 \cup e_4$ .

If this path leaves  $e_4 \setminus e_3$ , then applying Corollary 4.4.5 to  $e_1, e_2$  and  $e_3$ , we get another path of length  $\ell - 3$  away from center. Then we have four paths of length  $\ell - 3$  and four central edges. Call this situation (i).

If instead this path leaves a vertex in the symmetric difference  $e_2 \Delta e_3$ , then we may assume it is of length  $\ell - 2$  since  $\ell \geq 5$ . Assume without loss of generality this path leaves a vertex in  $e_3 \setminus e_2$ . Let  $e_0$  be the edge preceding  $e_1$  in  $P$ . Then  $e_0$  has a path of length  $\ell - 4$  leaving it. Applying Corollary 4.4.5 to edges  $e_0, e_1$  and  $e_2$  gives us a second path of length at least  $\ell - 4$ . Instead of considering  $e_4$  with a path of length  $\ell - 3$  away from  $e_4$ , we will instead consider this as a path of length  $\ell - 2$  away from  $e_3$ . Then we have two paths of length  $\ell - 2$ , two paths of length  $\ell - 4$  and four central edges. Call this situation (ii).

Finally if the path leaves the vertex in  $e_2 \cap e_3$ , the path must be of length at least  $\ell - 2$  for otherwise the edges would not be in a Berge- $P_{m-2}$ , contradicting Lemma 4.4.3. Now, let  $e_0$  be the edge preceding  $e_1$  in  $P$  and  $e_5$  be the edge after  $e_4$  in  $P$ . Then each of these edges has paths of length  $\ell - 4$  leaving them. Further, let  $e_1^*, e_2^*$  and  $e_3^*$  be the first three edges of the path leaving the vertex in  $e_2 \cap e_3$ . Then there is a path of length at least  $\ell - 5$  leaving  $e_3^*$ . Applying Corollary 4.4.5 to the three triplets  $e_0, e_1, e_2$ ;  $e_3, e_4, e_5$  and  $e_1^*, e_2^*, e_3^*$  gives us

two new paths of length  $\ell - 4$  and a path of length  $\ell - 5$ . Observe that if the paths of length  $\ell - 5$  are exactly length  $\ell - 5$ , then their edges are in no Berge- $P_{m-1}$  since the path is at distance at most 3 from the center, so the path, three edges, and a path of length  $\ell - 1$  gives us a length of at most  $\ell - 5 + 3 + \ell - 1 = 2\ell - 3 = m - 3$ . Thus if  $m \geq 14$ ,  $\ell - 5 \geq 2$ , and by (\*) we may count these paths as if they are length  $\ell - 4$ . In this case, we have nine edges in the center with six paths of length at least  $\ell - 4$ . Call this situation (iii).

If in this same setup though,  $m = 10$  or  $m = 12$ , we cannot assume this path is of length  $\ell - 4$ . In this case though, we still have the nine central edges and four paths of length  $\ell - 4 \geq 1$ . This gives us 13 edges, which for both  $m = 10$  and  $m = 12$  exceeds our result, so we are safe in assuming that in situation (iii),  $m \geq 14$ .

These three situations exhaust all possibilities for paths given by the first use of Corollary 4.4.5.

In each of these situations, we can use Lemma 4.4.9 to count how many edges are guaranteed by each path away from the center. To do so, we need to consider each situation in terms of the residue of  $m$  modulo 6. Let  $m = 6s + r$ . Then for even  $r$ , the edge counts for each situation is summarized in the table below:



r	Situation	Calculation	Edge Count
0	i	$4(2^s + 2^{s-1} - 3) + 4$	$6(2^s) - 8$
0	ii	$2(2^{s+1} - 3) + 2(2^s + 2^{s-2} - 3) + 4$	$6(2^s) + 2^{s-1} - 8$
0	iii	$6(2^s + 2^{s-2} - 3) + 9$	$7(2^s) + 2^{s-1} - 9$
2	i	$4(2^{s+1} - 3) + 4$	$8(2^s) - 8$
2	ii	$2(2^{s+1} + 2^{s-1} - 3) + 2(2^s + 2^{s-1} - 3) + 4$	$8(2^s) - 8$
2	iii	$6(2^s + 2^{s-1} - 3) + 9$	$9(2^s) - 9$
4	i	$4(2^{s+1} + 2^{s-1} - 3) + 4$	$10(2^s) - 8$
4	ii	$2(2^{s+1} + 2^s - 3) + 2(2^{s+1} - 3) + 4$	$10(2^s) - 8$
4	iii	$6(2^{s+1} - 3) + 9$	$12(2^s) - 9$

Taking the minimal value for each even  $r$  gives our result for even  $m$ .

Now let us consider odd  $m = 2\ell + 1$ . Let  $P$  be a Berge- $P_{m-1}$  in  $H$ . Note that  $P$  has odd length  $2\ell - 1$ . Let  $e_1, e_2, e_3, e_4$  and  $e_5$  be the five central edges of  $P$ , appearing in that order. Then there are paths of length  $\ell - 3$  leaving some vertex in  $e_1$  and some vertex in  $e_5$ . Applying Corollary 4.4.5 to edges  $e_3, e_4$  and  $e_5$ , we get that there is a path of length at least  $\ell - 3$  leaving a vertex in  $e_3 \cup e_4 \cup e_5$ .

If this path leaves a vertex in  $e_5$ , then we can apply Corollary 4.4.5 on edges  $e_2, e_3$  and  $e_4$  to get a path of length at least  $\ell - 2$  leaving a vertex in  $e_2 \cup e_3 \cup e_4$ . Observe that this path is not the same path found on the previous use of the corollary since the corollary guarantees that the path attaches at a vertex different from the two end vertices. If we then consider the edge  $e_1$  and the path away from it of length  $\ell - 3$  as a path of length  $\ell - 2$  away from  $e_2$ , we have two paths of length  $\ell - 2$ , two paths of length  $\ell - 2$  and four central edges. Call this situation (iv).

If the path instead leaves a vertex in  $e_4 \setminus e_3$ , then since  $\ell - 3 > 2$ , we can assume this path is of length  $\ell - 2$ . Applying Corollary 4.4.5 to edges  $e_1, e_2$  and  $e_3$  gives us another path

of length  $\ell - 3$ . If we then consider the edge  $e_5$  and the path away from it of length  $\ell - 3$  as a path of length  $\ell - 2$  away from  $e_4$ , we get a situation identical to situation (iv).

If the path leaves a vertex in  $e_3$ , then it is of length at least  $\ell - 2$  since otherwise the edges in it would be in no Berge- $P_{m-2}$ , contradicting Lemma 4.4.3. We can further assume by (\*), this path has length  $\ell - 1$  since otherwise it would be in no Berge- $P_{m-1}$  and  $\ell - 2 > 2$ . If we then consider the edges  $e_2, e_1$  and the path of length  $\ell - 3$  away from  $e_1$  as a path of length  $\ell - 1$  away from  $e_3$ , and similarly consider the edges  $e_4, e_5$  and the path of length  $\ell - 3$  away from  $e_5$  as a path of length  $\ell - 1$  away from  $e_3$ , then have three paths of length  $\ell - 1$  and one central edge. Let this be situation (v).

This exhausts all possibilities for the location of the path given by the first use of Corollary 4.4.5.

Now, let  $m = 6s + r$ . Similar to the even case, we can count the edges in each of these situations using Lemma 4.4.9 to count the edges given by each path away from center.

The counts are summarized in the table below for odd values of  $r$ :

r	Situation	Calculation	Edge Count
1	iv	$2(2^{s+1} - 3) + 2(2^s + 2^{s-1} - 3) + 4$	$7(2^s) - 8$
1	v	$3(2^{s+1} + 2^{s-1} - 3) + 1$	$7(2^s) + 2^{s-1} - 8$
3	iv	$2(2^{s+1} + 2^{s-1} - 3) + 2(2^{s+1} - 3) + 4$	$9(2^s) - 8$
3	v	$3(2^{s+1} + 2^s - 3) + 1$	$9(2^s) - 8$
5	iv	$2(2^{s+1} + 2^s - 3) + 2(2^{s+1} + 2^{s-1} - 3) + 4$	$11(2^s) - 8$
5	v	$3(2^{s+2} - 3) + 1$	$12(2^s) - 8$

Taking the minimal value for each  $r$  gives us our result. This completes the proof.  $\square$

**Theorem 4.4.15** *Let  $k \geq 6$  and  $m \geq 10$ . Let  $H$  be a Berge- $P_m$  saturated linear tree on at*

least  $k + 1$  vertices. Then

$$|E(H)| \geq \begin{cases} 2^{s+1} + 2^s + 2^{s-1} + 2^{s-2} - 6 & \text{if } m = 4s, \\ 2^{s+2} + 2^{s-1} - 6 & \text{if } m = 4s + 1, \\ 2^{s+2} + 2^s - 6 & \text{if } m = 4s + 2, \\ 2^{s+2} + 2^{s+1} + 2^{s-1} - 6 & \text{if } m = 4s + 3. \end{cases}$$

*Proof.* If  $H$  contains no Berge- $P_{m-1}$  then we are done by Lemma 4.4.12, so assume  $H$  contains a Berge- $P_{m-1}$ , say  $P$ .

Whenever we count edges, if we have a path away from the center of length exactly  $\beta$  whose edges are not in any Berge- $P_{m-1}$ , then by Lemma 4.4.11 we get that this path contributes at least  $2^\beta - 1$  edges. If instead this path was length  $\beta + 1$  with edges in some Berge- $P_{m-1}$ , then by Lemma 4.4.10, the path contributes  $2^{\ell+1} + 2^{\ell-1} - 2$  edges if  $\beta = 2\ell$  or  $2^{\ell+2} - 2$  edges if  $\beta = 2\ell + 1$ . Observe that if  $\beta \geq 2$ , we get a lower or equal count in all cases with the path of length  $\beta + 1$ . Thus any time we have a long path in no Berge- $P_{m-1}$ , we are justified in counting its contribution as if it were longer and in a Berge- $P_{m-1}$ . This argument will be used many times, so we will refer to it as  $(\star)$ .

Let  $m = 4s + r$ . Consider first when  $m = 2\ell$  is even, so  $r = 0$  or  $r = 2$ . Then the longest path in  $H$  is of even length  $2\ell - 2$ . Let  $e_1, e_2, e_3$  and  $e_4$  be the four central edges of  $P$ , appearing in that order. Then there are paths of length  $\ell - 3$  leaving some vertex in  $e_1$  and some vertex in  $e_4$ . Applying Lemma 4.4.4 on edges  $e_2, e_3$  and  $e_4$ , we get that there is either a path of length at least  $\ell - 3$  leaving  $e_3$  or there are  $k - 1$  paths of length  $\ell - 3$  leaving  $(e_2 \cup e_4) \setminus e_3$ .

If we have  $k - 1$  paths of length  $\ell - 3$  leaving  $(e_2 \cup e_4) \setminus e_3$ , we have a set of three edges with  $k - 1$  paths of length  $\ell - 3$  leaving them. We will call this situation (i).

If we are not in situation (i), then there must be a path of length at least  $\ell - 3$  leaving

$e_3$ . If this path attaches to  $P$  via the vertex in  $e_2 \cap e_3$ , then it is of length at least  $\ell - 2$  since otherwise the edges in this path would be in no Berge- $P_{m-2}$ , which contradicts Lemma 4.4.3. Furthermore, by  $(\star)$ , we can assume for counting purposes that this path is actually of length  $\ell - 1$ . Notice that in this case there are three paths of length  $\ell - 1$  all leaving the vertex in  $e_2 \cap e_3$ . Call this situation (ii) (see figure 4.13).

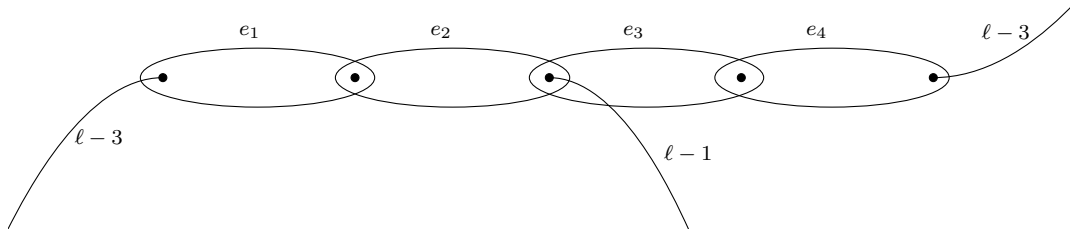


Figure 4.13: Edges  $e_1$ - $e_4$  when  $m = 2\ell$  in situation (ii) of Theorem 4.4.15.

If instead this path attaches to a vertex in  $e_3 \setminus e_2$ , we can apply Lemma 4.4.4 on edges  $e_1, e_2$  and  $e_3$ , and similarly to before either we are in situation (i) or we find another path of length at least  $\ell - 3$  attaching to a vertex in  $e_2$ . Assuming that we are not in situation (i), by  $(\star)$ , we can assume that the path attaching to a vertex in  $e_3 \setminus e_2$  and the path attaching to a vertex in  $e_2$  are of length  $\ell - 2$  so that these edges are in some Berge- $P_{m-1}$ .

Then, including the paths containing the edges  $e_1$  and  $e_4$ , there are at least four paths leaving vertices in  $e_2 \cup e_3$ , all of length  $\ell - 2$  in a Berge- $P_{m-1}$ . Call this situation (iii) (see figure 4.14).

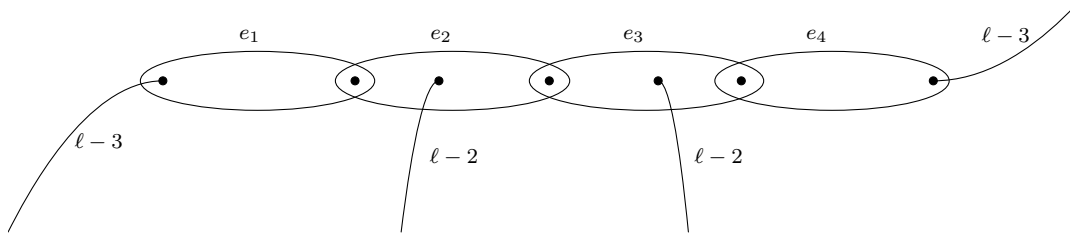


Figure 4.14: Edges  $e_1$ - $e_4$  when  $m = 2\ell$  in situation (iii) of Theorem 4.4.15.

These three situations exhaust all possibilities. Now for even  $m = 4s + r$ , we can use

Lemma 4.4.10 to count the edges contributed by the paths in each of these situations. The counts are summarized in the table below for even  $r$ :

r	Situation	Calculation	Edge Count
0	i	$(k-1)(2^{s-1} + 2^{s-3} - 2) + 3$	$(k-1)(2^{s-1} + 2^{s-3}) - 2k + 5$
0	ii	$3(2^s + 2^{s-2} - 2)$	$2^{s+1} + 2^s + 2^{s-1} + 2^{s-2} - 6$
0	iii	$4(2^s - 2) + 2$	$2^{s+2} - 6$
2	i	$(k-1)(2^s - 2) + 3$	$(k-1)(2^s) - 2k + 5$
2	ii	$3(2^{s+1} - 2) + 4$	$2^{s+2} + 2^{s+1} - 6$
2	iii	$4(2^s + 2^{s-2} - 2) + 2$	$2^{s+2} + 2^s - 6$

Since  $k \geq 6$ , situation (ii) gives the smallest edge count when  $r = 0$ , and situation (iii) gives the smallest when  $r = 2$ , giving us the result for  $m$  even.

Now let us consider when  $m = 2\ell + 1$  is odd. The longest path is of odd length  $2\ell - 1$ . Let  $e_1, e_2, e_3, e_4$  and  $e_5$  be the five central edges of a longest path appearing in that order. Then there are paths of length  $\ell - 3$  leaving  $e_1$  and  $e_5$ . If we apply Lemma 4.4.4 to edges  $e_1, e_2$  and  $e_3$ , we see that there is either a path of length at least  $\ell - 3$  leaving  $e_2$  or there are  $k - 1$  paths of length  $\ell - 3$  leaving  $(e_1 \cup e_3) \setminus e_2$ .

If we have the  $k - 1$  paths leaving  $(e_1 \cup e_3) \setminus e_2$ , we will call this situation (iv). If we are not in situation (iv), then we have a path of length at least  $\ell - 3$  leaving  $e_2$ . Now using Lemma 4.4.4 on the edges  $e_3, e_4$  and  $e_5$ , we either have  $k - 1$  paths of length  $\ell - 3$ , as in situation (iv) or one path of length  $\ell - 3$  leaving  $e_4$ .

Assuming we are not in situation (iv), we have a path leaving  $e_2$  and a path leaving  $e_4$ , each of length at least  $\ell - 3$ . If one of these paths attach at the vertex in either  $e_2 \cap e_3$  or  $e_3 \cap e_4$ , call this situation (v), and assume without loss of generality there is a path attached at  $e_2 \cap e_3$ . By  $(\star)$ , we can assume this path is actually of length  $\ell - 1$  since otherwise the edges in the path would be in no Berge- $P_{m-1}$ . Similarly, we can assume the path that attaches to a vertex in  $e_4$  is of length at least  $\ell - 2$ , as in Figure 4.15. Thus, in situation (v), we have

four paths leaving the two edges  $e_3$  and  $e_4$ , two of length  $\ell - 1$  and two of length  $\ell - 2$ .

If we are not in situation (v), then neither of the paths found by the first two uses of Lemma 4.4.4 attach in  $e_3$ . Thus if we apply Lemma 4.4.4 to the edges  $e_2$ ,  $e_3$  and  $e_4$ , we get that there must be either a path of length at least  $\ell - 2 \geq \ell - 3$  leaving  $e_3$ , which gives us situation (v), or we get at least  $k - 1$  paths of length at least  $\ell - 2 \geq \ell - 3$  leaving vertices in  $(e_2 \cup e_4) \setminus e_3$ , which gives us situation (iv). Thus these two situations are exhaustive.

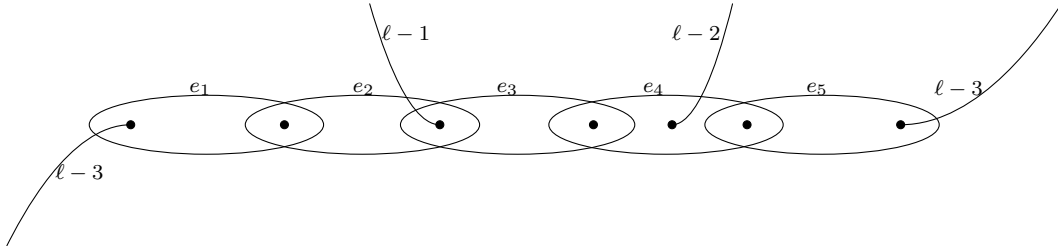


Figure 4.15: Edges  $e_1$ - $e_5$  when  $m = 2\ell + 1$  in situation (v) of Theorem 4.4.15.

Just as in the even case, we now use Lemma 4.4.10 to count the edges contributed by each path in each situation for odd  $m = 4s + r$ . This is summarized below:

r	Situation	Calculation	Edge Count
1	iv	$(k - 1)(2^{s-1} + 2^{s-3} - 2) + 3$	$(k - 1)(2^{s-1} + 2^{s-3}) - 2k + 5$
1	v	$2(2^s + 2^{s-2} - 2) + 2(2^s - 2) + 2$	$2^{s+2} + 2^{s-1} - 6$
3	iv	$(k - 1)(2^s - 2) + 3$	$(k - 1)(2^s) - 2k + 5$
3	v	$2(2^{s+1} - 2) + 2(2^s + 2^{s-2} - 2) + 2$	$2^{s+2} + 2^{s+1} + 2^{s-1} - 6$

Since  $k \geq 6$ , situation (v) gives the lowest bound for both  $r = 1$  and  $r = 3$ , completing the proof. □

## 4.5 Berge Saturation Numbers for $K_3$ , $C_m$ , $K_{1,m}$ , and $\ell K_2$

In this section we explore bounds on the Berge saturation numbers for many common classes of graphs. We begin with the saturation number for Berge matchings,  $\ell K_2$ .

**Theorem 4.5.1** *If  $k \geq 3$ , then  $\text{sat}_k(n, \text{Berge-}\ell K_2) = \ell - 1$  for  $n \geq k(\ell - 1)$ .*

*Proof.* Let  $H$  be a  $k$ -uniform Berge- $\ell K_2$ -saturated hypergraph with  $k \geq 3$  and  $n \geq k(\ell - 1)$ . First, note that any Berge- $\ell K_2$ -saturated hypergraph must have at least  $\ell - 1$  edges so  $|E(H)| \geq \ell - 1$ .

Now suppose that  $H$  is the hypergraph consisting of  $\ell - 1$  disjoint edges and  $n - (\ell - 1)k$  isolated vertices. Observe that for any edge  $e \in E(\overline{H})$ , we can find two vertices  $x, y \in e$  such that  $x$  and  $y$  are not adjacent in  $H$ . Then every edge in  $H$  contains a pair of vertices that are not  $x$  or  $y$ , so every edge in  $H$  and  $e$  can be used to create a copy of  $K_2$  disjoint from all the other copies of  $K_2$ . Thus  $H$  is Berge- $\ell K_2$ -saturated.  $\square$

Next, we will establish the saturation number for triangles. To do so, we need a definition. A connected component is *edge-minimal* if no connected hypergraph on the same number of vertices has fewer edges. Observe that an edge minimal component with  $n'$  vertices has  $\lceil \frac{n'-1}{k-1} \rceil$  edges. Now, the following lemma will be useful in the proof of  $\text{sat}_k(n, \text{Berge-}K_3)$ .

**Lemma 4.5.2** *Let  $k \geq 3$ . Let  $H$  be a Berge- $K_3$ -saturated  $k$ -uniform hypergraph. If  $H$  contains more than one component, then  $H$  does not have an edge-minimal component with  $n' > 1$  vertices such that  $(k - 1) \mid (n' - 1)$ .*

*Proof.* Assume  $H$  contains an edge-minimal component,  $H_1$ , with  $n' > 1$  vertices such that  $(k - 1) \mid (n' - 1)$ . Then, since  $H_1$  is connected with  $\frac{n'-1}{k-1}$  edges,  $H_1$  is a linear tree. Thus,  $H_1$  has at least one edge containing exactly  $k - 1$  vertices of degree 1, say  $v_1, v_2, \dots, v_{k-1}$ .

Let  $x$  be a vertex in another component of  $H$ . Then, adding the edge  $v_1v_2 \dots v_{k-1}x$  does not introduce a  $K_3$ . Thus,  $H$  is not Berge- $K_3$ -saturated.  $\square$

**Theorem 4.5.3** *Let  $k \geq 3$ . For all  $n \geq k + 1$ ,  $\text{sat}_k(n, \text{Berge-}K_3) = \lceil \frac{n-1}{k-1} \rceil$ .*

*Proof.* Let  $S = S_n^{(k)}$  be the  $k$ -uniform hypergraph with  $\lfloor \frac{n-1}{k-1} \rfloor$  edges that intersect only at a single vertex  $v$ , and if  $r = (n-1) \bmod (k-1) > 0$ , one more edge containing the remaining  $r$  vertices, the vertex  $v$ , and  $k-1-r$  other vertices, all from exactly one other edge of  $S$  (see Figure 4.16). Then  $S$  has  $\lceil \frac{n-1}{k-1} \rceil$  edges and is Berge- $K_3$ -free since a Berge- $K_3$  must have three edges, each of which contain two vertices of degree 2, but  $S$  only has two such edges. Further  $S$  is Berge- $K_3$ -saturated. Indeed, every edge  $e \in E(\overline{S})$  shares vertices that are not  $v$  with at least two distinct edges of  $S$ , say  $f$  and  $g$ , so  $S + e$  contains the Berge- $K_3$  with edges  $e, f$  and  $g$ . Hence,  $\text{sat}(n, \text{Berge-}K_3) \leq \lceil \frac{n-1}{k-1} \rceil$ .

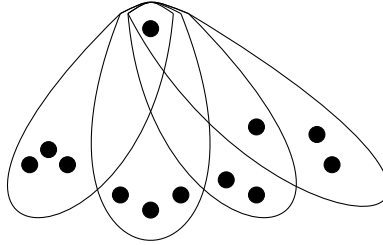


Figure 4.16: The Berge- $K_3$  saturated graph,  $S_{12}^{(4)}$ .

Now, let  $H$  be a minimal Berge- $K_3$ -saturated  $k$ -graph on  $n$  vertices. If  $H$  is connected, then  $H$  has at least  $\lceil \frac{n-1}{k-1} \rceil$  edges and we are done. Suppose now that  $H$  has  $j \geq 2$  components  $H_1, \dots, H_j$ , with  $H_i$  on  $n_i$  vertices for  $1 \leq i \leq j$ . By Lemma 4.5.2,  $H$  does not contain any edge minimal components  $H_i$  such that  $(k-1)|(n_i-1)$ .

By re-indexing the  $H_i$ 's if necessary, we can assume that  $H$  has  $\ell$  non-edge-minimal components where  $(k-1)|(n_i-1)$ ,  $H_1, \dots, H_\ell$ , that  $H$  has  $s$  non-edge-minimal components where  $(k-1) \nmid (n_i-1)$ ,  $H_{\ell+1}, \dots, H_{\ell+s}$ , that  $H$  has  $t$  isolated vertices,  $H_{\ell+s+1}, \dots, H_{\ell+s+t}$ , and



that  $H$  has  $j - \ell - s - t$  edge-minimal components where  $(k - 1) \nmid (n_i - 1)$ ,  $H_{\ell+s+t+1}, \dots, H_j$ .

Now we introduce a parameter that will help us compare the number of edges in  $H$  to the number of edges in  $S_n^{(k)}$ . For a  $k$ -uniform hypergraph  $G$ , let  $O_G := \sum_{v \in V(G)} (d(v) - 1)$ . This parameter will count the number of times a vertex is covered by more than one edge. Note that  $O_G = O_{S_{|V(G)|}^{(k)}} \pmod k$  since  $O_G = k|E(G)| - |V(G)|$  and  $O_{S_{|V(G)|}^{(k)}} = k|E(S_{|V(G)|}^{(k)})| - |V(G)|$ . Also note that for our construction  $S_n^{(k)}$  above, we have  $O_{S_n^{(k)}} \leq \lceil \frac{n-1}{k-1} \rceil + (k-2) - [(n-1) \pmod{(k-1)}]$ .

We will calculate a bound for  $O_{H_i}$  for each connected component  $H_i$  individually based on the following procedure: Let  $E^* \subseteq E(H_i)$  be a smallest set of edges such that the hypergraph  $H^* = (V(H_i), E^*)$  is connected. There is an ordering of the edge of  $E^*$  such that each edge except the first intersects some preceding edge at least once. Thus, each edge except one in this ordering must contribute at least one to the parameter  $O_{H_i}$ , and in fact if  $(k - 1) \nmid (n_i - 1)$ , the edges of  $E^*$  must give at least an additional contribution of  $k - 1 - [(n_i - 1) \pmod{(k - 1)}]$  since some edges must overlap preceding edges in more than one vertex. Finally, the edges in  $E(H_i) \setminus E^*$  will contribute  $k$  to this parameter for each edge.

For the components  $H_i$  of  $H$  we get the following bounds on  $O_{H_i}$ :

$$O_{H_i} \geq \begin{cases} \lceil \frac{n_i-1}{k-1} \rceil + k - 1 & \text{if } 1 \leq i \leq \ell \text{ (non-minimal, divisible),} \\ \lceil \frac{n_i-1}{k-1} \rceil + 2k - 2 - [(n_i - 1) \pmod{(k - 1)}] & \text{if } \ell + 1 \leq i \leq \ell + s \text{ (non-minimal, non-divisible),} \\ -1 & \text{if } \ell + s + 1 \leq i \leq \ell + s + t \text{ (isolated vertices),} \\ \lceil \frac{n_i-1}{k-1} \rceil + k - 2 - [(n_i - 1) \pmod{(k - 1)}], & \text{if } \ell + s + t + 1 \leq i \leq j \text{ (minimal, non-divisible).} \end{cases}$$

Let  $r_i = (n_i - 1) \pmod{(k - 1)}$  for  $1 \leq i \leq j$  and let  $\sum_{i=1}^j r_i = \alpha(k - 1) + r^*$  for some  $\alpha$  and  $r^* = \left( \sum_{i=1}^j r_i \right) \pmod{(k - 1)}$ . Note that the average remainder of a non-isolate component,  $\frac{1}{j-t} \sum_{i=1}^j r_i \geq \frac{\alpha(k-1)}{j-t}$ , but any remainder is less than  $k - 1$ , so  $\alpha < j - t$ .

Then, we have

$$\begin{aligned}
\sum_{i=1}^j O_{H_i} &\geq \sum_{i=1}^{\ell} \left( \left\lceil \frac{n_i-1}{k-1} \right\rceil - 1 + k \right) + \sum_{i=\ell+1}^{\ell+s} \left( \left\lceil \frac{n_i-1}{k-1} \right\rceil + k - 2 - r_i + k \right) \\
&\quad + \sum_{i=\ell+s+1}^{j-t} \left( \left\lceil \frac{n_i-1}{k-1} \right\rceil + k - 2 - r_i \right) - t \\
&= \sum_{i=1}^j \left\lceil \frac{n_i-1}{k-1} \right\rceil + \ell(k-1) + s(k-2+k) + (j-t-\ell-s)(k-2) - t - \sum_{i=1}^j r_i \\
&\geq \left\lceil \frac{n-j}{k-1} \right\rceil - \ell - t + (\ell+s)k + (j-\ell-t)(k-2) - \sum_{i=1}^j r_i \\
&\geq \left\lceil \frac{n-1}{k-1} \right\rceil - 1 + k - 2 - \left( \sum_{i=1}^j r_i + j - 1 - \alpha(k-1) \right) + j - 1 - \alpha(k-1) \\
&\quad + (j-1)(k-2) + sk + \ell - t(k-1) \\
&\geq O_{S_n^{(k)}} + (j - \alpha - t)(k-1) + \ell + k(s-1) \\
&\geq O_{S_n^{(k)}} + k - 1 + \ell + k(s-1) \\
&\geq O_{S_n^{(k)}} - 1.
\end{aligned}$$

Furthermore, as  $O_H - O_{S_n^{(k)}} = 0 \pmod k$ , and  $k > 1$ , we have  $O_H \neq O_{S_n^{(k)}} - 1$ . Hence  $O_H \geq O_{S_n^{(k)}}$ , and so  $E(H) \geq E(S_n^{(k)})$ .  $\square$

The upper bound for the  $\text{sat}_k(n, \text{Berge-}K_3)$  is based on the existence of a saturated construction with the desired number of edges. For many  $k$  and  $n$  combinations, this construction seems to be unique, but this is not true for some uniformity and size pairs. For example, for  $k = 3$  and  $n = 8$  there is a second construction, a linear cycle, which also uses  $\left\lceil \frac{8-1}{3-1} \right\rceil = 4$  edges, see Figure 4.17.

We now turn our attention to an upper bound for cycles. We present three constructions based on the relationship between  $k$  and  $m$ , then prove that these constructions are indeed saturated.

**Construction 4.5.4** *If  $k \geq m-1$ , let  $S_{n,m}^k$  be a  $k$ -uniform hypergraph containing  $\left\lfloor \frac{n-(m-2)}{k-(m-2)} \right\rfloor$  hyperedges that pairwise intersect in a set  $I$  of  $m-2$  vertices and the remaining  $((n-(m-2)) \pmod{(k-(m-2)))$  vertices are in an edge that contains  $I$ , and some vertices not in  $I$  from*

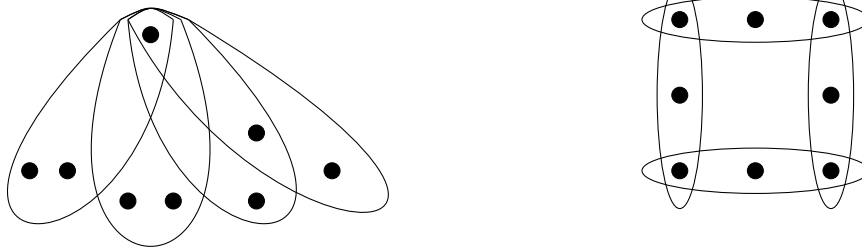


Figure 4.17: Minimal Berge- $K_3$  saturated hypergraphs on 8 vertices with uniformity 3.

*exactly one edge.*

It is worth noting here that  $S_{n,3}^k = S_n^k$  from Theorem 4.5.3, so our work for cycles can be thought of as a generalization of our work done on triangles.

**Construction 4.5.5** Let  $k = m - 2$  and  $n \geq m^2$ . Let  $F_{n,m}^{(m-2)}$  be the  $k$ -uniform hypergraph obtained by first identifying  $\lfloor \frac{n-1}{m-2} \rfloor$  cliques of  $m - 1$  vertices at a single vertex  $v$ . If  $(m - 2) \nmid (n - 1)$ , then we will absorb the remaining vertices into one clique. Let  $r = (n - 1) \bmod (m - 2)$  and let  $K$  be one of the  $(m - 1)$ -cliques. Let  $x \neq v$  be some vertex in  $K$ . Add  $\lceil \frac{r}{k-2} \rceil$  edges,  $e_1, \dots, e_r$ , that contain both  $x$  and  $v$  so that the  $r$  remaining vertices are each in exactly one of these edges, and the edges only contain these  $r$  vertices and vertices from  $K$ .

**Construction 4.5.6** Let  $k \leq m - 3$ ,  $\ell = \max\{m/2 + 1, k + 1\}$  and  $n \geq \ell^2$ . Let  $F_{n,m}^{(k)}$  be the  $k$ -uniform hypergraph obtained by identifying  $\lfloor \frac{n-1}{\ell-1} \rfloor$  cliques of  $\ell$  or  $\ell + 1$  vertices at a single vertex  $v$ . There are exactly  $(n - 1) \bmod (\ell - 1)$  cliques on  $\ell + 1$  vertices.

In Construction 4.5.6, we could have chosen any  $\ell \in [m/2 + 1, m - 1]$  as long as  $\ell \geq k + 1$  and gotten a saturated construction, but minimizing  $\ell$  under these constraints gives us the smallest number of edges.

**Theorem 4.5.7** *Let  $m \geq 4$ . If  $k \geq m - 1$  and  $n > m(k - (m - 2)) + (m - 2)$ , then*

$$\text{sat}_k(n, \text{Berge-}C_m) \leq \left\lceil \frac{n - m + 2}{k - m + 2} \right\rceil. \quad (4.1)$$

*If  $k = m - 2$  and  $n \geq m^2$ ,*

$$\text{sat}_k(n, \text{Berge-}C_m) \leq \left\lfloor \frac{n - 1}{m - 2} \right\rfloor \binom{m - 1}{k} + \frac{(n - 1) \bmod (m - 2)}{k - 2}. \quad (4.2)$$

*If  $k \leq m - 3$ ,  $\ell = \max\{m/2 + 1, k + 1\}$  and  $n \geq \ell^2$ , then*

$$\text{sat}_k(n, \text{Berge-}C_m) \leq \left\lfloor \frac{n - 1}{\ell - 1} \right\rfloor \binom{\ell}{k} + ((n - 1) \bmod (\ell - 1)) \binom{\ell}{k - 1}. \quad (4.3)$$

*Proof.* First we will prove (4.1). Assume  $k \geq m - 1$  and  $n > m(k - (m - 2)) + (m - 2)$ . We need to show that  $S = S_{n,m}^{(k)}$  from Construction 4.5.4 is indeed Berge- $C_m$ -saturated. If  $(k - m + 2) \nmid (n - m + 2)$ , let  $e_1$  and  $e_2$  be the two edges of  $S$  that intersect outside of  $I$ . To see that  $S$  contains no Berge- $C_m$ , first note that the only vertices of degree at least 2 are in either  $I$  or  $e_1 \cap e_2$ . Since  $|I| = m - 2$ , any Berge- $C_m$  would need to use at least two vertices from  $e_1 \cap e_2$ , but there are only two edges incident to any such pair of vertices, so they could not both be used as the degree 2 vertices in the same Berge- $C_m$ .

Let  $e \in E(\overline{S})$ .  $e$  must contain at least two vertices outside of  $I$ . Let  $x$  be one such vertex, and let  $e'$  be an edge of  $S$  such that  $x \in e'$ . Note that  $e \setminus I$  cannot be contained in  $e'$  since  $I \subseteq e'$ , and so that would imply that  $e = e'$ . Thus, there is some  $y \in e \setminus I$  such that  $y \notin e'$ . Let  $e''$  be some edge that contains  $y$ . Then by our choice of  $n$ , there are enough edges incident with  $I$  to create a Berge- $P_m$  in  $S$  that goes from  $x$  to  $y$ , using all the vertices of  $I$  in the middle. With the addition of  $e$ , this path can be extended to a Berge- $C_m$  in  $S + e$ . Thus  $S$  is saturated. This proves (4.1).

Now let us focus on (4.2). Let  $k = m - 2$  and  $n \geq m^2$ . Let  $F = F_{n,m}^{(m-2)}$ ,  $v, r, K, x$

and  $e_1, \dots, e_r$  be as in Construction 4.5.5. Each clique has  $m - 1$  vertices in it, so no clique contains a Berge- $C_m$ , and since  $v$  is the only vertex incident with more than one clique, no cycles use vertices from more than one clique. Finally, the  $r$  vertices not in any  $(m - 1)$ -clique are all of degree 1, so they cannot be used in a cycle, so  $F$  does not contain a Berge- $C_m$ .

Now let  $e \in E(\overline{F})$ . If  $e$  contains two vertices from different  $(m - 1)$ -cliques, say  $u_1$  and  $u_2$ , then there is a Berge- $P_m$  from  $x$  to  $y$  in  $F$  that uses as many vertices in each  $(m - 1)$ -clique as needed. We can then use  $e$  to close up the cycle, giving us a Berge- $C_m$  in  $F + e$ . This also works if we have one vertex from some  $(m - 1)$ -clique that is not  $K$ , and one of the  $r$  vertices in no  $(m - 1)$ -clique. If  $e$  contains a vertex in  $K$ , say  $u_1$  and one of the  $r$  vertices, say  $u_2 \in e_1$ , in no  $\ell$ -clique, then there is a Berge  $P_{m-1}$  from  $u_1$  to  $x$  inside  $K$ , and we can use  $e_1$  to extend this from  $x$  to  $u_1$ , so there is a  $F + e$  contains a Berge- $C_m$ . Finally, if  $e$  contains only vertices from the  $r$  vertices that are not in any  $(m - 1)$ -clique,  $e$  must contain two vertices from two different edges, say  $u_1 \in e_1$  and  $u_2 \in e_2$ . In this case, we can find a Berge- $P_{m-2}$  in  $K$  from  $v$  to  $x$ , then extend each end of this path using  $e_1$  and  $e_2$  to build a Berge- $P_m$  that goes from  $u_1$  to  $u_2$  in  $F$ . Thus,  $e$  can be used to close up the cycle, so  $F + e$  contains a Berge- $C_m$ . The only other possibility for  $e$  is that  $e$  is contained in some  $(m - 1)$ -clique, but all those edges are already present in  $F$ , so we are done.

Finally, we will show (4.3). Let  $k \leq m - 3$ ,  $\ell = \max\{m/2 + 1, k + 1\}$  and  $n \geq \ell^2$ . Let  $F = F_{n,m}^{(k)}$  and  $v$  be as in Construction 4.5.6. First note that since  $k \leq m - 3$ ,  $\ell \leq m - 2$ . Thus, the  $\ell$ -cliques and the  $(\ell + 1)$ -cliques do not have enough vertices to contain a Berge- $C_m$ . Further, no cycle can use vertices from two different cliques since  $v$  is the only vertex incident with more than one clique. Thus  $F$  does not contain a Berge- $C_m$ .

Now, let  $e \in E(\overline{F})$ .  $e$  must contain two vertices from two different cliques, say  $x$  and  $y$ . Since  $\ell \geq m/2 + 1$ , and thus  $2\ell - 1 > m$ , using the vertices from these two cliques, we can build a  $P_m$  from  $x$  to  $y$ . We can then use  $e$  to complete the cycle, so  $F + e$  contains a Berge- $C_m$ , and thus  $F$  is saturated.

The final thing to note is that  $F$  has the number of edges claimed in (4.3) since each of the  $\lfloor \frac{n-1}{\ell-1} \rfloor$  cliques contain  $\ell$  vertices, they contain a total of  $\lfloor \frac{n-1}{\ell-1} \rfloor \binom{\ell}{k}$  edges, and additionally  $(n-1) \bmod (\ell-1)$  of these cliques each have one extra vertex, which is in  $\binom{\ell}{k-1}$  edges.  $\square$

Finally, we turn to the saturation number for stars on  $k+2$  vertices. A  $k$ -uniform tight cycle is a hypergraph whose vertex set has a cyclic ordering such that any  $k$  consecutive vertices form a hyperedge, and these are the only hyperedges. Here the construction will be a tight cycle along with  $k-1$  isolated vertices.

**Theorem 4.5.8** *If  $k \geq 3$ , then  $\text{sat}_k(n, \text{Berge-}K_{1,k+1}) = n - k + 1$  for  $n \geq k^2$ .*

*Proof.* Let  $C_{n-k+1}^{(k)}$  be a  $k$ -uniform tight cycle on  $n-k+1$  vertices. Let  $\mathcal{C}$  be the hypergraph on  $n$  vertices that has  $C_{n-k+1}^{(k)}$  as a component along with  $k-1$  isolated vertices. We will first show that  $\mathcal{C}$  is saturated. Indeed, given any edge  $e \notin E(\mathcal{C})$ ,  $e$  must contain some vertex  $v$  in the tight cycle. Let  $\{u_1, \dots, u_{k-1}\}$  be the  $k-1$  vertices preceding  $v$  in the tight cycle. We can assume that  $e \neq v \cup \{u_1, \dots, u_{k-1}\}$  since otherwise we can take the  $k-1$  vertices succeeding  $v$  instead. Let  $w \in e$  be some vertex such that  $w \neq u_i$  for  $1 \leq i \leq k-1$ . Then  $\mathcal{C} + e$  contains a Berge- $K_{1,k+1}$  since we can take the edges containing the pairs  $vu_i$  for  $1 \leq i \leq k-1$ , without using the edge  $e'$  of  $\mathcal{C}$  that contains  $v$  and the  $k-1$  vertices succeeding  $v$ , the edge  $vw$  and then the edge  $vx$  for some  $x \in e'$ ,  $x \neq w$ . Thus  $\mathcal{C} + e$  contains a Berge- $K_{1,k+1}$  for any edge  $e \notin E(\mathcal{C})$ . Hence,  $\text{sat}_k(n, \text{Berge-}K_{1,k+1}) \leq E(\mathcal{C}) = n - (k-1)$ .

Now, let  $H$  be a  $k$ -uniform Berge- $K_{1,k+1}$ -saturated graph. If there are at least  $k$  vertices all of degree at most  $k-1$ , then these vertices must all be in a clique together, since otherwise we could add an edge only containing such vertices, which would not create a Berge- $K_{1,k+1}$ . Thus,  $H$  either consists of a clique of  $k$  vertices each with degree at least 1 and  $n-k$  vertices of degree at least  $k$ , or  $H$  has of  $\ell < k$  vertices of degree less than  $k$  and

$n - \ell$  vertices of degree at least  $k$ . In the former case, if we count degrees, we have that

$$|E(H)| \geq \frac{k(1) + (n - k)k}{k} = n - k + 1.$$

In the latter case, we have

$$|E(H)| \geq \frac{(n - \ell)k}{k} \geq n - k + 1.$$

□

The following theorem establishes linearity for a larger class of stars than Theorem 4.5.8. However, we do not believe the bound to be tight.

**Theorem 4.5.9** *If  $k \leq m - 1$ ,  $\text{sat}_k(n, \text{Berge-}K_{1,m}) \leq \lceil \frac{n}{m} \rceil \binom{m}{k}$ .*

*Proof.* Let  $\mathcal{K}_n^{(k)}$  be the  $k$ -uniform hypergraph that is the disjoint union of  $\lfloor \frac{n}{m} \rfloor$  cliques on  $m$  vertices, and a clique on  $n \bmod m$  vertices if  $n \bmod m \geq k$ , or  $n \bmod m$  isolated vertices if  $n \bmod m < k$ . Note that the maximum size of the neighborhood (the set of vertices adjacent to a given vertex) of any vertex in  $\mathcal{K}_n^{(k)}$  is  $m - 1$ , and so the graph does not contain a Berge- $K_{1,m}$ . Consider adding an edge  $e \notin E(\mathcal{K}_n^{(k)})$ . It must contain at least one vertex from a clique on  $m$  vertices, say  $v$ , and a vertex in a different component, say  $u$ . The clique on  $m$  vertices contains a Berge- $K_{1,m-1}$  with  $v$  at the center, so adding  $\{u, v\}$  yields a Berge- $K_{1,m}$ . Since  $E(\mathcal{K}_n^{(k)}) \leq \lceil \frac{n}{m} \rceil \binom{m}{k}$ , we have that  $\text{sat}_k(n, \text{Berge-}K_{1,m}) = O(n)$  when  $m - 1 \geq k$ . □

## 4.6 Concluding Remarks

There are many directions that can be explored involving Berge saturation. The most pressing question in the authors' minds is about the asymptotic growth of Berge saturation

numbers. For classical saturation, we have that the saturation numbers grow at most linearly with  $n$ .

**Theorem 4.6.1** (Kászonyi and Tuza, 1986, [56]) *For any fixed finite family of graphs  $\mathcal{F}$ ,*

$$\text{sat}(n, \mathcal{F}) = O(n).$$

On the other hand, for  $k$ -uniform hypergraphs, saturation numbers grow with at most  $n^{k-1}$ .

**Theorem 4.6.2** (Pikhurko, 1999, [68]) *For any fixed finite family of  $k$ -uniform hypergraphs  $\mathcal{F}^{(k)}$ ,*

$$\text{sat}(n, \mathcal{F}^{(k)}) = O(n^{k-1}).$$

Even though  $k$ -uniform Berge saturation is a special case of hypergraph saturation, the authors conjecture that asymptotically it resembles the graph case.

**Conjecture 4.6.3** *For any fixed family of graphs  $\mathcal{F}$ ,*

$$\text{sat}_k(n, \text{Berge-}\mathcal{F}) = O(n).$$

This conjecture is supported by the results in this chapter and our explorations, as every lower bound we have found is at most linear. Furthermore, preliminary work has shown that for any graph  $F$ , and  $k \in \{3, 4, 5\}$ ,  $\text{sat}_k(n, \text{Berge-}F) = O(n)$  in [24].

It would also be interesting to see if Berge saturation exhibits any monotonicity irregularities similar to those discussed by Kászonyi and Tuza in [56].



# Chapter 5

## A Random Variant of the Game of Plates and Olives

### 5.1 Introduction

The results of this chapter are joint work with Andrzej Dudek and Alan Frieze [18].

The game of plates and olives is a purely combinatorial process that has an interesting application to topology and Morse theory. Morse theory involves the study of topological manifolds by considering the smooth functions on the manifolds. An *excellent Morse function* on the 2-sphere is a smooth function from  $S^2 \rightarrow \mathbb{R}$  such that all the critical points are non-degenerate (i.e. the matrix of second partial derivatives is non-singular) and take distinct values.

If  $f$  is an excellent Morse function on the sphere,  $S^2$ , with critical points  $x_1, \dots, x_m$  with  $f(x_1) < \dots < f(x_m)$ , a *slicing* of  $f$  is an increasing sequence  $a_0, \dots, a_m$  such that  $a_0 < f(x_1) < a_1 < f(x_2) < \dots < a_{m-1} < f(x_m) < a_m$ . Then two excellent Morse functions  $f$  and  $g$ , with the same number of critical points, are said to be *topologically equivalent* if for any slicing  $a_0, \dots, a_m$  of  $f$  and  $b_0, \dots, b_m$  of  $g$ , there is an order-preserving diffeomorphism (i.e. an isomorphism of smooth manifolds) between the sublevel sets  $\{x \in S^2 \mid f(x) \leq a_i\}$  and  $\{x \in S^2 \mid g(x) \leq b_i\}$  for each  $0 \leq i \leq m$ .

Loosely speaking, two excellent Morse functions are topologically equivalent if when their critical values are ordered as mentioned above, both functions have the same types of

critical points (in terms of being local minima, maxima, or saddle points), appearing in the same order, and in a rough sense the same location relative to other critical values, which is necessary for the sublevel sets described above to be diffeomorphic.

Morse functions on the sphere have exactly  $2n+2$  critical points,  $n$  of which are saddle points. It was shown in [65] that the sublevel sets  $\{f(x) \leq a\}$  are topologically equivalent to either all of  $S^2$ , or a finite (possibly empty) disjoint union of disks, each with at most a finite number of punctures (i.e. isolated “missing” points). As the value  $a$  crosses a critical point, one of the following four things will take place:

- (1) a new disk may appear,
- (2) two such disks may merge (preserving the punctures in both disks),
- (3) a new puncture may appear, or
- (4) a puncture may disappear.

Given a slicing  $a_0, \dots, a_m$  of the excellent Morse function  $f$ , we have that the first sublevel set,  $\{f(x) \leq a_0\} = \emptyset$ , and the second,  $\{f(x) \leq a_1\}$  is a disk. The second to last sublevel,  $\{f(x) \leq a_{m-1}\}$  is also a disk, and the last sublevel set,  $\{f(x) \leq a_m\}$ , is the entire sphere, and this is the only sublevel set that is topologically equivalent to the sphere.

The game of plates and olives was originally formulated by Nicolaescu in [65] and encodes the evolution of the topology of the sublevel sets in a purely combinatorial process in which plates play the role of disks and olives represent punctures in the disks. The moves in the game of plates and olives are designed to resemble exactly the possible transformations that happen when  $a$  crosses a critical point:

- (1) add a plate,
- (2) combine two plates while keeping all the olives,

(3) add an olive to a plate, or

(4) remove an olive from a plate.

The game of plates and olives begins at an empty table and ends the first time we return to an empty table, signifying that the level sets of a Morse function start with the empty set and end with the entire sphere.

Let  $T_n^2$  denote the number of excellent Morse functions on the 2-sphere with  $n$  saddle points, up to topological equivalence. A lower bound for  $T_n^2$  was given by Nicolaescu in [65] by studying walks on Young's lattice. An upper bound on  $T_n^2$  was given by Carroll and Galvin in [12] from studying the game of plates and olives directly. The bounds of these two papers give

$$(2/e)^{n+o(n)}n^n \leq T_n^2 \leq (4/e)^{n+o(n)}n^n.$$

Here we will study a random variant of the game of plates and olives.

## The Model

The process starts with an empty table. There are four different types of moves that can happen in the process. The four main moves are as follows:

( $P^+$ ) Add a plate; from every configuration we can add one empty plate to the table.

( $P^-$ ) Combine two plates; assuming there are at least two plates, we can choose two of them (order does not matter), combine their olives onto one plate, and remove the other plate from the table.

( $O^+$ ) Add an olive to a plate; we can choose any plate and add one olive to it.

( $O^-$ ) Remove an olive from a plate; we can choose any non-empty plate and remove one olive from it.

In our model, the plates are distinguishable, but the olives are not. At each time step in the process, one of the available moves will be chosen to be performed uniformly at random.

In addition to the random aspect, our model differs from the game of plates and olives only in that in our model, the plates are distinguishable, and we do not allow for the process to return to an empty table.

Our main result shows that the number of olives grows linearly with the number of steps in the process, and that the number of olives is concentrated.

**Theorem 5.1.1** *Let  $O_t$  be the total number of olives on the table in the preceding model at time  $t$ .*

(a) *There exist absolute constants  $C > 0$  and  $\frac{1}{342} \leq c_1 \leq c_2 \leq \frac{2}{3}$  such that*

$$\Pr(c_1 t \leq O_t \leq c_2 t) \geq 1 - e^{-Ct}. \quad (5.1)$$

(b) *Furthermore, there exists an absolute constants  $A > 0$  such that for every  $\delta \geq 0$  we have*

$$\Pr(|O_t - \mathbf{E}(O_t)| \geq \delta t) \leq e^{-A\delta^2 t} \quad (5.2)$$

(c) *Also, a.a.s. no plate, except for the first plate, has more than  $B \log t$  olives at any time, for some absolute constant  $B > 0$ .*

We prove in Section 5.2 that

$$1/342 \leq \mathbf{E}(O_t) \leq 2/3. \quad (5.3)$$

Next in Section 5.3 we derive the concentration result (5.2), which altogether will imply (5.1).

In Section 5.4 we consider an auxiliary Markov chain process. It follows from our proofs that

constants  $1/342$  and  $2/3$  are not optimal. As a matter of fact a computer simulation suggests that the number of olives  $O_t$  is concentrated around  $ct$ , where  $c \approx 0.096$ .

## 5.2 Bounds on the Expected Number of Olives

### 5.2.1 Lower Bound

We would like to show that the number of olives at a given time grows linearly with time  $t$ .

Towards this, we will establish two facts:

- we expect to return to a single plate a linear number of times, and
- each time we return to a single plate, we expect to gain a positive number of olives.

This will give us a linear expectation.

Now let us show that we expect to return to a single plate a linear number of times.

If we have  $\ell \geq 1$  plates, then the probability we do a plate move is at least

$$\frac{\binom{\ell}{2} + 1}{2\ell + \binom{\ell}{2} + 1} \geq 1/3.$$

Let  $t_{plate}$  be the random variable that counts the number of plate moves we have after  $t$  moves overall. Then

$$\mathbf{E}(t_{plate}) \geq \sum_{i=1}^t 1/3 = t/3. \tag{5.4}$$

Now let us consider only plate moves to get a lower bound on the random variable  $X$ , which counts the number of times we transition from two plates to one plate.

We consider a related Markov chain. In this process, we will consider a random walk on the positive integers. We will start this walk at 1 (plate). If we are currently at 1, then we will move to 2 with probability 1. If we are currently at 2, we will move to 1 with probability  $1/2$  and to 3 with probability  $1/2$ . If we are at  $k \geq 3$ , we will move to  $k - 1$  with probability

3/4 and to  $k + 1$  with probability 1/4. This Markov chain will be indexed by time  $t_{plate}$  as it only models moves made when there is at least one plate.

Observe that in our model, for  $\ell \geq 3$ ,

$$\Pr(P^- | \text{there are } \ell \text{ plates currently and we perform a plate move}) = \frac{\binom{\ell}{2}}{\binom{\ell}{2} + 1} \geq 3/4, \quad (5.5)$$

and

$$\Pr(P^+ | \text{there are } \ell \text{ plates currently and we perform a plate move}) = \frac{1}{\binom{\ell}{2} + 1} \leq 1/4.$$

Thus the Markov chain gives an underestimate for how often we transition from two plates to one plate. Let  $N_{1,1}(t_{plate})$  be the random variable that tracks the total number of times this Markov chain returns to a state with a single plate, given that we start at a state with a single plate and a total of  $t_{plate}$  plate moves have been made.

By Theorem 5.4.1, we have that  $\mathbf{E}(N_{1,1}(t_{plate})) \geq \mathbf{E}(t_{plate})/19$ . Furthermore, note that  $\mathbf{E}(N_{1,1}(t_{plate})) \leq \mathbf{E}(X)$ , so

$$\mathbf{E}(X) \geq \mathbf{E}(t_{plate})/19 \geq t/57.$$

Now we explore what happens each time we transition from two plates to one plate. Consider a state in the process that currently has two plates. If the second plate currently has olives on it, then the probability that next time we make a plate move, plate 2 still has olives is at least 1/2. (We can immediately make the plate move.) If the second plate currently has no olives on it, then the probability there is at least one olive on it when we make the next plate move is at least 1/6. (1/3 probability to add an olive, 1/2 probability to perform a plate move once we have added the olive.) Thus we have at least a 1/6 chance of adding an olive to the first plate each time we reduce the number of plates to 1. Let  $Y$

be a random variable that counts the number of times we add at least one olive to the first plate from a plate move, given that we transition from 2 plates to 1 plate  $X$  times. Then

$$\mathbf{E}(Y) \geq \mathbf{E}(X)/6.$$

Now we put everything together. We will only consider olives on the first plate. Let  $O_t^{(1)}$  denote the number of olives on plate 1 at time  $t$ . Let  $O_t^{(1)+}$  and  $O_t^{(1)-}$  denote the total number of olives that were added to (respectively subtracted from) plate 1 from  $O^+$  (respectively  $O^-$ ) moves. Note that  $\mathbf{E}(O_t^{(1)+} - O_t^{(1)-}) \geq 0$  since the probability of performing an  $O^+$  move is always at least the probability of performing an  $O^-$  move. Finally, let  $O_t^{plate}$  denote the total number of olives added to the first plate from plate moves. Then  $O_t^{plate} \geq Y$ . This gives us that

$$\mathbf{E}(O_t) \geq \mathbf{E}(O_t^{(1)}) = \mathbf{E}(O_t^{(1)+} - O_t^{(1)-} + O_t^{plate}) \geq \mathbf{E}(Y) \geq \mathbf{E}(X)/6 \geq t/342. \quad (5.6)$$

## 5.2.2 Upper Bound

Now we bound the expected value of  $O_t$  from above. Let  $O_t^+$  and  $O_t^-$  are the random variables that count the number of  $O^+$  moves and the number of  $O^-$  moves after  $t$  total moves, respectively. Clearly,

$$O_t = O_t^+ - O_t^- = t - t_{plate} - 2O_t^- \leq t - t_{plate}.$$

Thus, by (5.4) we conclude that

$$\mathbf{E}(O_t) \leq t - \mathbf{E}(t_{plate}) \leq 2t/3.$$

And this proves (5.3).

### 5.3 Concentration

Suppose that we transition from a state with two plates to a state with a unique plate at times  $t_1, t_2, \dots, t_m$  and recall that  $O_t$  denotes the number of olives at time  $t$ . Define  $t_0 := 1$ . Let  $X_i = O_{t_{i+1}} - O_{t_i}$ . Then the  $X_i$  are independent random variables and based on the previous section we have  $\mathbf{E}(X_i) \geq 1/342$ . Then  $S_m := O_{t_m} = \sum_{i=0}^m X_i$ . We can argue for concentration of  $S_m$  as follows.

Note that from (5.4)  $\mathbf{E}(t_{plate}) \geq t/3$ , and by the Chernoff bound we have for any  $0 \leq \delta \leq 1$ ,

$$\mathbf{Pr}(t_{plate} < (1 - \delta)t/3) \leq e^{-\frac{\delta^2 t}{6}}.$$

As we are not trying to optimize the constant  $A$ , we can be imprecise here and choose  $\delta = 1/4$ , giving

$$\mathbf{Pr}(t_{plate} < t/4) \leq e^{-\frac{t}{96}}. \tag{5.7}$$

If  $t_{plate} \geq t/4$ , then the probability that we start at a unique plate, add plates, and then return to a unique plate before  $t$  moves is at least  $F_{1,1}(t/4)$ , which is the probability that our related Markov chain, defined in Section 5.2 and studied in Section 5.4, returns to 1 at least once in the first  $t/4$  moves, assuming it started at 1. Clearly,

$$\mathbf{Pr}(X_i \geq k) \leq \mathbf{Pr}(t_{i+1} - t_i \geq k)$$

and also

$$\mathbf{Pr}(t_{i+1} - t_i < k) \geq F_{1,1}(k/4) - \mathbf{Pr}(\text{less than } k/4 \text{ plate moves happen after } k \text{ moves}).$$



Thus, if  $k' = k/4$ , then by (5.15) we get that

$$\begin{aligned} \Pr(X_i \geq k) &\leq 1 - F_{1,1}(k') + e^{-\frac{k}{96}} = \sum_{j=\lfloor \frac{k'}{2} \rfloor}^{\infty} 4 \left(\frac{3}{16}\right)^{j-1} \binom{2j-3}{j-1} + e^{-\frac{k}{96}} \\ &\leq 4 \sum_{j=\lfloor \frac{k'}{2} \rfloor}^{\infty} \left(\frac{3}{4}\right)^{j-1} + e^{-\frac{k}{96}} \leq C' \zeta^{k'} + e^{-\frac{k}{96}} \leq C \rho^k, \end{aligned} \quad (5.8)$$

where  $\zeta = 3^{1/2}/2 < 1$ ,  $\rho = \max\{\zeta^{1/4}, e^{-1/96}\}$  and  $C, C' > 0$  are constants. Let  $\mu_i = \mathbf{E}(X_i)$  and  $\mu = \mu_1 + \dots + \mu_m$ . Note that we have

$$\mu_i \leq \sum_{k=1}^{\infty} k C \rho^k = \frac{C \rho}{(\rho - 1)^2} < \infty. \quad (5.9)$$

Now we can easily prove a concentration result for this situation. We modify an argument from [34]. We prove

$$\Pr(|S_m - \mu| \geq \delta m) \leq e^{-A \delta^2 m} \quad (5.10)$$

for some constant  $A > 0$ . That means we have replaced (5.6) by a concentration inequality.

We write, for  $\lambda > 0$  such that  $e^\lambda < 1/\rho$ ,

$$\mathbf{E}(X_i^2 e^{\lambda X_i}) = \sum_{k=0}^{\infty} k^2 e^{\lambda k} \Pr(X_i = k) \leq C \sum_{k=0}^{\infty} k^2 (\rho e^\lambda)^k \leq \frac{3C}{(1 - \rho e^\lambda)^3}.$$

Now  $e^x \leq 1 + x + x^2 e^x$  for  $x \geq 0$ , and so, using the above, we have

$$\mathbf{E}(e^{\lambda X_i}) \leq 1 + \lambda \mu_i + \lambda^2 \left( \frac{3C}{(1 - \rho e^\lambda)^3} \right) < 1 + \lambda \mu_i + \lambda^2 \left( 1 + \frac{3C}{(1 - \rho e^\lambda)^3} \right).$$

The generic Chernoff bound (A.1) implies

$$\begin{aligned}
\Pr(S_m \geq \mu_i m + \delta m) &\leq e^{-\lambda(\mu_i m + \delta m)} \prod_{i=1}^m \mathbf{E}(e^{\lambda X_i}) \\
&\leq \exp\{-\lambda(\mu_i m + \delta m)\} \cdot \left(1 + \lambda\mu_i + \lambda^2 \left(1 + \frac{3C}{(1 - \rho e^\lambda)^3}\right)\right)^m \\
&\leq \exp\{-\lambda(\mu_i m + \delta m)\} \cdot \exp\left\{\left(\lambda\mu_i + \lambda^2 \left(1 + \frac{3C}{(1 - \rho e^\lambda)^3}\right)\right) m\right\} \\
&= \exp\left\{-\lambda\delta m + \lambda^2 \left(1 + \frac{3C}{(1 - \rho e^\lambda)^3}\right) m\right\} \\
&\leq \exp\{-\lambda\delta m + \lambda^2(1 + 3C\varepsilon^{-3})m\},
\end{aligned}$$

where  $\varepsilon = \varepsilon(\delta) > 0$  is a constant such that

$$e^\lambda \leq (1 - \varepsilon)/\rho. \quad (5.11)$$

Now choose  $\lambda = \delta/(2(1 + 3C\varepsilon^{-3}))$  and  $\varepsilon$  such that (5.11) holds. Such a choice of  $\varepsilon$  is always possible since as  $\varepsilon \rightarrow 0$ ,  $\exp\{\delta/(2(1 + 3C\varepsilon^{-3}))\} \rightarrow 1$  and  $(1 - \varepsilon)/\rho \rightarrow 1/\rho > 1$ . Then

$$\Pr(S_m \geq \mu + \delta m) \leq \max_{1 \leq i \leq m} \Pr(S_m \geq \mu_i m + \delta m) \leq \exp\left\{-\frac{\delta^2 m}{4(1 + 3\varepsilon^{-3})}\right\}.$$

To bound  $\Pr(S_m \leq \mu - \delta m)$ , we proceed similarly. We have  $e^{-x} \leq 1 - x + x^2 e^x$ , so

$$\mathbf{E}(e^{-\lambda X_i}) \leq 1 - \lambda\mu_i + \lambda^2 \left(\frac{3C}{(1 - \rho e^\lambda)^3}\right) < 1 - \lambda\mu_i + \lambda^2 \left(1 + \frac{3C}{(1 - \rho e^\lambda)^3}\right)$$

and so by the generic Chernoff bound (A.2),

$$\begin{aligned}
\Pr(S_m \leq \mu_i m - \delta m) &\leq e^{\lambda(\mu_i m - \delta m)} \prod_{i=1}^m \mathbf{E}(e^{-\lambda X_i}) \\
&\leq \exp\{\lambda(\mu_i m - \delta m)\} \cdot \left(1 - \lambda\mu_i + \lambda^2 \left(1 + \frac{3C}{(1 - \rho e^\lambda)^3}\right)\right)^m \\
&\leq \exp\{\lambda(\mu_i m - \delta m)\} \cdot \exp\left\{\left(-\lambda\mu_i + \lambda^2 \left(1 + \frac{3C}{(1 - \rho e^\lambda)^3}\right)\right) m\right\} \\
&\leq \exp\{-\lambda\delta m + \lambda^2(1 + 3C\varepsilon^{-3})m\},
\end{aligned}$$

and we can proceed as before. This completes the proof of (5.10).

For (5.10) to be useful, we need to show that a.a.s.  $m$  is linear in  $t$ . We condition on performing a plate move. Let the random variables  $\tau_i$  for  $1 \leq i < \infty$  count how many times we have exactly  $i$  plates after  $t$  steps, where we only count when the number of plates change. So, if we are at e.g. two plates and we make three olive moves before the next plate move, we only count this as having two plates once. Then  $t_{plate} = \sum_{i=1}^{\infty} \tau_i$  and  $\tau_1 = m$ . Note that we can express  $\tau_1$  as a sum of  $\tau_2$  indicator random variables that denotes if on the  $j$ th time we are at two plates, we then transition to one plate. Note that the probability of such a transition is  $1/2$  (when we have exactly two plates, there is one way to remove a plate and one way to add a plate, giving equal probability of moving to 1 plate vs. 3 plates) and so  $\mathbf{E}(\tau_1) = \tau_2/2$ .

We will consider two cases based on the value of  $\tau_2$ . If  $\tau_2 \geq 3t_{plate}/19$ , we have from equation (5.7) and the Chernoff bound (A.4),

$$\begin{aligned}
\Pr(\tau_1 \leq t/76 = (1/3)(1/4)(3/19)t) &\leq \Pr(\tau_1 < \tau_2/3) + \Pr(t_{plate} < t/4) \\
&\leq e^{-\tau_2/18} + e^{-t/96} \leq e^{-t_{plate}/114} + e^{-t/96} \leq e^{-t/456} + e^{-t/96}.
\end{aligned}$$

Now, if  $\tau_2 < 3t_{plate}/19$  and  $\tau_1 < t_{plate}/19$ , then  $\tau_{\geq 3} := \sum_{i=3}^{\infty} \tau_i \geq 15t_{plate}/19$ . Now,

let  $L_{\geq 3}$  be the random variable that counts how many times we have at least three plates and we remove a plate. Due to (5.5) we have at least a  $3/4$  probability of removing a plate whenever we make a plate move, so by the Chernoff bound (A.4), we have

$$\Pr(L_{\geq 3} \leq 3\tau_{\geq 3}/5) \leq e^{-3\tau_{\geq 3}/200} \leq e^{-9t_{\text{plate}}/950} \leq e^{-9t/3800} + e^{-t/96}.$$

Thus a.a.s. when we have at least three plates, we remove plates at least  $3/5$  of the time and add them at most  $2/5$  of the time. This implies that we must transition from two plates to three plates at least  $\tau_{\geq 3}/5$  times to make up for the discrepancy. This implies trivially that  $\tau_2 \geq \tau_{\geq 3}/5 \geq 3t_{\text{plate}}/19 > \tau_2$ , a contradiction. Thus, there exists an absolute constant  $D > 0$  such that

$$m = \tau_1 \geq \frac{t}{76} \text{ with probability at least } 1 - e^{-Dt}. \quad (5.12)$$

Now observe that (5.8) also implies that for  $k = \log_\rho(1/Ct^2) = B \log t$  (for some constant  $B > 0$ ) we have

$$\Pr(t_{i+1} - t_i \geq k) \leq C\rho^k = 1/t^2 \quad (5.13)$$

and so

$$\Pr\left(\bigcup_{1 \leq i \leq m} (t_{i+1} - t_i \geq k)\right) \leq t \cdot \frac{1}{t^2} = o(1). \quad (5.14)$$

Note that between time  $t_i$  and  $t_{i+1}$  the number of olives at any plate different from the first one is at most  $t_{i+1} - t_i$  and so (5.14) implies that a.a.s. no plate, except for the first plate has more than  $B \log t$  olives at any time. Part (c) of Theorem 5.1.1 follows directly from (5.14).

Now let  $T = O_t - O_{t_m} = O_t - S_m$ . Then, by (5.8) we have  $\Pr(T \geq k) \leq C\rho^k$  and the triangle inequality implies

$$|O_t - \mathbf{E}(O_t)| = |T + S_m - \mathbf{E}(O_t)| \leq |S_m - \mu| + |T + \mu - \mathbf{E}(O_t)| = |S_m - \mu| + |T - \mathbf{E}(T)|.$$

Thus,

$$\Pr(|O_t - \mathbf{E}(O_t)| \geq \delta t) \leq \Pr(|S_m - \mu| \geq \delta t/2) + \Pr(|T - \mathbf{E}(T)| \geq \delta t/2).$$

Furthermore, since  $T \geq 0$  and  $\mathbf{E}(T) = O(1)$  (cf. (5.9)) we get that  $\Pr(|T - \mathbf{E}(T)| \geq \delta t/2) \leq \Pr(T \geq \delta t/2)$ . Hence,

$$\begin{aligned} \Pr(|O_t - \mathbf{E}(O_t)| \geq \delta t) &\leq \Pr(|S_m - \mu| \geq \delta t/2) + \Pr(T \geq \delta t/2) \\ &\leq \Pr(|S_m - \mu| \geq \delta m/2) + \Pr(T \geq \delta m/2) \\ &\leq e^{-A\delta^2 m/4} + C\rho^{\delta m/2}. \end{aligned}$$

This together with (5.12) proves Theorem 5.1.1(b) and this completes the proof of Theorem 5.1.1.

## 5.4 A Related Markov Chain

**Theorem 5.4.1** *Consider a random walk on the positive integers: If we are currently at 1, then we will move to 2 with probability 1. If we are currently at 2, we will move to 1 with probability 1/2 and to 3 with probability 1/2. If we are at  $k \geq 3$ , we will move to  $k - 1$  with probability 3/4 and to  $k + 1$  with probability 1/4.*

*Let  $N_{1,1}(t)$  be the random variable that counts the number of times we return to state 1 when we start the walk at state 1. Then a.a.s. we have that  $N_{1,1}(t) \geq t/19$ .*

*Proof.* Notice that we cannot return to 1 in an odd number of steps. Let  $f_{1,1}(2t)$  be the probability that the first time we return to state 1 after  $2t$  steps, given that we start at state 1 i.e. the probability that the *first return time* is  $2t$ . Let  $X_j$  be the location at time  $j$ . So  $X_0 = X_{2t} = 1$ ,  $X_1 = X_{2t-1} = 2$  and  $X_j \neq 1$  for each  $2 < j < 2t - 2$ . Furthermore, we

need to control the number of steps at which  $X_j = 2$ . Assume that at exactly  $(i + 1)$  steps we have that  $X_j = 2$ , where  $1 \leq i \leq t - 1$ . That means

$$X_1 = X_{1+2a_1} = X_{1+2a_1+2a_2} = \cdots = X_{1+2a_1+\cdots+2a_i} = 2,$$

where  $1 + 2a_1 + \cdots + 2a_i = 2t - 1$  and  $a_j \geq 1$ . Hence,

$$\begin{aligned} f_{1,1}(2t) &= \mathbf{Pr}(X_{2t} = 1, X_1 \neq 1, \dots, X_{2t-1} \neq 1 \mid X_0 = 1) \\ &= \sum_{i=1}^{t-1} \sum_{a_1+\cdots+a_i=t-1} \prod_{j=1}^i C_{a_j-1} \left(\frac{1}{2}\right)^{i+1} \left(\frac{1}{4}\right)^{t-i-2} \left(\frac{3}{4}\right)^{t-1}, \end{aligned}$$

where  $C_k = \frac{1}{k+1} \binom{2k}{k}$  is the *Catalan number*. Now the Catalan  $i$ -fold convolution formula (see, e.g., [71]) gives that

$$\sum_{a_1+\cdots+a_i=t-1} \prod_{j=1}^i C_{a_j-1} = \frac{i}{2t-i-2} \binom{2t-i-2}{t-1}.$$

Thus,

$$f_{1,1}(2t) = \sum_{i=1}^{t-1} \frac{i}{2t-i-2} \binom{2t-i-2}{t-1} \left(\frac{1}{2}\right)^{i+1} \left(\frac{1}{4}\right)^{t-i-2} \left(\frac{3}{4}\right)^{t-1}$$

and equivalently by replacing  $i$  by  $k = t - i - 1$  we get

$$f_{1,1}(2t) = 4 \left(\frac{3}{16}\right)^{t-1} \cdot \sum_{k=0}^{t-2} \binom{(t-1)+k}{k} \frac{(t-1)-k}{(t-1)+k} 2^{(t-2)-k}.$$

Now we apply the following identity (see, e.g., (1.12) in [43])

$$\sum_{k=0}^n \binom{x+k}{k} \frac{x-k}{x+k} 2^{n-k} = \binom{x+n}{n}$$

with  $n = t - 2$  and  $x = t - 1$  to conclude that

$$f_{1,1}(2t) = 4 \left( \frac{3}{16} \right)^{t-1} \binom{2t-3}{t-1}.$$

Consequently, the probability  $F_{1,1}(t)$ , given  $X_0 = 1$ , that we return to 1 at some point in the first  $t$  steps is given by

$$F_{1,1}(t) = \sum_{j=1}^{\lfloor \frac{t}{2} \rfloor} f_{1,1}(2j) = \sum_{j=1}^{\lfloor \frac{t}{2} \rfloor} 4 \left( \frac{3}{16} \right)^{j-1} \binom{2j-3}{j-1}. \quad (5.15)$$

By (5.7) in [38], we can calculate the mean time  $\bar{T}_{1,1}$ , to return to state 1 after starting at state 1 by

$$\bar{T}_{1,1} = \sum_{t=0}^{\infty} \Pr(T_{1,1} \geq t) = 1 + \sum_{t=1}^{\infty} (1 - F_{1,1}(t)) = 1 + \sum_{t=1}^{\infty} \sum_{j=\lfloor \frac{t}{2} \rfloor + 1}^{\infty} 4 \left( \frac{3}{16} \right)^{j-1} \binom{2j-3}{j-1}.$$

Now we will evaluate the latter double sum. First observe that

$$\sum_{t=1}^{\infty} \sum_{j=\lfloor \frac{t}{2} \rfloor + 1}^{\infty} a_j = 2 \sum_{j=1}^{\infty} j a_j - \sum_{j=1}^{\infty} a_j = 2 \sum_{j=1}^{\infty} (j-1) a_j + \sum_{j=1}^{\infty} a_j$$

assuming that all  $a_j$ s are nonnegative. Thus,

$$\begin{aligned} \bar{T}_{1,1} &= 1 + 8 \sum_{j=1}^{\infty} (j-1) \left( \frac{3}{16} \right)^{j-1} \binom{2j-3}{j-1} + 4 \sum_{j=1}^{\infty} \left( \frac{3}{16} \right)^{j-1} \binom{2j-3}{j-1} \\ &= 1 + 8 \sum_{k=1}^{\infty} k \left( \frac{3}{16} \right)^k \binom{2k-1}{k} + 4 \sum_{k=0}^{\infty} \left( \frac{3}{16} \right)^k \binom{2k-1}{k}. \end{aligned}$$

Now we will use the following identities that can be obtained by an application of Newton's

binomial theorem (see, e.g., (1.30) and (1.3) in [41]):

$$\sum_{k=1}^{\infty} k \binom{2k}{k} \left(\frac{x}{4}\right)^k = \frac{x}{2(1-x)^{3/2}} \quad (5.16)$$

and

$$\sum_{k=0}^{\infty} \binom{2k}{k} \left(\frac{x}{4}\right)^k = \frac{1}{\sqrt{1-x}} \quad (5.17)$$

for  $|x| < 1$ . Applying (5.16) with  $x = \frac{3}{4}$  and using  $\binom{2k-1}{k} = \frac{1}{2}\binom{2k}{k}$  for  $k \geq 1$  yields

$$8 \sum_{k=1}^{\infty} k \left(\frac{3}{16}\right)^k \binom{2k-1}{k} = 4 \sum_{k=1}^{\infty} k \left(\frac{3}{16}\right)^k \binom{2k}{k} = 4 \cdot \frac{3/4}{2(1-3/4)^{3/2}} = 12.$$

Similarly, we apply (5.17) to get

$$4 \sum_{k=0}^{\infty} \binom{2k}{k} \left(\frac{3}{16}\right)^k = 4 + 2 \sum_{k=1}^{\infty} \binom{2k}{k} \left(\frac{3}{16}\right)^k = 4 + 2 \left( \frac{1}{\sqrt{1-3/4}} - 1 \right) = 6.$$

Consequently,

$$\bar{T}_{1,1} = 1 + 12 + 6 = 19.$$

Let  $N_{1,1}(t)$  be the number of times that, given starting at state 1, we return to state 1 in the first  $t$  moves. By (5.8) in [38], we have that a.a.s.

$$\lim_{t \rightarrow \infty} N_{1,1}(t)/t = 1/\bar{T}_{1,1} = 1/19.$$

□



## 5.5 Concluding Remarks

In the original game of plates and olives, the plates are indistinguishable from each other, so it would be interesting to explore a random version of this. This may involve more complicated ideas since the number of distinguishable plates is harder to track than the total number of plates.

The game of plates and olives was formulated to describe the evolution of the topology of sublevel sets of different excellent Morse functions on the sphere  $S^2$ . If instead we consider the simpler problem involving Morse functions on the circle,  $S^1$ , it ends up that the number of excellent Morse functions with  $n$  critical points is counted by the well-studied Catalan numbers  $C_n = \frac{1}{n+1} \binom{2n}{n}$  [65].

Since the study of the topological equivalence classes of Morse functions on  $S^1$  and  $S^2$  has led to interesting combinatorial structures, it could be interesting to investigate this problem for more complicated manifolds. One natural choice of the next manifold to explore would be the 3-sphere  $S^3$ . It could also be interesting to explore the torus, or other surfaces of higher genus.

# Appendix A

## Some Tools and Inequalities

The following is a short list of some of the tools used throughout this dissertation. For more information about these tools and estimates, see e.g. [55],[1],[54],[52].

**First Moment Method** *Let  $X$  be a non-negative random variable that takes only integer values. Then  $\Pr(X \geq 1) \leq \mathbf{E}[X]$ . If  $\mathbf{E}[X] \rightarrow 0$ , then  $X = 0$  asymptotically almost surely.*

**Second Moment Method** *Let  $X$  be a non-negative random variable that takes only integer values. If  $\mathbf{E}(X^2)/\mathbf{E}(X)^2 = 1 + o(1)$ , then  $X \geq 1$  asymptotically almost surely.*

**The Generic Chernoff Bound** *Let  $X = \sum_{i=1}^n X_i$  be a sum of independent random variables. Then*

$$\Pr(X \geq a) \leq \min_{t>0} e^{-ta} \prod_{i=1}^n \mathbf{E}(e^{tX_i}), \quad (\text{A.1})$$

and

$$\Pr(X \leq a) \leq \min_{t>0} e^{ta} \prod_{i=1}^n \mathbf{E}(e^{-tX_i}). \quad (\text{A.2})$$

In addition to the generic Chernoff bound, we will use two specific bounds.

**Chernoff bound (v1)** *Let  $X$  be a binomial random variable with mean  $\mu = \mathbf{E}(X)$ , and let  $\gamma < 3/2$ . Then*

$$\Pr(|X - \mu| \geq \gamma\mu) \leq 2e^{-\gamma^2\mu/3}. \quad (\text{A.3})$$

**Chernoff bound (v2)** *Let  $X = \sum_{i=1}^n X_i$  be a sum of independent random indicator vari-*

ables with mean  $\mu = \mathbf{E}(X)$ . Then for all  $0 \leq \delta \leq 1$ ,

$$\Pr(X < (1 - \delta)\mu) \leq e^{-\delta^2\mu/2}. \quad (\text{A.4})$$

**Bernoulli's inequality** Let  $k > 1$ , then

$$(1 + (k - 1))^{k/(k-1)} \geq 1 + k. \quad (\text{A.5})$$

**Stirling's formula**

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n. \quad (\text{A.6})$$

# Bibliography

- [1] Alan A. Frieze and M. Karoński, *Introduction to random graphs*, Cambridge University Press, Cambridge, 2016.
- [2] M. Ajtai, J. Komlós, and E. Szemerédi, *First occurrence of Hamilton cycles in random graphs*, Cycles in graphs (Burnaby, B.C., 1982), North-Holland Math. Stud., vol. 115, North-Holland, Amsterdam, 1985, pp. 173–178.
- [3] P. Allen, J. Böttcher, Y. Kohayakawa, and Y. Person, *Tight Hamilton cycles in random hypergraphs*, Random Structures Algorithms **46** (2015), no. 3, 446–465.
- [4] D. Bal and L. DeBiasio, *Partitioning random graphs into monochromatic components*, Electron. J. Combin. **24** (2017), no. 1, Paper 1.18, 25.
- [5] D. Bal and A. Frieze, *Rainbow matchings and Hamilton cycles in random graphs*, Random Structures Algorithms **48** (2016), no. 3, 503–523.
- [6] I. Ben-Eliezer, M. Krivelevich, and B. Sudakov, *Long cycles in subgraphs of (pseudo)random directed graphs*, J. Graph Theory **70** (2012), no. 3, 284–296.
- [7] I. Ben-Eliezer, M. Krivelevich, and B. Sudakov, *The size Ramsey number of a directed path*, J. Combin. Theory Ser. B **102** (2012), no. 3, 743–755.
- [8] P. Bennett, L. DeBiasio, A. Dudek, and S. English, *Large monochromatic components and long monochromatic cycles in random hypergraphs*, arXiv:1709.02990 (2017), submitted.
- [9] B. Bollobás, *The evolution of sparse graphs*, Graph theory and combinatorics (Cambridge, 1983), Academic Press, London, 1984, pp. 35–57.

- [10] B. Bollobás and E. Győri, *Pentagons versus triangles*, Discrete Mathematics **308** (2008), no. 19, 4332–4336.
- [11] B. Bollobás and A. Thomason, *Threshold functions*, Combinatorica **7** (1987), no. 1, 35–38.
- [12] T. Carroll and D. Galvin, *The game of plates and olives*, arXiv:1711.10670 (2017).
- [13] F. Chung, *Regularity lemmas for hypergraphs and quasi-randomness*, Random Structures Algorithms **2** (1991), no. 2, 241–252.
- [14] D. Conlon, *Combinatorial theorems relative to a random set*, Proceedings of the International Congress of Mathematicians, vol. 4, Kyung Moon SA, 2014, pp. 303–328.
- [15] C. Cooper and A. Frieze, *Multi-coloured Hamilton cycles in random edge-coloured graphs*, Combin. Probab. Comput. **11** (2002), no. 2, 129–133.
- [16] A. Davoodi, E. Győri, A. Methuku, and C. Tompkins, *An Erdős-Gallai type theorem for hypergraphs*, arXiv:1608.03241 (2016).
- [17] A. Dudek, S. English, and A. Frieze, *On rainbow hamilton cycles in random hypergraphs*, arXiv:1708.08975 (2017), submitted.
- [18] A. Dudek, S. English, and A. Frieze, *A random variant of the game of plates and olives*, arXiv:1803.10278 (2018), submitted.
- [19] A. Dudek and A. Frieze, *Tight Hamilton cycles in random uniform hypergraphs*, Random Structures Algorithms **42** (2013), no. 3, 374–385.
- [20] A. Dudek, A. Frieze, P. Loh, and S. Speiss, *Optimal divisibility conditions for loose Hamilton cycles in random hypergraphs*, Electron. J. Combin. **19** (2012), no. 4, Paper 44, 17.

- [21] A. Dudek and A. Frieze, *Loose Hamilton cycles in random uniform hypergraphs*, Electron. J. Combin. **18** (2011), no. 1, Paper 48, 14.
- [22] A. Dudek and P. Prałat, *An alternative proof of the linearity of the size-Ramsey number of paths*, Combin. Probab. Comput. **24** (2015), no. 3, 551–555.
- [23] A. Dudek and P. Prałat, *On some multicolor ramsey properties of random graphs*, SIAM J. Discrete Math. **31** (2017), 2079–2092.
- [24] S. English, D. Gerbner, A. Methuku, and M. Tait, *Linearity of Berge saturation*, In preperation.
- [25] S. English, N. Graber, P. Kirkpatrick, A. Methuku, and E. Sullivan, *Saturation of Berge hypergraphs*, arXiv:1710.03735 (2017), submitted.
- [26] P. Erdős and A. Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc **52** (1946), 1087–1091.
- [27] B. Ergemlidze, E. Györi, and A. Methuku, *Asymptotics for Turán numbers of cycles in 3-uniform linear hypergraphs*, arXiv:1705.03561 (2017).
- [28] A. Ferber, *Closing gaps in problems related to Hamilton cycles in random graphs and hypergraphs*, Electron. J. Combin. **22** (2015), no. 1, Paper 1.61, 7.
- [29] A. Ferber and M. Krivelevich, *Rainbow Hamilton cycles in random graphs and hypergraphs*, Recent trends in combinatorics, IMA Vol. Math. Appl., vol. 159, Springer, [Cham], 2016, pp. 167–189.
- [30] A. Ferber, R. Nenadov, and U. Peter, *Universality of random graphs and rainbow embedding*, Random Structures Algorithms **48** (2016), no. 3, 546–564.

- [31] A. Frieze, *Loose Hamilton cycles in random 3-uniform hypergraphs*, Electron. J. Combin. **17** (2010), no. 1, Note 28, 4.
- [32] A. Frieze and M. Krivelevich, *Packing Hamilton cycles in random and pseudo-random hypergraphs*, Random Structures Algorithms **41** (2012), no. 1, 1–22.
- [33] A. Frieze and P. Loh, *Rainbow Hamilton cycles in random graphs*, Random Structures Algorithms **44** (2014), no. 3, 328–354.
- [34] A. Frieze and W. Pegden, *Traveling in randomly embedded random graphs*, Approximation, randomization, and combinatorial optimization. Algorithms and techniques, LIPIcs. Leibniz Int. Proc. Inform., vol. 81, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017, pp. Art. No. 45, 17.
- [35] Z. Füredi, *Maximum degree and fractional matchings in uniform hypergraphs*, Combinatorica **1** (1981), no. 2, 155–162.
- [36] Z. Füredi and A. Gyárfás, *Covering  $t$ -element sets by partitions*, European J. Combin. **12** (1991), no. 6, 483–489.
- [37] Z. Füredi and L. Özkahya, *On 3-uniform hypergraphs without a cycle of a given length*, Discrete Applied Mathematics **216** (2017), 582–588.
- [38] R. Gallager, *Stochastic processes*, Cambridge University Press, Cambridge, 2013, Theory for applications.
- [39] D. Gerbner, A. Methuku, and M. Vizer, *Asymptotics for the Turán number of Berge- $K_{2,t}$* , arXiv:1705.04134 (2017).
- [40] D. Gerbner and C. Palmer, *Extremal results for Berge-hypergraphs*, arXiv:1505.08127 (2015).

- [41] H. Gould, *Tables of combinatorial identities. eight tables based on seven unpublished manuscript notebooks (1945-1990) of h. w. gould, edited and compiled by prof. jocelyn quaintance*, <http://www.math.wvu.edu/gould/Vol.2.PDF>.
- [42] R. Gould, *Recent advances on the Hamiltonian problem: Survey III*, *Graphs Combin.* **30** (2014), no. 1, 1–46.
- [43] H. Gould, *Tables of combinatorial identities. eight tables based on seven unpublished manuscript notebooks (1945-1990) of h. w. gould, edited and compiled by prof. jocelyn quaintance*, <http://www.math.wvu.edu/gould/Vol.3.PDF>.
- [44] W. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, *Ann. of Math. (2)* **166** (2007), no. 3, 897–946.
- [45] A. Gyárfás, *Particiófedések és lefogóhalmazok hipergráfokban*, *Tanulmányok-MTA Számítástech. Automat. Kutató Int. Budapest* (1977), no. 71, 62.
- [46] A. Gyárfás and P. Haxell, *Large monochromatic components in colorings of complete 3-uniform hypergraphs*, *Discrete Math.* **309** (2009), no. 10, 3156–3160.
- [47] A. Gyárfás, G. Sárközy, and E. Szemerédi, *The Ramsey number of diamond-matchings and loose cycles in hypergraphs*, *Electron. J. Combin.* **15** (2008), no. 1, Research Paper 126, 14.
- [48] E. Győri, *Triangle-free hypergraphs*, *Combinatorics, Probability and Computing* **15** (2006), no. 1-2, 185–191.
- [49] E. Győri, G. Katona, and N. Lemons, *Hypergraph extensions of the Erdős-Gallai theorem*, *European Journal of Combinatorics* **58** (2016), 238–246.
- [50] E. Győri and N. Lemons, *3-uniform hypergraphs avoiding a given odd cycle*, *Combinatorica* **32** (2012), no. 2, 187–203.



- [51] A. Hajnal, P. Erdos, and J. Moon, *A Problem in Graph Theory*, The American Mathematical Monthly **71** (1964), no. 10, 1107–1110.
- [52] G. Hardy, J. Littlewood, and G. Pólya, *Inequalities*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1988, Reprint of the 1952 edition.
- [53] P. Haxell, T. Łuczak, Y. Peng, V. Rödl, A. Ruciński, M. Simonovits, and J. Skokan, *The Ramsey number of hypergraph cycles. I*, J. Combin. Theory Ser. A **113** (2006), no. 1, 67–83.
- [54] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.
- [55] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.
- [56] L. Kászonyi and Z. Tuza, *Saturated graphs with minimal number of edges*, J. Graph Theory **10** (1986), no. 2, 203–210.
- [57] Y. Kohayakawa, *Szemerédi’s regularity lemma for sparse graphs*, Foundations of computational mathematics (Rio de Janeiro, 1997), Springer, Berlin, 1997, pp. 216–230.
- [58] J. Komlós and E. Szemerédi, *Limit distribution for the existence of Hamiltonian cycles in a random graph*, Discrete Math. **43** (1983), no. 1, 55–63.
- [59] F. Lazebnik and J. Verstraëte, *On hypergraphs of girth five*, The Electronic Journal of Combinatorics **10** (2003), no. 1, R25.
- [60] S. Letzter, *Path Ramsey number for random graphs*, Combin. Probab. Comput. **25** (2016), no. 4, 612–622.

- [61] H. Liu, R. Morris, and N. Prince, *Highly connected monochromatic subgraphs of multi-colored graphs*, J. Graph Theory **61** (2009), no. 1, 22–44.
- [62] W. Mantel, *Problem 28*, Wiskundige Opgaven **10** (1907), no. 60-61, 320.
- [63] D. Mubayi, *Generalizing the Ramsey problem through diameter*, Electron. J. Combin. **9** (2002), no. 1, Research Paper 42, 10.
- [64] R. Nenadov and N. Škorić, *Powers of Hamilton cycles in random graphs and tight hamilton cycles in random hypergraphs*, arXiv:1601.04034 (2016).
- [65] L. Nicolaescu, *Counting Morse functions on the 2-sphere*, Compos. Math. **144** (2008), no. 5, 1081–1106.
- [66] C. Palmer, M. Tait, C. Timmons, and A. Wagner, *Turán numbers for Berge-hypergraphs and related extremal problems*, arXiv:1706.04249 (2017).
- [67] O. Parczyk and Y. Person, *Spanning structures and universality in sparse hypergraphs*, Random Structures Algorithms **49** (2016), no. 4, 819–844.
- [68] O. Pikhurko, *The minimum size of saturated hypergraphs*, Combinatorics, Probability and Computing **8** (1999), no. 5, 483–492.
- [69] A. Pokrovskiy, *Partitioning edge-coloured complete graphs into monochromatic cycles and paths*, J. Combin. Theory Ser. B **106** (2014), 70–97.
- [70] F. Ramsey, *On a Problem of Formal Logic*, Proc. London Math. Soc. (2) **30** (1929), no. 4, 264–286.
- [71] A. Regev, *A proof of Catalan’s convolution formula*, Integers **12** (2012), no. 5, 929–934.
- [72] V. Rödl and J. Skokan, *Regularity lemma for  $k$ -uniform hypergraphs*, Random Structures Algorithms **25** (2004), no. 1, 1–42.

- [73] A. Scott, *Szemerédi's regularity lemma for matrices and sparse graphs*, *Combin. Probab. Comput.* **20** (2011), no. 3, 455–466.
- [74] E. Szemerédi, *Regular partitions of graphs*, *Problèmes combinatoires et théorie des graphes* (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), *Colloq. Internat. CNRS*, vol. 260, CNRS, Paris, 1978, pp. 399–401.
- [75] T. Tao, *A variant of the hypergraph removal lemma*, *J. Combin. Theory Ser. A* **113** (2006), no. 7, 1257–1280.
- [76] C. Timmons, *On  $r$ -uniform linear hypergraphs with no Berge- $K_{2,t}$* , arXiv:1609.03401 (2016).
- [77] P. Turán, *Einige Extremalaufgabe aus der Graphentheorie*, *Mat. Fiz. Lapok* **48** (1941), no. 436-452, 137.